

# Guía Práctica del Curso de Bioestadística Aplicada a las Ciencias de la Salud

Dr. Jacobo Díaz Portillo

Guía Práctica del Curso  
de Bioestadística Aplicada  
a las Ciencias de la Salud  
Dr. Jacobo Díaz Portillo



GOBIERNO  
DE ESPAÑA

MINISTERIO  
DE SANIDAD, POLÍTICA SOCIAL  
E IGUALDAD



# Guía Práctica del Curso de Bioestadística Aplicada a las Ciencias de la Salud

**Dr. Jacobo Díaz Portillo**

Coordinador de Formación Continuada  
Hospital Universitario de Ceuta

Dirección Territorial de Ceuta  
Hospital Universitario de Ceuta  
Urb. Loma Colmenar, s/n  
51003 Ceuta

Edita: © Instituto Nacional de Gestión Sanitaria  
Subdirección General de Gestión Económica y Recursos Humanos  
Servicio de Recursos Documentales y Apoyo Institucional  
Alcalá, 56 28014 Madrid

El copyright y otros derechos de propiedad intelectual pertenecen al Instituto Nacional de Gestión Sanitaria.  
Se autoriza la reproducción total o parcial para uso no comercial, siempre que se cite explícitamente su procedencia.

Depósito Legal: M-33564-2011  
NIPO: 865-11-010-6  
Colección Editorial de Publicaciones del INGESA: 1.921



“La medicina es una ciencia de probabilidades y el arte de manejar la incertidumbre..”

“La estadística es la ciencia que permite tomar decisiones en situaciones de incertidumbre.”



Bailar JC III. Some uses of statistical thinking. En: Bailar JC III, Mostoller F, editors. Medical uses of statistics. Boston: N Engl J Med Books, 1992; p. 25-57.

# Índice

## Guía Práctica del Curso de Bioestadística Aplicada a las Ciencias de la Salud

### Capítulo 1

Página 10

Publicación Bioestadística: Introducción.

### Capítulo 2

Página 72

Publicación Bioestadística: Variables y gráficos.

## Capítulo 3

Página 161

Publicación Bioestadística: Tamaño muestral, Error  $\alpha$  y  $\beta$ .

## Capítulo 4

Página 214

Publicación Bioestadística: Error  $\alpha$  y  $\beta$ . Evaluación de pruebas diagnósticas.

## Capítulo 5

Página 355

Publicación Bioestadística: Análisis bivariante. Test de contraste de hipótesis. Cualitativa versus cualitativa.

## Capítulo 6

Página 450

Publicación Bioestadística: Análisis bivariante. Test de contraste de hipótesis. Cualitativa versus cualitativa. Medidas de asociación.

## Capítulo 7

Página 640

Publicación Bioestadística: Análisis bivariante. Test de contraste de hipótesis. Cuantitativa versus cuantitativa.

## Capítulo 8

Página 727

Publicación Bioestadística: Análisis bivariante. Test de contraste de hipótesis. Cualitativa dicotómica versus cuantitativa.

## Capítulo 9

Página 874

Publicación Bioestadística: Análisis bivariante. Test de contraste de hipótesis. Cualitativa policotómica versus cuantitativa.

## Capítulo 10

Página 992

Publicación Bioestadística:

Medidas de concordancia. Variables cualitativas.

## Capítulo 11

Página 1021

Publicación Bioestadística: Medidas de concordancia.

Variables cuantitativas.

## Capítulo 12

Página 1158

Publicación Bioestadística: Análisis multivariante.

## Capítulo 13

Página 1337

Publicación Bioestadística: Bibliografía.

“Es una verdad muy cierta que, cuando no esté a nuestro alcance determinar lo que es verdad, deberemos buscar lo que es más probable.”

Discurso del Método. René Descartes.

“No existe ni el negro ni el blanco, solo existe una completa escala de grises...”

“La respuesta está dibujada en el aire...”

Si, según Montaigne, “para juzgar cosas grandes y nobles, es necesario poseer un alma igual de grande y noble”(\*), para juzgar un artículo científico escrito por un colega es necesario poseer, además, conocimientos, experiencia, imparcialidad, confidencialidad, diligencia, una cierta dosis de pedagogía y otra no menos esencial de compañerismo (\*\*).

(\*) Montaigne M. Les essais [libro en internet]. Versión HTML d'après l'édition de 1595 [acceso 31 de octubre de 2008]. Disponible en: <http://www.bribes.org/trismegiste/montable.htm>

(\*\*) F. Antoja Ribó. Revisores. Revista del laboratorio clínico. 2009; 2: 63.

# Capítulo I

## Publicación Bioestadística: Introducción.

**Dr. Jacobo Díaz Portillo**  
Coordinador de Formación Continuada  
Hospital Universitario de Ceuta

# Bioestadística aplicada a las Ciencias de la Salud

## Definición

La Estadística es la ciencia que se ocupa de los métodos y procedimientos para recoger, clasificar, resumir, hallar regularidades y analizar los datos, siempre y cuando la variabilidad e incertidumbre sea una causa intrínseca de los mismos; así como de realizar inferencias a partir de ellos, con la finalidad de ayudar a la toma de decisiones y en su caso formular predicciones.

# Bioestadística aplicada a las Ciencias de la Salud

## Estadística descriptiva

Describe, analiza y representa un grupo de datos utilizando métodos numéricos y gráficos que resumen y presentan la información contenida en ellos.

# Bioestadística aplicada a las Ciencias de la Salud

## Estadística inferencial

Apoyándose en el cálculo de probabilidades y a partir de datos de una muestra (procedente de una población), efectúa estimaciones, decisiones, predicciones u otras generalizaciones sobre un conjunto mayor de datos. Permite generalizar los datos obtenidos a partir de una muestra a la población de la que proceden. Se basa en la teoría de las probabilidades y trabaja con los datos que le proporciona la estadística descriptiva.

# Bioestadística aplicada a las Ciencias de la Salud

## Estadística en Medicina

La medicina es una ciencia estadística, pues siempre se trabaja con la probabilidad de enfermar y con la oportunidad incierta del diagnóstico y del tratamiento. El médico debe adecuar el conocimiento científico y tecnológico a la situación clínica personal y social del paciente del que se ocupa en ese momento, ya que «no existen enfermedades sino enfermos». El buen juicio médico logra un equilibrio entre el riesgo que comporta toda intervención diagnóstica y/o terapéutica y el beneficio esperable de dicha intervención en el paciente concreto, lo que obliga a combinar ciencia y arte, pues no es fácil trabajar con la incertidumbre, y sólo la estadística nos proporciona el instrumento adecuado que nos permite convivir con el azar, y con las limitaciones del conocimiento médico.

Hay verdades que por sabidas no se dicen, y por no decirlas se olvidan...

Hay mentiras (mitos o leyendas) que por decirlas pueden terminar siendo pseudorealidades o realidades virtuales.

Dr. Jacobo Díaz Portillo  
Coordinador de Formación Continuada.  
Hospital Universitario de Ceuta.

## Mitos y Leyendas en Bioestadística

Bobenrieth Astete M.A . Mitos y realidades en torno a la publicación científica. Med Clin (Barc) 2000; 114: 339-341.

Bobenrieth Astete M.A . Lectura crítica de artículos originales en Salud. Medicina de Familia (And) 2001; 2: 81-90.

Profesor Bobenrieth Astete M.A.  
Profesor de Metodología de Investigación y Escritura Científica.  
Escuela Andaluza de Salud Pública.

# Mitos y Realidades en Bioestadística

MITO 1: “La sonrisa de la esfinge”.

¿Cuál es el ser que por la mañana anda con cuatro patas, luego por la tarde con dos, y después por la noche con tres patas, y que se vuelve más débil cuanto más patas tiene?

Hay dos hermanas una de las cuales engendra a la otra, y ésta a su vez engendra a la primera ¿Quiénes son?

## Mitos y Realidades en Bioestadística

REALIDAD 1: “Respuesta de Edipo”.

¿Cuál es el ser que por la mañana anda con cuatro patas, luego por la tarde con dos, y después por la noche con tres patas, y que se vuelve más débil cuanto más patas tiene?

*Edipo respondió correctamente y dijo: “el HOMBRE, porque la mañana, la tarde y la noche traducen las etapas de la vida. El día o comienzo de la vida muestra al bebé gateando en cuatro pies o patas, la tarde o mitad de la vida es la adultez del ser humano caminando en dos pies y la noche o final de la vida muestra al anciano con un bastón o sea en tres pies.”*

Hay dos hermanas una de las cuales engendra a la otra, y ésta a su vez engendra a la primera ¿Quiénes son?

*Edipo contestó: el DÍA y la NOCHE.*

Furiosa, la Esfinge se suicida lanzándose al vacío y Edipo fue aclamado como el salvador de Tebas, y como premio, Edipo fue nombrado rey de Tebas.

## Mitos y Realidades en Bioestadística

MITO 1: “La sonrisa de la esfinge”.

¿Cómo voy a ser capaz de presentar y defender una comunicación científica en cualquier foro público o mandar a publicar un artículo original a una revista sanitaria ante tan docto auditorio?

¿Qué me van a preguntar? ¿Qué me van a decir si no respondo o no lo hago adecuadamente? Seguro que hago el ridículo...

Seguro que me rechazan el artículo original cuando lo mande a publicar a la revista...

Si lo publican seguro que alguien me lo critica con una carta al director que se publica en el número siguiente, y todo el mundo va a conocer mis deficiencias...

Dr. Jacobo Díaz Portillo  
Coordinador de Formación Continuada  
Hospital Universitario INGESA de Ceuta

## Realidad I: “Respuesta de Edipo”

### FORMACIÓN CONTINUADA

METODOLOGÍA DE INVESTIGACIÓN

BIOESTADÍSTICA

ESCRITURA CIENTÍFICA

PRESENTACIONES CIENTÍFICAS EN POWER POINT

BÚSQUEDA DE BIBLIOGRÁFICA SEGÚN CRITERIOS  
DE MEDICINA BASADA EN LA EVIDENCIA

# Bioestadística

“La bioestadística es la punta visible del iceberg de la metodología de investigación”.

Dr. Jacobo Díaz Portillo  
Coordinador de Formación Continuada  
Hospital Universitario de Ceuta

## Mitos y Realidades en Bioestadística

REALIDAD 1: Muchos investigadores tienen bien definido su estudio y manejan correctamente las variables (datos), pero no aciertan cuando escogen las pruebas estadísticas pertinentes, y llegan a la presentación oral (comunicación) o a la redacción de las revistas científicas (artículo original) trabajos que adolecen de una manifiesta mala sincronía entre los datos obtenidos y su posterior análisis estadístico.

Antoja Ribó F. The innocence of the storks.  
Rev Lab Clin 2010; 3 (2): 51.

## Mitos y Leyendas en Bioestadística

REALIDAD 1: Los auditores de los congresos sanitarios y los revisores de las revistas científicas suelen ser colegas con un alto perfil profesional, aceptable formación en Metodología de Investigación, pero quizás sin un alto nivel de conocimientos de análisis estadístico de datos.

Antoja Ribó F. The innocence of the storks.  
Rev Lab Clin 2010; 3 (2): 51.

## Mitos y Leyendas en Bioestadística

REALIDAD 1: El Comité Científico de los congresos y los editores de las revistas ante trabajos de difícil evaluación metodológica incorporan un tercer revisor especialista en Bioestadística que dictamina definitivamente sobre la idoneidad de las herramientas metodológicas empleadas por los autores.

Antoja Ribó F. The innocence of the storks.  
Rev Lab Clin 2010; 3 (2): 51.

## Enseñanza de la Bioestadística Aplicada

MITO 2: Enseñanza de la investigación clínica.



Antonio Fraguas de Pablo (Forges).

¿Qué le pasa, Doctor? ¿Eh?. Ediciones Doyma 1988 (Barcelona).

# Mitos y Realidades en Bioestadística

## Enseñanza autodidacta de la Bioestadística aplicada

REALIDAD 2: La educación autodidacta en Bioestadística es el conjunto de procesos que realizamos para incorporar nuevos conocimientos necesarios para cada momento, donde uno mismo selecciona la información que necesita y se evalúa a sí mismo con el método «ensayo-error».

“La educación autodidacta es, creo firmemente, el único tipo de educación que existe.”

Isaac Asimov.

## Mitos y Realidades en Bioestadística

MITO 3: Una publicación sobre las ventajas de un nuevo tratamiento con otro vigente que demuestra significación estadística ( $p < 0,05$ ) posee, en consecuencia, importancia y significación clínica.

## Mitos y Realidades en Bioestadística

REALIDAD 3: Mediciones y afirmaciones de probabilidad pueden ser de gran ayuda en la práctica médica, pero nunca deben reemplazar en muchos casos a...

- La experiencia.
- La perceptividad.
- El sentido común.
- Y la intuición.

## Mitos y Realidades en las publicaciones científicas

REALIDAD 3: El “valor de p”, debe ser valorado con precaución y teniendo en cuenta el contexto del estudio, su diseño, las características de la muestra o la población analizada, de los potenciales sesgos, etc. Y NUNCA como una cifra mágica que nos seduzca tanto, que nos invite o autorice a tomar decisiones o cambiar conductas relacionadas con la práctica clínica cotidiana.

REALIDAD 3: Antes de tomar decisiones o cambiar conductas clínicas basadas en un “valor de p”, se ha de considerar también la VALIDEZ EXTERNA (generalización de los resultados obtenidos en ese estudio respecto de la población inicial y, si esas conclusiones pueden ser extrapoladas a nuestros pacientes o nuestra realidad laboral (que pueden ser distinta a la descrita en el estudio publicado y valorado por nosotros.

## Realidad 3

### ...de lo estadísticamente significativo a lo clínicamente relevante

#### Hallazgos estadísticos

1. Reflejan la influencia del azar en el resultado
2. Derivan de grupos de individuos
3. Requieren cantidades adecuadas de datos para ser válidas
4. Las respuestas estadísticas son probabilísticas
5. El análisis estadístico siempre requiere medición.

#### Hallazgos clínicos

1. Reflejan el valor biológico del resultado
2. Se practica en individuos específicos
3. Las decisiones clínicas deben tomarse a menudo con datos insuficientes
4. Los tratamientos médicos requieren decisiones oportunas comprometidas
5. La práctica clínica no puede medir fácilmente cada función, cada sistema, cada signo o síntoma clínico

## Realidad 3

...de lo estadísticamente significativo a lo clínicamente relevante

Por todo ello, es más apropiado que hablar de “SIGNIFICACIÓN ESTADÍSTICA” es más correcto utilizar el concepto de “RELEVANCIA CLÍNICA”; porque la relevancia clínica de un fenómeno va más allá de cálculos matemáticos y depende de...

1. La gravedad del problema,
2. La morbilidad y mortalidad generada por el mismo,
3. La magnitud de la diferencia,
4. La vulnerabilidad,
5. Los costes involucrados, etc.

## Realidad 3

...de lo estadísticamente significativo a lo clínicamente relevante

Realidad 3: ...de lo estadísticamente significativo a lo clínicamente relevante.	¿Es clínicamente relevante?	
¿Es estadísticamente significativo?	Sí (éxito)	No (fracaso)
Sí ( $p < 0,05$ )	Clínicamente relevante Estadísticamente significativo	Clínicamente irrelevante ( $p < 0,05$ ) Estadísticamente significativo ¿Error Alfa?
No ( $p \geq 0,1$ )	Clínicamente relevante Estadísticamente NO significativo ¿Error Beta?	Clínicamente irrelevante Estadísticamente NO significativo

## Mitos y Realidades en Bioestadística

REALIDAD 3: Un resultado estadísticamente significativo ( $p < 0,05$ ) sin relevancia clínica no deja de ser una ANÉCDOTA; un hallazgo clínicamente importante sin significación clínica ( $p > 0,1$ ) no puede ser asumido como concluyente, ya que no podemos atribuir con seguridad el hallazgo clínico a la intervención realizada (no podemos descartar la influencia del AZAR).

## Mitos y Realidades en Bioestadística

REALIDAD 3: “Una prueba estadística NO es una prueba de relevancia científica. Bajo ningún concepto la relevancia del dato se encuentra en la técnica de análisis, sino en la repercusión del mismo en el conocimiento del objeto de estudio. Es perfectamente compatible encontrar diferencias estadísticamente significativas que no suponen relevancia clínica, y viceversa”.

## Mitos y Realidades en Bioestadística

MITO 4: “La ausencia de evidencia es evidencia de ausencia”.

REALIDAD 4: “La ausencia de evidencia NO es evidencia de ausencia”.

## Mitos y Realidades en Bioestadística

MITO 4: La ausencia de significación estadística ( $p > 0,05$ ) implica que dos tratamientos comparados tienen la misma eficacia y por tanto pueden prescribirse indistintamente.

## Mitos y Realidades en Bioestadística

REALIDAD 4: Altman y Bland alertaron sobre un error común en la bibliografía biomédica: interpretar un resultado estadísticamente no significativo, cuando se comparan dos tratamientos, como sinónimo de su equivalencia terapéutica.

Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995;311:485.

Argimon JM. La ausencia de significación estadística en un ensayo clínico no significa equivalencia terapéutica. *Med Clin (Barc)* 2002; 118: 701-3.

## Mitos y Realidades en Bioestadística

REALIDAD 4: Un valor de significación estadística mayor  $p > 0,05$  se etiqueta como no significativo. Estos estudios se les denomina negativos, un término que implica erróneamente que los tratamientos son iguales, cuando lo único que sucede es que no se ha podido demostrar una diferencia. Ambos conceptos son distintos. Un resultado que no alcance significación estadística quiere decir que (si en realidad no existe esa diferencia), es probable que la observada en el estudio pueda haberse producido simplemente por el AZAR (variabilidad aleatoria). No debe interpretarse como indicativo de que no existe una diferencia en la realidad, sino tan sólo de que no puede descartarse esta posibilidad, en especial en los estudios que han incluido a pocos sujetos.

## Mitos y Realidades en Bioestadística

MITO 5: La estadística es una religión que conduce a la salvación a través del mágico ritual de la búsqueda de valores significativos de la  $p$  ( $p < 0,05$ ), alcanzando entonces los autores la verdad absoluta y el paraíso prometido de la publicación del trabajo en una revista de alto índice de impacto.



## Mitos y Realidades en Bioestadística

REALIDAD 5: En la bibliografía científica se está abusando del término significación estadística. Numerosos investigadores dan por supuesto que si en un estudio hallan una probabilidad ( $p$ ) inferior a 0,05, ello certifica la existencia tanto de la validez científica de su observación como de la relevancia clínica de la misma. Y ello está lejos de ser cierto. Su fe ciega en el valor de la  $p$  contrasta con los detractores de su significado (ateos de la estadística).

## Mitos y Realidades en Bioestadística

REALIDAD 5: La estadística es una herramienta útil para el diseño y realización de estudios sanitarios siempre que sea utilizada adecuadamente.

Salsburg DS. The religion of statistics as practiced in medical journals. Am Statistician 1985; 39: 220-223.

## Mitos y Realidades en Bioestadística

REALIDAD 5: Los métodos estadísticos NO son un sustituto del sentido común y la objetividad. Nunca deberían estar dirigidos a impresionar o confundir al lector, sino que deben ser una contribución importante a la claridad de los conocimientos científicos.

## Mitos y Realidades en Bioestadística

REALIDAD 5: Es fundamental hacer buenas preguntas de investigación, lo cual nos llevará a trabajos interesantes y respuestas con sentido común, no habrá necesidad de “SALIR DE PESCA” para presentar al editor muchas  $p$  menores de 0.05. Todo, con mesura y utilizado racionalmente, es bueno, y la estadística es un instrumento indispensable pero, como todo en la vida, con SENTIDO COMÚN (y este sentido estadísticamente hablando no parece ser a veces muy frecuente).

Dr. Juan M. García Torrecillas  
Médico de Especialista en Medicina Familiar y Comunitaria  
Experto en Epidemiología y Nuevas Tecnologías



## El Borracho

“Utiliza la estadística como un borracho usa una farola, para aguantarse en lugar de para iluminarse”

## Objetivo principal de la estadística:

¿Cuál es la probabilidad de que la diferencia observada en el estudio se deba al AZAR?

- Una  $p < 0,05$  indica que en menos de 5 veces de cada 100 que repitiéramos el mismo estudio, nuestro resultado se debería al AZAR.
- Cuanto mayor sea la  $p$  ( $p > 0,1$ ) más fuerte y segura será la evidencia a favor de la hipótesis nula (igualdad).
- El valor de la  $p$  no es un indicador de fuerza de asociación, ni tampoco de su importancia clínica.

## La Llave Dorada

“La estadística no es la llave dorada que abre la puerta de la verdad absoluta e infinita”

La verdad absoluta e infinita tiene una sonrisa diabólica. Si consigo la llave dorada podré abrir la puerta , y así podré conocer la verdad absoluta e infinita.

Dr. Jacobo Díaz Portillo  
Coordinador de Formación Continuada  
Hospital Universitario de Ceuta

## Mitos y Realidades en Bioestadística

MITO 6 (*significantitis*): Si  $p$  es mayor que el 5%, el resultado del estudio no vale para nada; sólo si es menor de 0,05, hemos hecho un gran descubrimiento.



Chia KS. "Significant-itis": an obsession with the P value. Scand J Work Environ Health. 1997; 23:152-4.  
Christenson P. To p or not to p. J Child Adolesc Psychiatr Nurs. 1995; 8:42.

## Mitos y Realidades en Bioestadística

REALIDAD 6: Algunos, de forma irónica, a esta postura tan drástica y que pocas veces está realmente justificada la consideran una auténtica enfermedad llamada significantitis o es como un problema de conciencia: “to p or not to p”. Neyman et al. al considerar que había que establecer un límite claro en el valor  $p$  para aplicar los razonamientos anteriores de forma absoluta. Este valor crítico de  $p$  es el famoso 0,05 y marca el mítico límite de la significación estadística.

Neyman J, Pearson E. On the problem of the most efficient tests of statistical hypothesis. Philosophical trans of the Royal Society of London (A) 1933;231:289-337.

## Mitos y Realidades en Bioestadística

REALIDAD 6: Neyman et al lanzaron este lema exclusivamente bajo el punto de vista de la toma de decisión ante un problema que fuera trascendente y en el que, urgiera tal toma de decisión: rechazar o no la hipótesis nula con sus consecuencias derivadas. Lo que ocurre es que se ha universalizado demasiado y se lo ha sacado de su contexto, ya que se aplica a cualquier inferencia estadística, aunque el fin de la investigación no suponga obligación alguna de tomar una decisión tan drástica, basada en la ley del todo o nada.

Neyman J, Pearson E. On the problem of the most efficient tests of statistical hypothesis. Philosophical trans of the Royal Society of London (A) 1933;231:289-337.

## Mitos y Realidades en Bioestadística

MITO 7 (adoradores de la p): La esencia de la estadística es la búsqueda de los valores más bajos de la p, (una  $p < 0,001$  es mejor que una  $p < 0,01$ , y ésta superior a una  $p < 0,05$ ) ya que cuanto más pequeña sea esta p mayor intensidad de asociación y causalidad existirá entre las variables y mayor credibilidad tendrán las conclusiones de mi estudio.



## Mito 7: “Adoradores de la p”

“El cual los tomó de las manos, y formólo con buril, e hizo de ello un becerro de fundición. Entonces dijeron: Israel, estos son tus dioses, que te sacaron de la tierra de Egipto”

## Mitos y Realidades en Bioestadística

REALIDAD 7: La utilidad de la  $p$  es una medida de la influencia del AZAR en la posible asociación entre las variables analizadas, este azar debe ser siempre contemplado, evaluado y cuantificado mediante el valor de la  $p$ . La “ $p$ ” NO es un indicador de fuerza de la asociación entre variables, ni de su importancia. Los valores de la “ $p$ ” deben ser considerados solo como una guía y no como base de conclusiones definitivas e irrevocables.

## Mitos y Realidades en Bioestadística

MITO 8: Hace poco leí en la prensa médica la siguiente frase: “*Yo no creo en la estadística*”, (ATEOS de la estadística) porque está claro que la estadística, como religión que es, es solo una cuestión de fe, que como tal debe ser respetada (según dice el artículo 16 de la Constitución Española sobre la libertad religiosa).



## Mitos y Realidades en Bioestadística

REALIDAD 8: La estadística constituye un instrumento científico que requiere el cumplimiento de NUMEROSAS CONDICIONES para producir resultados válidos.

1. Es preciso que los experimentos tengan un diseño muy escrupuloso.
2. Es inexcusable aplicar en cada circunstancia aquellos métodos que en ella son idóneos y controlar rigurosamente su validez.
3. NUNCA es legítimo deducir que un resultado significativo es clínicamente relevante.
4. Una correlación entre dos variables acabará siendo siempre significativa ( $p < 0,05$ ) si la muestra es lo suficientemente grande.
5. Por eso en algunos casos, se dice que: “La estadística es un instrumento, que si no se usa adecuadamente, puede demostrar cualquier MENTIRA”.

## Mitos y Realidades en Bioestadística

REALIDAD 8: Aplicando los métodos estadísticos de forma abusiva (como lo hacía la Santa Inquisición), si torturamos suficientemente a nuestros datos, acabaran por confesarnos lo que nosotros queramos. El fin nunca justifica los medios.

“La estadística es un instrumento, que si no se usa adecuadamente, puede demostrar cualquier mentira”.

En Bioestadística el fin nunca justifica los medios.

## Mitos y Realidades en Bioestadística

REALIDAD 8: Si los datos de un estudio se analizan con muchos enfoques distintos y con la suficiente intensidad hasta TORTURARLOS, acaban dando los resultados que al investigador le hubiese gustado obtener, aunque tales resultados suelen ser científicamente NEFASTOS.

## Mitos y Realidades en Bioestadística

MITO 9: “La inocencia de las cigüeñas”

“Una buena correlación estadística (coeficiente de correlación  $r$  elevado) ( $p < 0,05$ ) siempre significa que existe una relación entre las variables evaluadas y, por tanto, existe una relación causa - efecto”.

Rev Lab Clin. 2010;3(2):51



ELSEVIER  
DOYMA

# Revista del Laboratorio Clínico

[www.elsevier.es/LabClin](http://www.elsevier.es/LabClin)



EDITORIAL

## La inocencia de las cigüeñas

The innocence of the storks



## Bibliografía

1. Arís A. Tómese una antes de acostarse, 1ª ed. Barcelona: Planeta; 1998.
2. Escrig Sos J. Errores de grueso calibre en la aplicación de pruebas estadísticas. Cir Esp. 2010;87:127–9.

Felip Antoja Ribó  
*Director*

## Mitos y Realidades en Bioestadística

### REALIDAD 9:

- Una buena correlación estadística ( $r$  elevado) ( $p < 0,05$ ) NO siempre significa que exista una relación entre las variables evaluadas y, por tanto, NO implica una relación causa-efecto.
- El trabajo estaría mal planteado, aunque se hubiera aplicado un tratamiento estadístico aparentemente correcto.
- Los autores de un estudio deben ser muy cautos en la aplicación y posterior interpretación del análisis estadístico.

## Estudio ecológico

Variable PREDICTORA  
Factor de exposición (F)

CIGÜEÑAS

Número de ejemplares de cigüeñas

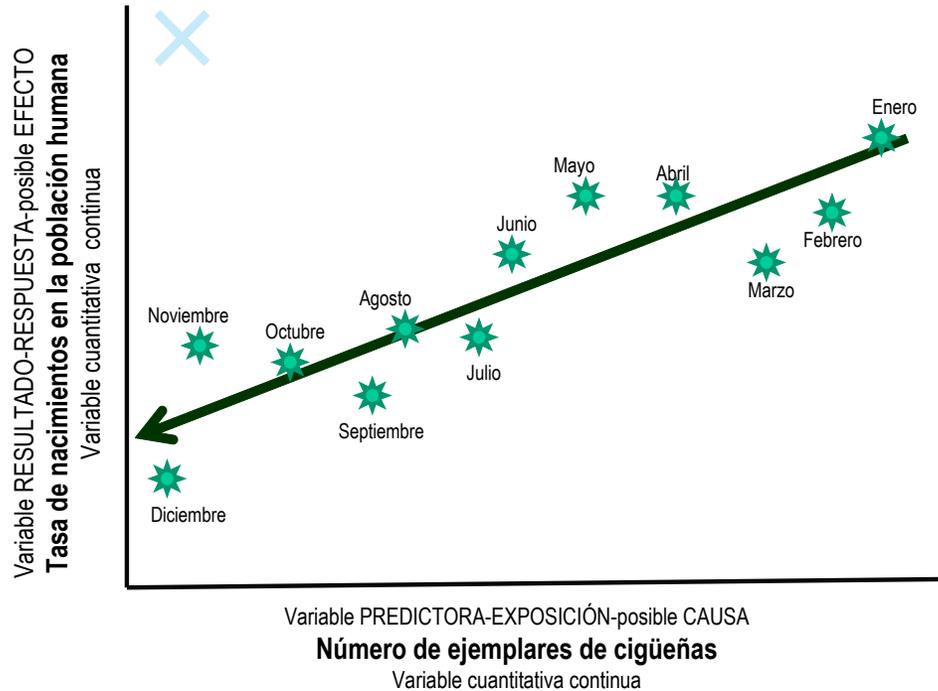


Variable RESULTADO  
Efecto o Enfermedad (E)

NATALIDAD HUMANA

Tasa de natalidad de la población humana

## Estudio de Bartels (Alemania)



## Estudio de Bartels (Alemania)

**RESULTADOS:** En Alemania, en un mismo año, descendieron, de forma significativa, la tasa de nacimientos de la población humana y el número de ejemplares de cigüeñas. La correlación entre ambas variables era alta, lo que demostraba científicamente su relación.



**CONCLUSIÓN:** La baja natalidad alemana en el año de estudio se explica por la disminución del número de cigüeñas repartidoras de bebés. La culpa era de las cigüeñas.

## Comentarios

- Es muy saludable encontrar detalles de humor.
- Si leyéramos un supuesto artículo científico con esa conclusión tan extravagante pensaríamos, sin duda, que se trata de una parodia y como tal la aceptaríamos.
- Es fácil entender el *error* porque sabemos que las cigüeñas no tienen nada que ver con la reproducción humana y, por tanto, se ha buscado erróneamente una relación entre dos variables sin plausibilidad biológica.
- El trabajo está mal planteado, y aunque se hubiera aplicado un tratamiento estadístico aparentemente correcto.
- Una buena correlación estadística ( $r$  elevado) ( $p < 0,05$ ) NO siempre significa que existe una relación entre las variables evaluadas, y por tanto NO implica una relación causa efecto.

## Reflexiones

¿Qué ocurre cuando el estudio se lleva a cabo con variables (datos) menos conocidos, de poblaciones que nos son menos familiares, obtenidos en estudios complejos?

¿Cómo sabemos si el autor sabe certeramente entre qué grupos se puede establecer relaciones y entre cuáles no tiene sentido hacerlas?

Esto sólo es posible si el investigador conoce muy bien lo que está estudiando y ha partido de un planteamiento y el diseño del estudio bien estructurado.

Las conclusiones que se obtienen al aplicar las pruebas estadísticas sólo tienen sentido cuando las variables (datos) están bien definidas y tienen plausibilidad biológica.

Estos tipos de errores se originan cuando los autores NO tienen claros algunos conceptos básicos de METODOLOGÍA de INVESTIGACIÓN pero disponen de programas estadísticos potentes, en los que sólo se necesita señalar y aceptar con el ratón entre un amplio abanico de pruebas estadísticas.

Parfraseando a Séneca , debemos convenir que...

“todo lo honesto es difícil”

... y la investigación clínica no es una excepción.

# Curso de Bioestadística Aplicada

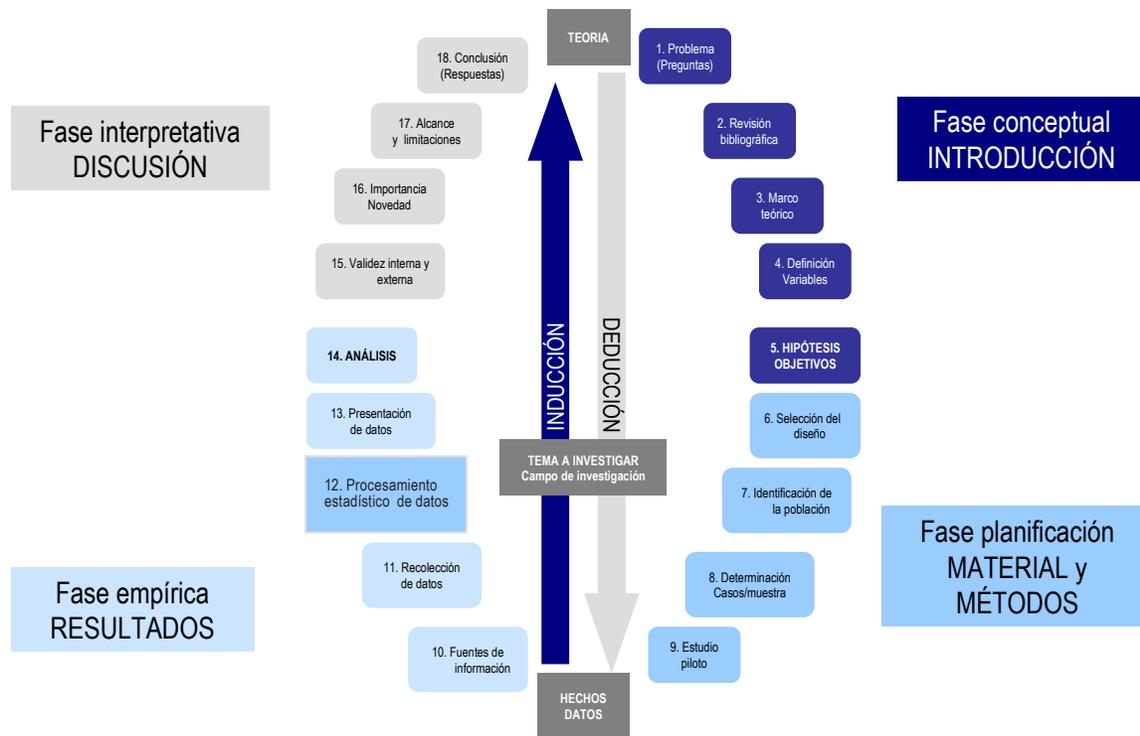
1. Introducción a la bioestadística.
2. Estadística descriptiva. Variables y gráficos.
3. Tamaño muestral: Error alfa y error beta.
4. Evaluación de pruebas diagnósticas.
5. Test de contraste de hipótesis: variable cualitativa versus cualitativa.
6. Test de contraste de hipótesis: variable cuantitativa versus cuantitativa.
7. Test de contraste de hipótesis: variable cuantitativa versus cualitativa.
8. Análisis estadístico multivariante.
9. Utilidad de la bioestadística en los resultados y discusión del estudio.

Salir de nuestra cueva de BENZÚ, de nuestro nicho ecológico...

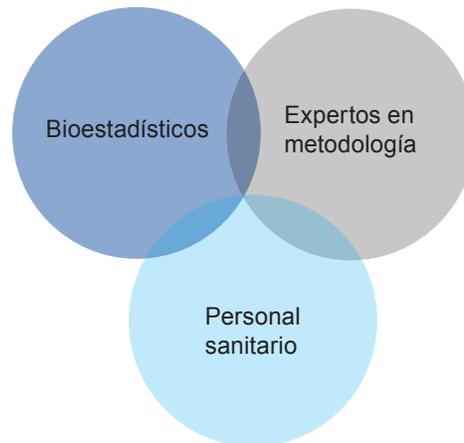
Comunicar nuestra experiencia a los demás...

Utilizando la bioestadística como herramienta de trabajo.

# Etapas del proceso de investigación



La bioestadística y el método epidemiológico aportan más ciencia al arte de la medicina y enfermería y es por ello que la colaboración conjunta de expertos en metodología, bioestadísticos y los clínicos deben trabajar en equipo en colaboración para en un esfuerzo conjunto reducir la variabilidad que rodea al ejercicio de la medicina y enfermería que será siempre una ciencia de probabilidades y un arte de manejar la incertidumbre.



Pita Fernández S. Dificultades de los médicos para la realización de estudios clínico-epidemiológicos. Congreso Galego de Estadística e Investigación de Operacións. Santiago de Compostela, A Coruña (España). 11-13 Novembre-1999.

## Capítulo 2

# Publicación Bioestadística: Variables y gráficos.

**Dr. Jacobo Díaz Portillo**  
Coordinador de Formación Continuada  
Hospital Universitario de Ceuta

# Materiales

## Pacientes

1. Criterios de selección de la muestra (inclusión y exclusión).
2. Procedencia.
3. Técnica de reclutamiento utilizada.
4. Tiempo empleado para reunir a todos los participantes.
5. *Principales características de los pacientes.*
  - *Edad*
  - *Sexo*
  - *Procedencia*
  - *Gravedad*
  - *Años de evolución de la enfermedad*
  - *Patologías asociadas...*
6. Cálculo de tamaño muestral.

## Tipos de variables

- Variables cualitativas
  - Categóricas nominales
    - Dicotómicas o Binarias
    - Policotómicas
  - Categóricas ordinales
- Variables cuantitativas
  - Discretas
  - Continuas



Categorización

## Tipos de variables

- Variables CUALITATIVAS: Representan una cualidad o atributo que clasifica a cada individuo en una de varias categorías. Pueden ser de dos tipos:
  - Dicotómicas o binarias: La situación más sencilla es aquella en la que se clasifica cada individuo en uno de dos grupos (hombre/mujer, enfermo/sano, fumador/no fumador).
  - Policotómicas: Es necesario un mayor número de categorías (color de los ojos, grupo sanguíneo, profesión, etcétera).
  
- Variables CUANTITATIVAS: Son aquellas que pueden medirse, cuantificarse o expresarse numéricamente. Pueden ser de dos tipos:
  - Continuas: Son aquellas que si admiten tomar cualquier valor dentro de un rango numérico determinado, con uno o varios decimales (peso, talla, edad).
  - Discretas: Son aquellas que NO admiten todos los valores intermedios en un rango. Suelen tomar solamente valores enteros (número de hijos, abortos, partos, hermanos).

## Tipos de variables según la escala de medida

Escalas nominales: Variables NOMINALES: ésta es una forma de observar o medir en la que los datos se ajustan por categorías que no mantienen una relación de orden entre sí.

- Sexo.
- Grupo sanguíneo.
- Presencia o ausencia de una enfermedad.
- Presencia o ausencia de un factor de riesgo.

Escalas ordinales: Variables ORDINALES: En las escalas utilizadas para medirlas existe un cierto orden, grado o jerarquía entre las categorías.

- Grado de disnea.
- Grado de dolor.
- Intensidad del hábito tabáquico.
- Tipo de fumador.

## VARIABLES CUALITATIVAS CATEGÓRICAS ORDINALES

- Grado de disnea.
- Grado de dolor.
- Intensidad del hábito tabáquico.
  - 1-9 cigarrillos/día.
  - 10-20 cigarrillos/día.
  - > 20 cigarrillos/día.
- Tipo de fumador.
  - Exfumador.
  - Fumador pasivo.
  - Fumador activo.

## Categorización de variables cuantitativas

Variable

Cuantitativa continua

Triglicéridos



Variable

Cualitativa dicotómica

Hipertriglicéridemia  
Normotriglicéridemia

Categorización

Variable

Cuantitativa discreta

Test de APGAR



Variable

Cualitativa ordinal

Patológico < 7  
Dudoso = 7  
Fisiológico > 7

Categorización

# Tipos de variables

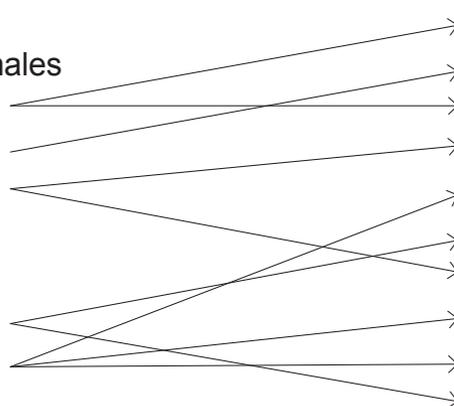
## Ejemplos

### Variables cualitativas

- Categóricas nominales
  - Binarias
  - Policotómicas
- Ordinales

### Variables cuantitativas

- Discretas
- Continuas



- Sexo
- Credo religioso en Ceuta
- Hipertensión
- Tipo de fumador
- Colesterolemia
- Número de abortos
- Nivel de instrucción
- Edad gestacional
- Frecuencia cardiaca
- Test de apgar

# Tipos de variables

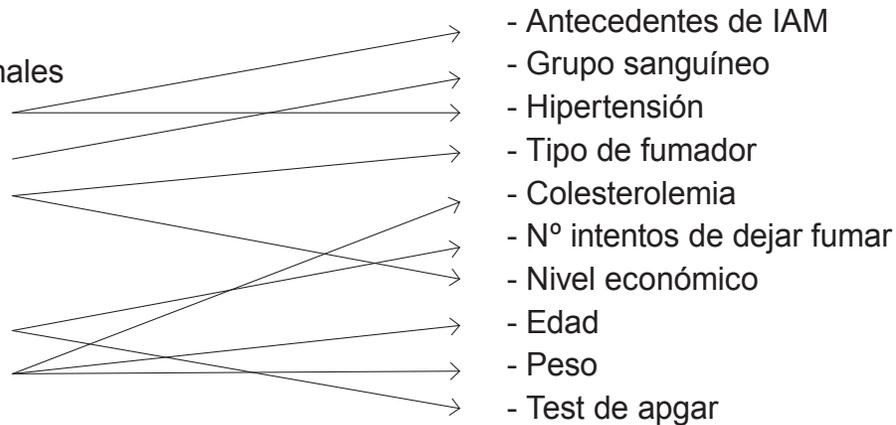
## Ejemplos

### Variables cualitativas

- Categóricas nominales
  - Binarias
  - Policotómicas
- Ordinales

### Variables cuantitativas

- Discretas
- Continuas



# Tipos de variables

## Ejemplos

### Variables cualitativas

- Categóricas nominales

- Binarias

- Policotómicas

- Ordinales

### Variables cuantitativas

- Discretas

- Continuas

- Desenlace
- Estado civil
- Inmunización toxoplasma
- Intensidad del hábito tabáquico
- LDL colesterol
- Número de cigarrillos
- Grado de disnea
- Años de evolución
- Hemoglobina
- Número de hijos

## Tipos de variables

Variable independiente (X)	Variable dependiente (Y)
Controlada por el investigador	Respuesta que no se controla
Información disponible	Información que se desea conseguir
Variable predictora	Variable desenlace
Variable previa	Variable criterio
Exposición (F)	Resultado (E)
Posible causa	Posible efecto
Factor	Consecuencia
Regreso	Regresando

## Pruebas estadísticas de contraste de hipótesis más aplicables a cada situación

Variable Independiente o PREDICTORA	Variable Dependiente o RESULTADO	Pruebas de contraste de hipótesis	Observaciones y/o limitaciones del test
Cualitativa	Cualitativa	Ji cuadrado Prueba exacta de Fisher Test de McNemar Q de Cochran REGRESIÓN LOGÍSTICA	Si el tamaño muestral es grande Si el tamaño muestral es pequeño Dos grupos de muestras repetidas o apareadas Más de dos grupos de muestras repetidas o apareadas Varias variables predictoras cualitativas
Cualitativa	Cuantitativa	t de Student Análisis de la Varianza Test de Mann-Whitney Test de Wilcoxon Test de Kruskal-Wallis Test de Friedman	Compara medias entre dos grupos independientes Compara medias entre más de dos grupos independientes Dos grupos independientes No paramétricos Dos grupos medidas repetidas No paramétrico Más de dos grupos No paramétrico, grupos independientes Más de dos grupos No paramétrico medidas repetidas
Cuantitativa	Cualitativa	Regresión logística simple Regresión logística multivariante	Una sola variable predictora cuantitativa Varias variables predictoras cuantitativa o cualitativas
Cuantitativa	Cuantitativa	Regresión lineal simple Correlación de Pearson Correlación de Spearman REGRESIÓN MÚLTIPLE	Paramétrico Paramétrico No paramétrico Varias variables predictoras cuantitativas
Cualitativa	Supervivencia	Kaplan-Meier Long-Rank Regresión de COX	Curvas de supervivencia Compara curvas de supervivencia Análisis Multivariante

## Fases del análisis estadístico

1.- ANÁLISIS DESCRIPTIVO



2.-ANÁLISIS BIVARIANTE



3.- ANÁLISIS MULTIVARIANTE

# Estadísticos descriptivos

## Medidas de tendencia central

- Media aritmética
- Media geométrica
- Media armónica
- Media ponderada
- Mediana
- Moda

## Medidas de forma

- Asimetría
- Curtosis o apuntamiento

## Medidas de dispersión

- Varianza
- Desviación estándar
- Coeficiente de variación
- Error estándar de la media

## Medidas de posición

- Cuantiles
- Percentiles

## Medidas de tendencia central

- Media aritmética [x]
- Media geométrica
- Media armónica
- Media ponderada
- Mediana (P50)
- Moda

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

$$\text{Media geométrica} = \sqrt[n]{x_1 x_2 \dots x_n} = \sqrt[n]{\prod_1^n x_i}$$

$$\text{Media armónica} = \frac{n}{\sum \left( \frac{1}{x_i} \right)}$$

$$\text{Media ponderada} = \frac{\sum (w_i x_i)}{\sum (w_i)} = \frac{w_1 x_1 + w_2 x_2 + w_3 x_3}{w_1 + w_2 + w_3}$$

## Media aritmética

- Principal medida de tendencia central, pero menos robusta que la media geométrica o media armónica.
- Se deja influir mucho por los valores extremos de la distribución.
- Se suman todos los valores y se divide por el número total de individuos.
- Expresión con pocos decimales.

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

Ejemplo: ¿Cuál es la nota de un alumno del curso de metodología de investigación que ha obtenido las siguientes calificaciones?

xi	valor	Peso	valor
Asistencia a clase (X1)	7,5	W1	20%
Tesina (prácticas) (X2)	5,5	W2	20%
Examen inicial (X3)	3/14 (2,1)	W3	10%
Examen final (X4)	11/14 (7,8)	W4	20%
Incremento en sus conocimientos (X4 - X3)	8	W5	30%

$$\text{Media ponderada} = \frac{\sum(w_i x_i)}{\sum(w_i)} = \frac{w_1 x_1 + w_2 x_2 + w_3 x_3}{w_1 + w_2 + w_3}$$

Media ponderada = 6,7

## Mediana $P_{50}$

- Medida de tendencia central robusta, de cálculo fácil.
- No se deja influir mucho por los valores extremos de la distribución.
- Coincide con el valor central del conjunto ordenado de observaciones que deja por debajo y por encima a la mitad de las observaciones realizadas.
- Se debe utilizar siempre en dos casos:
  - Bajo tamaño muestral (para muestras pequeñas) y/o
  - Cuando existen observaciones extremas (periféricos)

## Medidas de dispersión

- Varianza [s<sup>2</sup>]

$$\text{Varianza muestral } (s^2) = \frac{\sum(x_i - \bar{x})^2}{n - 1}$$

- Desviación típica o estándar [s]

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n - 1}}$$

Muestral

- Coeficiente de variación [CV]

$$\text{Coef. de variación} = \frac{\text{Desviación típica}}{\text{media}} = \frac{s}{\bar{x}} \times 100$$

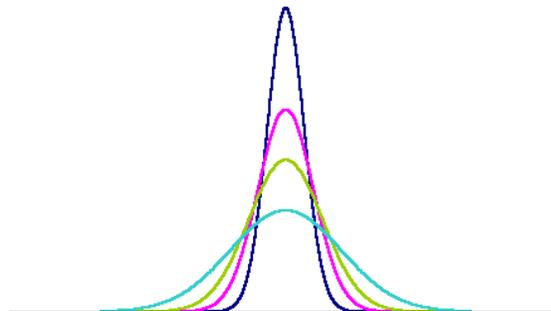
- Error estándar de la media [EE]

$$EE = \frac{s}{\sqrt{n}}$$

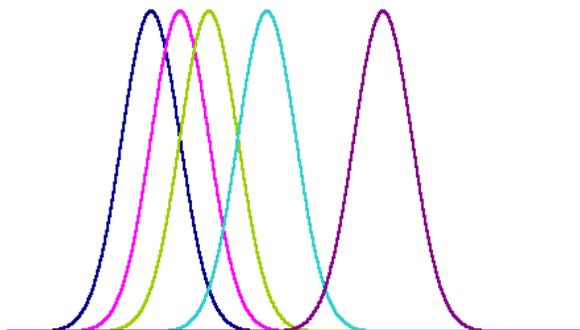
¿Cuál de las siguientes distribuciones presenta un mayor valor de la media?

¿Cuál de las siguientes distribuciones presenta una mayor dispersión?

(a) Distribuciones normales con distinta desviación estándar e igual media



(b) Distribuciones normales con diferentes medias e igual desviación estándar



¿Cuál de las siguientes distribuciones presenta una mayor dispersión?

¿Cuál de las siguientes distribuciones presenta un mayor valor de la media?

## Desviación estándar (s)

- Medida de dispersión que nos suministra mucha información sobre la naturaleza de una variable.
- Cuanto más separados están los valores de la media mayor será su dispersión, y mayor será su desviación estándar.
- Se debe representar como media (desviación estándar), y NO como media  $\pm$  desviación estándar.
- Su magnitud SI depende de las unidades de medida de la variable. NO permite la comparación directa de la dispersión de varias variables.

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n - 1}}$$

Muestral

## Coeficiente de variación (CV)

- Medida de dispersión que relaciona que proporción porcentual de la media representa la desviación estándar.
- Representa la desviación estándar medida en unidades de magnitud relativas a la media.
- Permite la comparación directa de la dispersión de varias variables.
- Su magnitud NO depende de las unidades de medida de la variable.

$$\text{Coef. de variación} = \frac{\text{Desviación típica}}{\text{media}} = \frac{s}{\bar{x}} \times 100$$

¿Cuál de las siguientes variables presenta un mayor grado de dispersión?

Repetibilidad interdiaria de los procedimientos de análisis del metabolismo férrico				
Componente	N	Media	Desviación estándar	Coefficiente de variación (CV %)
Hemoglobina	20	121,70	0,64	0,53
Hierro	20	0,99	0,01	1,20
Ferritina	20	94,50	4,24	4,49
Transferrina	20	3,25	0,094	2,89
Receptor soluble de la transferrina	20	2,39	0,13	5,44

¿Cuál de las siguientes variables presenta un menor grado de dispersión?

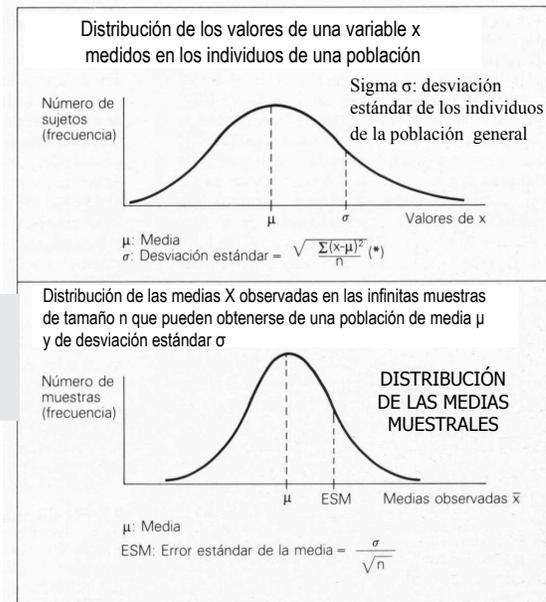
$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n - 1}}$$

**Muestral**

Desviación estándar: Mide el grado de dispersión de todos los individuos que forman la muestra (s) o la población ( $\sigma$ ).

Media poblacional  $\mu$  desconocida de la que proceden todas las muestras extraídas

$$EE = \frac{s}{\sqrt{n}}$$



Error Estándar de la Media EEM: Mide el grado de dispersión de las medias de todas las posibles muestras que pudieran extraerse de la población. Mide el grado de dispersión de las medias muestrales respecto a la media poblacional  $\mu$  (desconocida).

## Medidas de forma

- Asimetría (coeficiente de asimetría  $\gamma_1$  de Fisher)
  - Simétrica (coeficiente de asimetría  $\gamma_1 = 0$ )
  - Asimétrica (coeficiente de asimetría  $\gamma_1 \neq 0$ )
    - Asimetría positiva (coeficiente de asimetría  $\gamma_1 > 0$ )
    - Asimetría negativa (coeficiente de asimetría  $\gamma_1 < 0$ )
- Curtosis o apuntamiento (coeficiente  $\gamma_2$  curtosis de Fisher)
  - Mesocórtica ( $\gamma_2 = 0$ )
  - Platicúrtica ( $\gamma_2 < 0$ )
  - Leptocúrtica ( $\gamma_2 > 0$ )

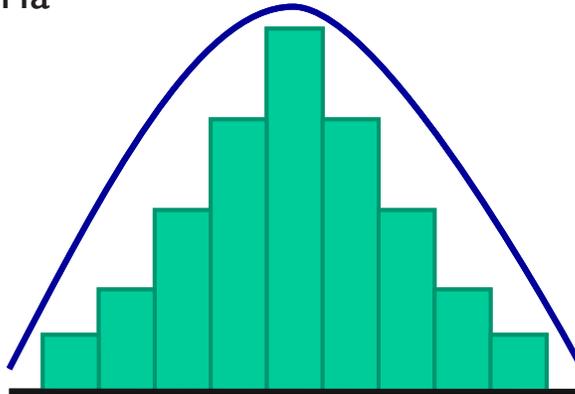
$$\gamma_1 = \frac{\sum_{i=1}^n \varepsilon_i^3}{\sigma_R^3}$$

$$\gamma_2 = \frac{\sum_{i=1}^n \varepsilon_i^4}{\sigma_R^4} - 3$$

Coeficiente  $\gamma_2$  mide el grado de APLASTAMIENTO de una curva

# Medidas de forma

## Asimetría



Histograma de frecuencias de una variable cuantitativa con distribución simétrica (coeficiente de asimetría  $\gamma_1 = 0$ )

$$\gamma_1 = \frac{\sum_{i=1}^n \varepsilon_i^3}{\sigma_R^3}$$

Distribución normal

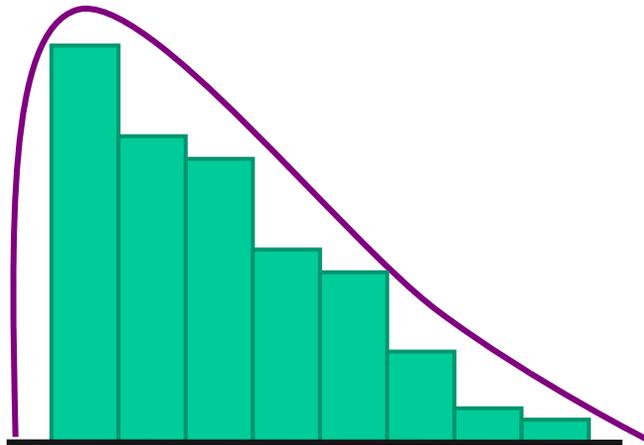


[Asimetría] < 2 EE<sub>asimetría</sub>

$$\gamma_1 < 2 EE \gamma_1$$

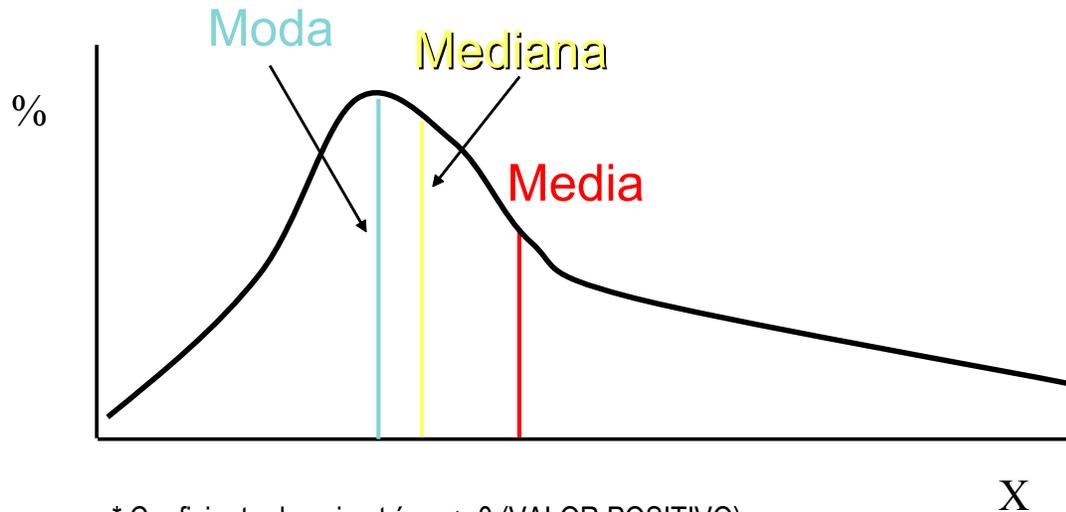
# Medidas de forma

## Asimetría



Histograma de frecuencias de una variable cuantitativa  
con distribución Asimétrica Positiva (coeficiente de asimetría  $\gamma_1 > 0$ )

Histograma de frecuencias de una variable cuantitativa con distribución Asimétrica positiva (coeficiente de asimetría  $\gamma_1 > 0$ )



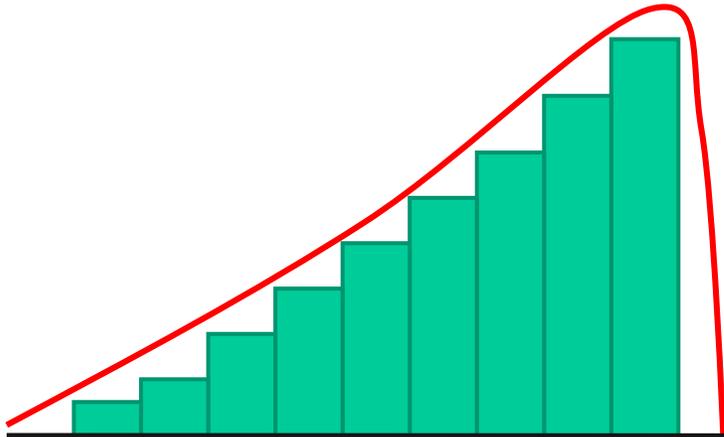
\* Coeficiente de asimetría  $\gamma_1 > 0$  (VALOR POSITIVO)

\* La Mediana es un valor más pequeño que la Media

\* La cola de la distribución se aleja por la derecha

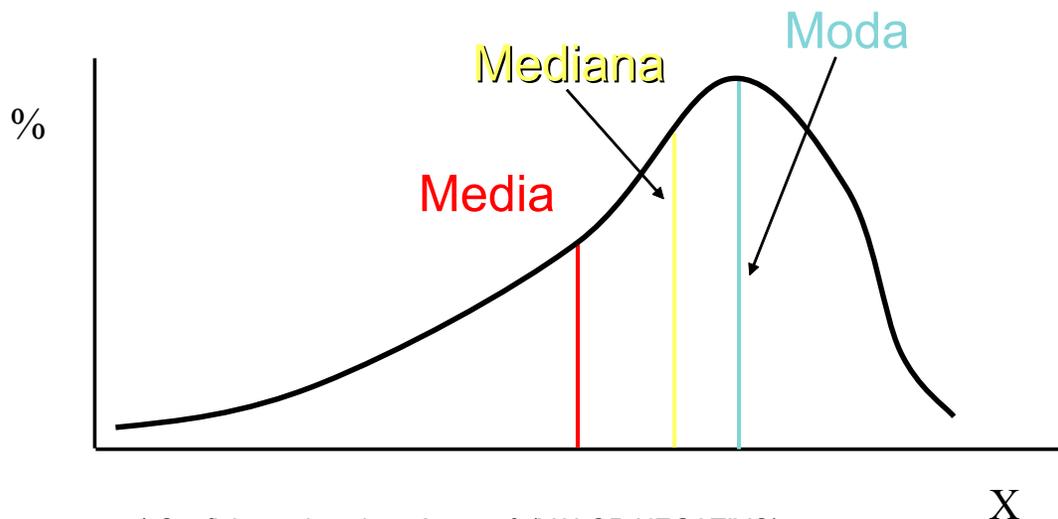
# Medidas de forma

## Asimetría



Histograma de frecuencias de una variable cuantitativa con distribución Asimétrica Negativa (coeficiente de asimetría  $\gamma_1 < 0$ )

Histograma de frecuencias de una variable cuantitativa con distribución Asimétrica negativa (coeficiente de asimetría  $\gamma_1 < 0$ )



\* Coeficiente de asimetría  $\gamma_1 < 0$  (VALOR NEGATIVO)

\* La Mediana es un valor más grande que la Media

\* La cola de la distribución se aleja por la izquierda

# Medidas de forma

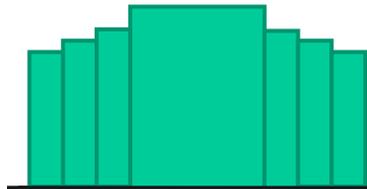
## Curtosis o apuntamiento

Coefficientes  $\gamma_2$  de apuntamiento o curtosis de Fisher

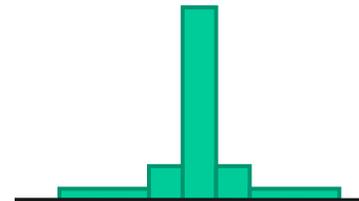
Mesocúrtica ( $\gamma_2 = 0$ )



Platicúrtica ( $\gamma_2 < 0$ )



Leptocúrtica ( $\gamma_2 > 0$ )



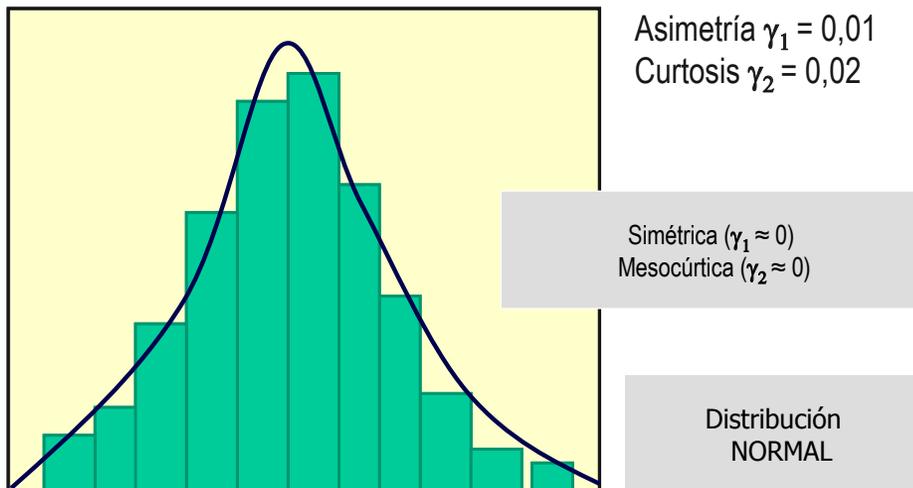
Distribución normal



[curtosis]  $< 2$  EE<sub>curtois</sub>

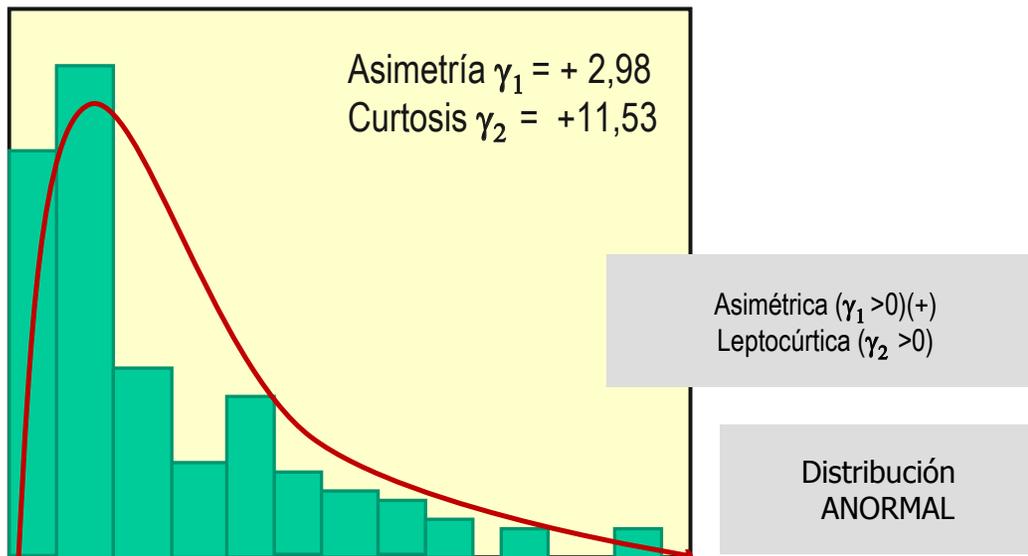
$\gamma_2 < 2$  EE  $\gamma_2$

Según las medidas de forma ¿Cómo puede clasificarse esta variable?



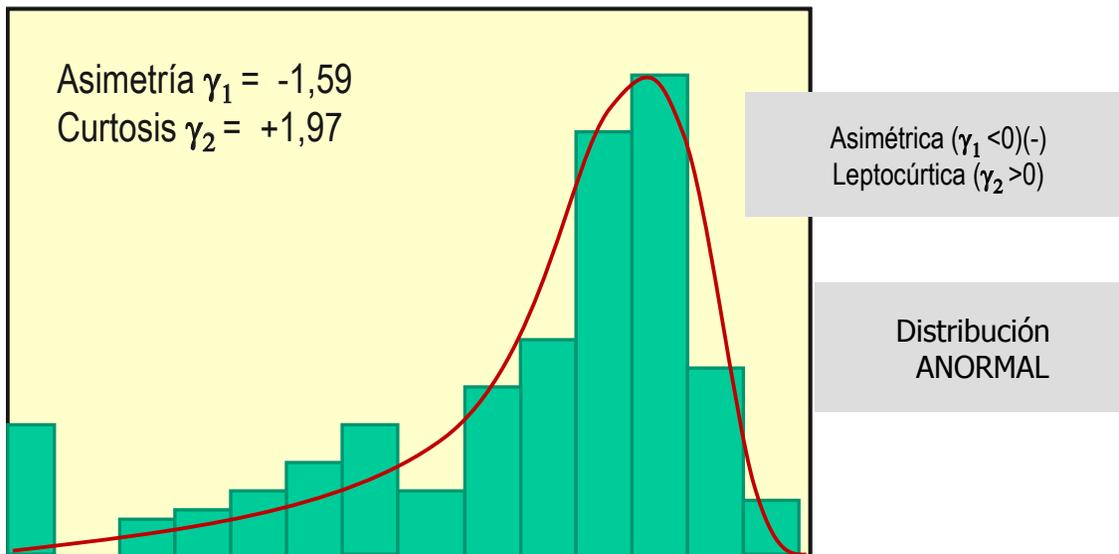
Coeficientes de asimetría  $\gamma_1$  y de apuntamiento o curtosis  $\gamma_2$   
en diversos tipos de distribuciones

Según las medidas de forma ¿Cómo puede clasificarse esta variable?



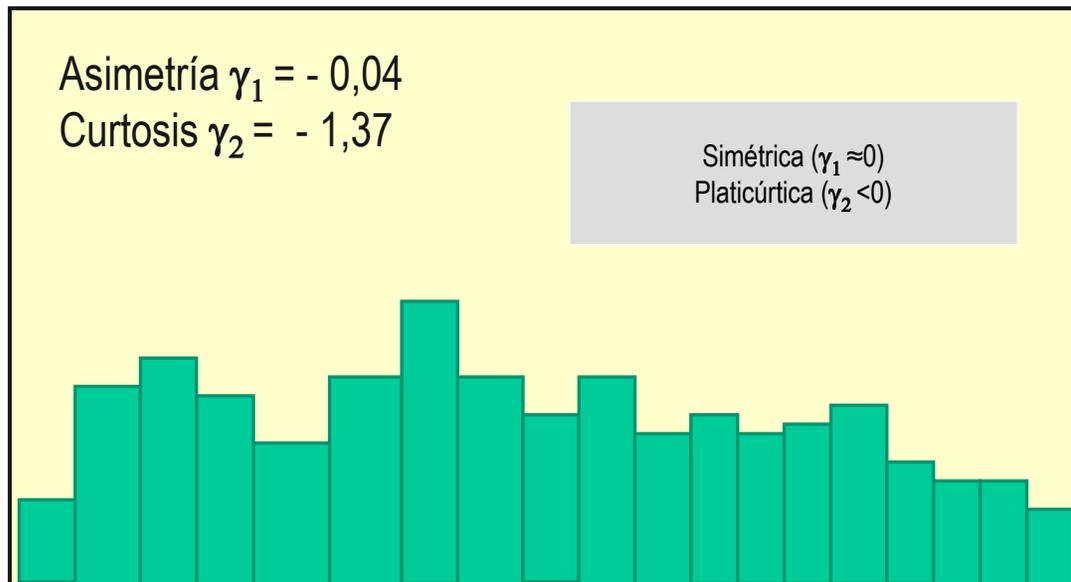
Coefficientes de asimetría  $\gamma_1$  y de apuntamiento o curtosis  $\gamma_2$   
en diversos tipos de distribuciones

Según las medidas de forma ¿Cómo puede clasificarse esta variable?



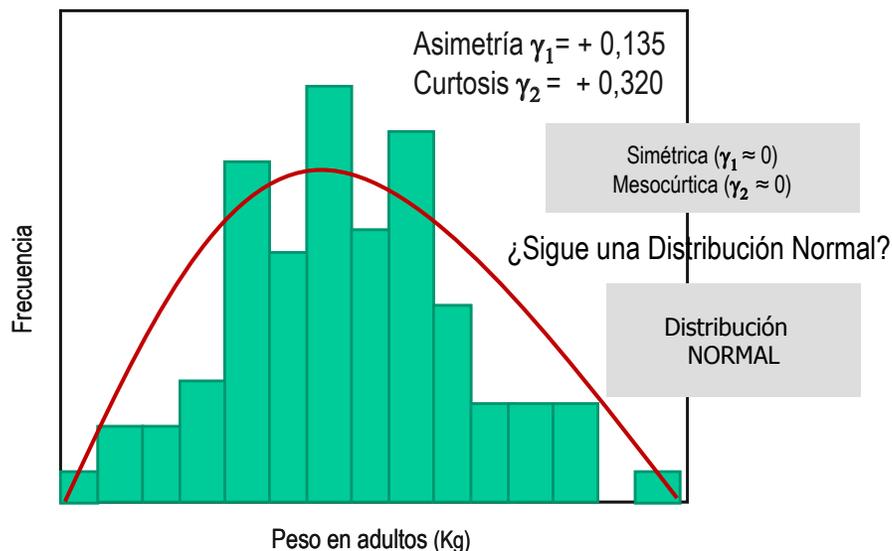
Coeficientes de asimetría  $\gamma_1$  y de apuntamiento o curtosis  $\gamma_2$   
en diversos tipos de distribuciones

Según las medidas de forma ¿Cómo puede clasificarse esta variable?



Coefficientes de asimetría  $\gamma_1$  y de apuntamiento o curtosis  $\gamma_2$   
en diversos tipos de distribuciones

Según las medidas de forma ¿Como puede clasificarse esta variable?

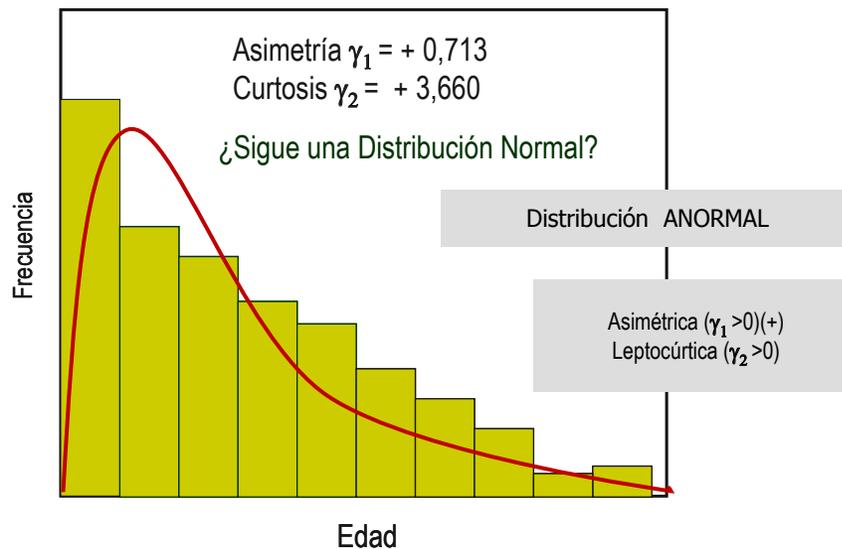


Test de Kolmogorov - Smirnov = 0,705



p = 0,702

Según las medidas de forma ¿Cómo puede clasificarse esta variable?

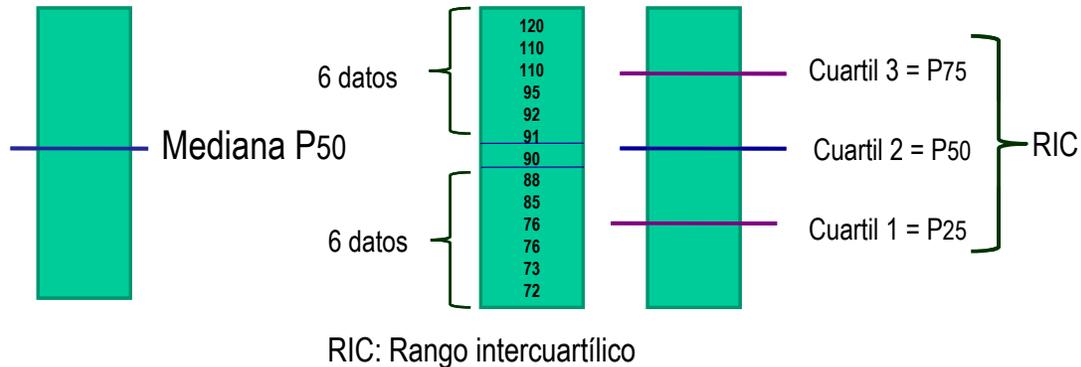


Test de Kolmogorov - Smirnov = 1,498



p = 0,022

# Medidas de posición: Cuartiles y percentiles



Representación esquemática de la Mediana P50  
(percentil 50) y de los cuartiles P75 y P25

## Gráficos de caja (box-plot)

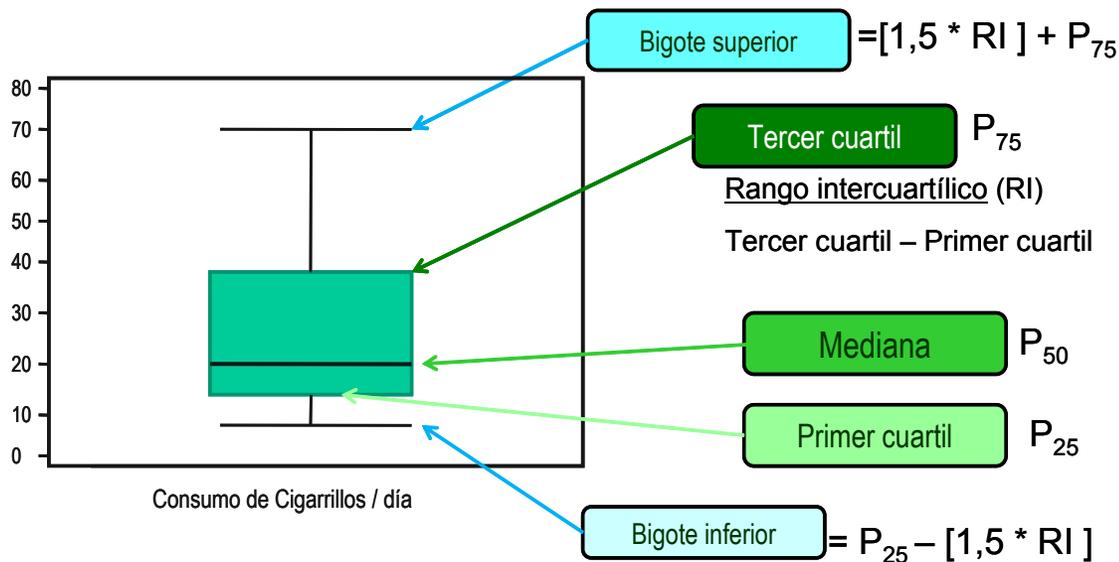


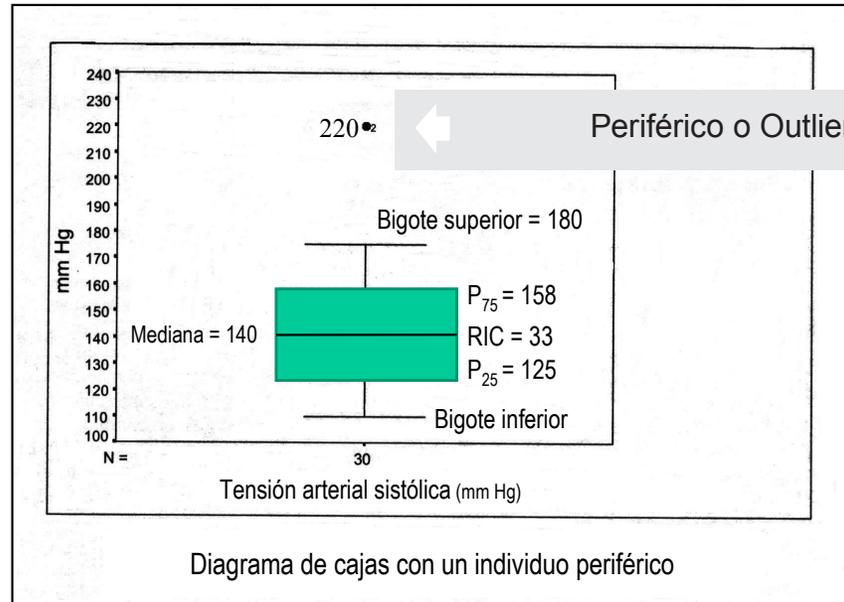
Diagrama de caja (box & whister plot)

## Gráficos de caja (box-plot)

¿Es legítimo un dato puntual cuyo valor es excepcionalmente grande o pequeño?

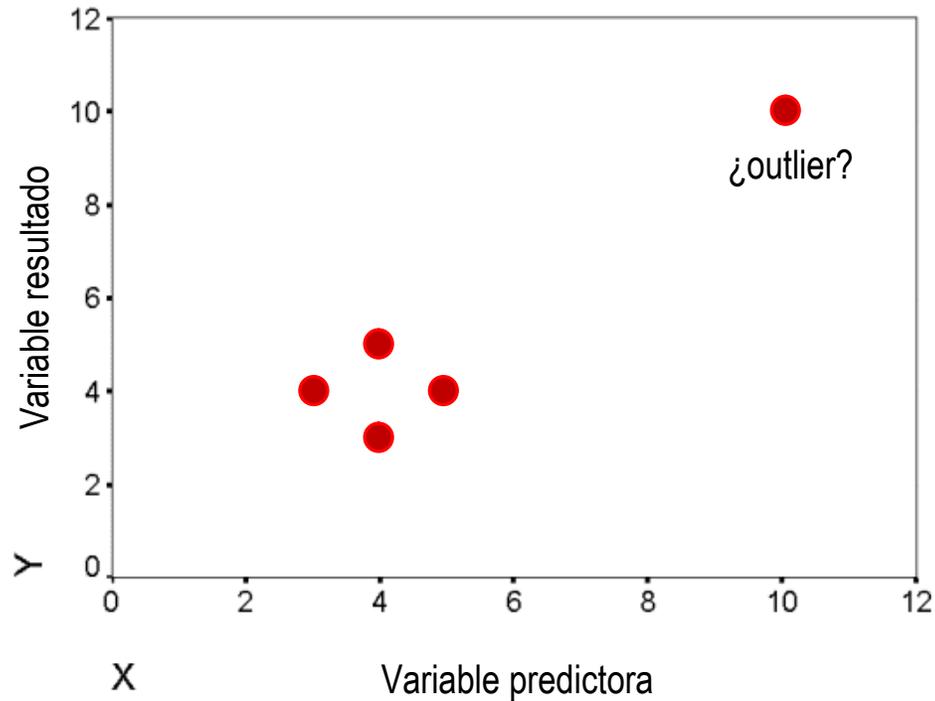
¿Es 220 un valor mal registrado?

¿Es el resultado de un error o accidente en la experimentación?



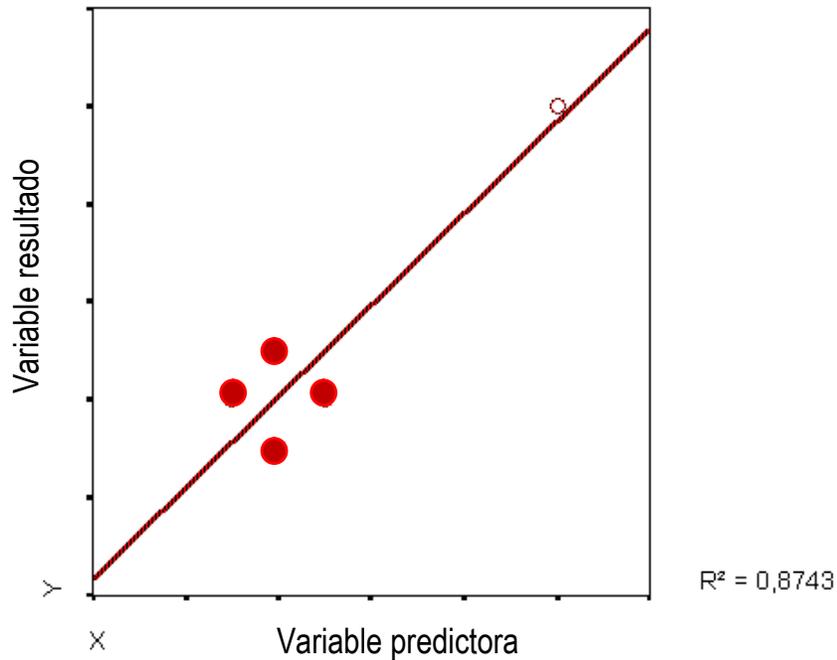
Martínez González MA, Jokin de Irala Estévez, Faulin Fajardo FJ. Bioestadística amigable. Madrid: Ediciones Díaz de Santos; 2001.

## Técnicas de detección de Outlier



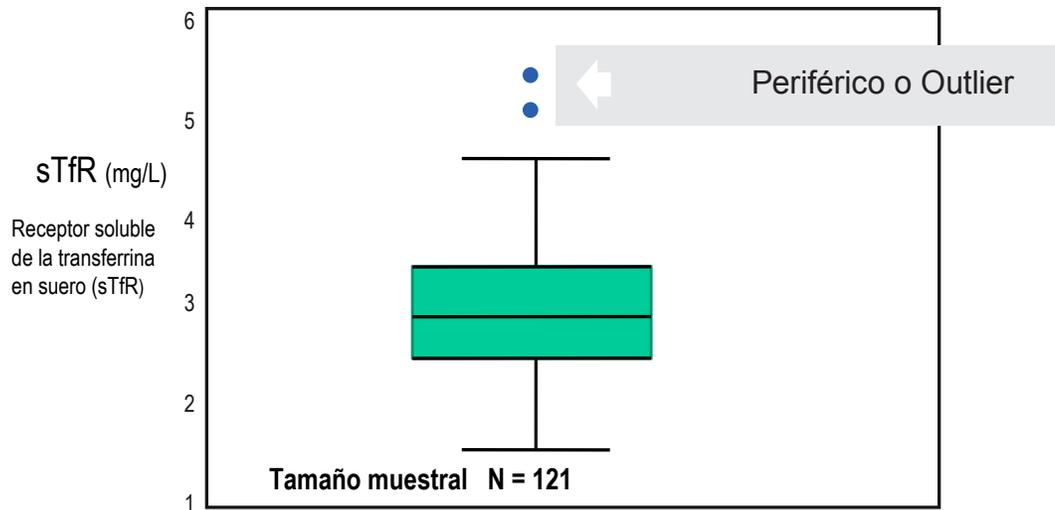
# Técnicas de detección de Outlier

Coeficiente de determinación  $R^2$



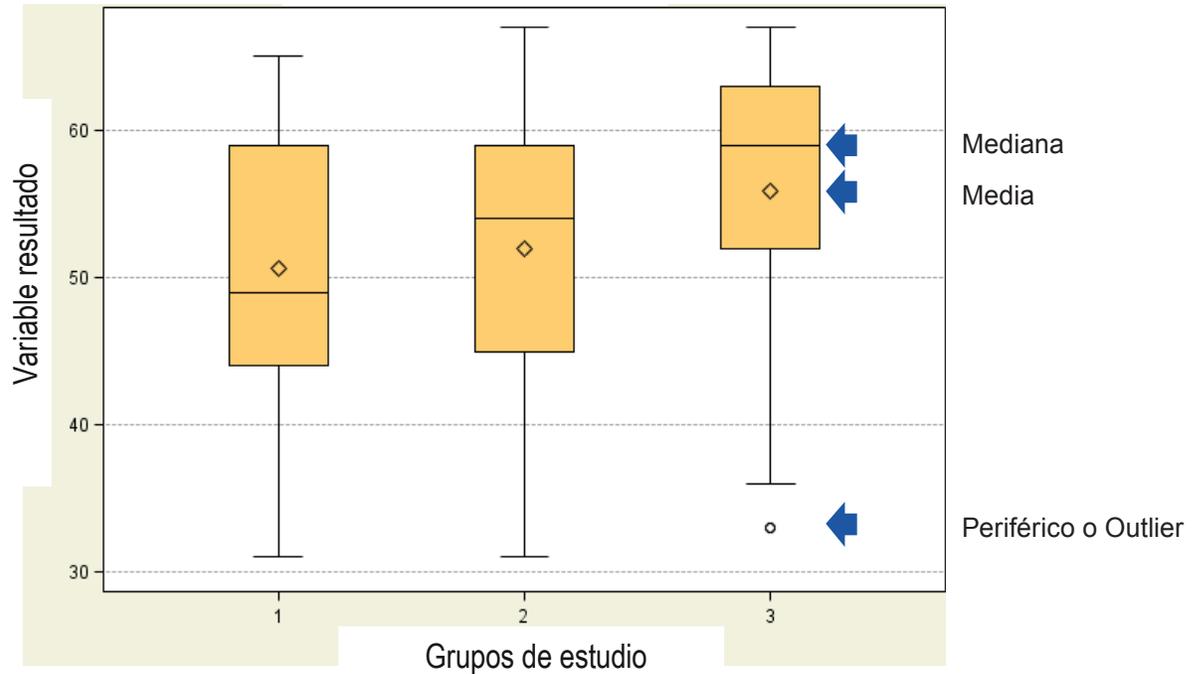
# Gráficos de caja (box-plot)

¿Sigue el sTfR una Distribución Normal?

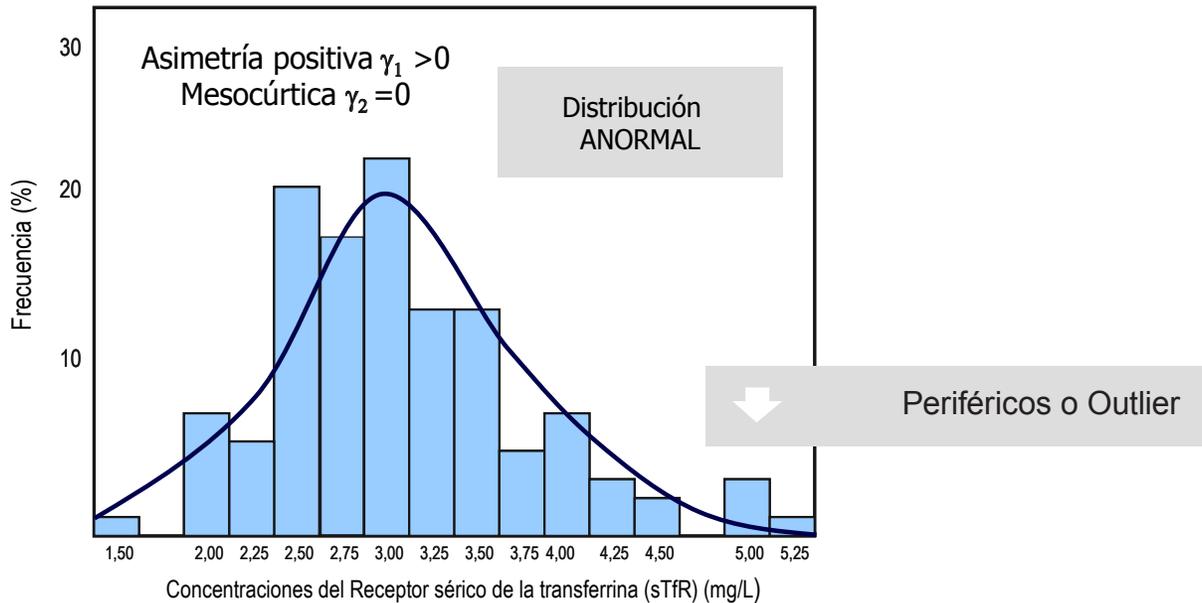


Distribución de los valores de sTfR en el grupo de referencia

## Gráficos de caja (box-plot)

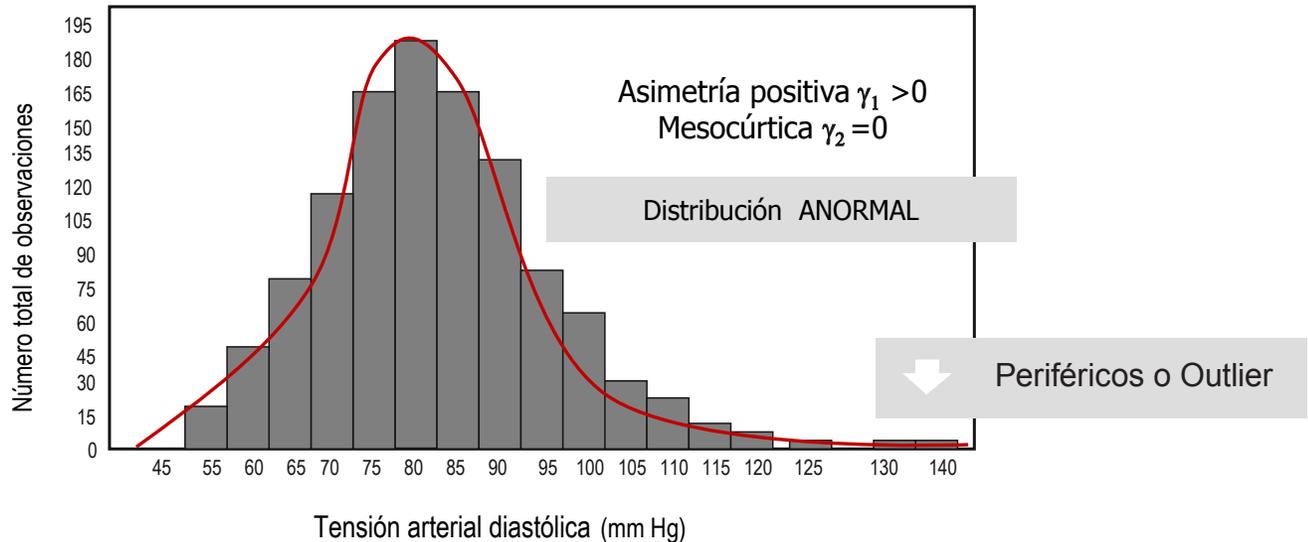


# Gráficos de histograma



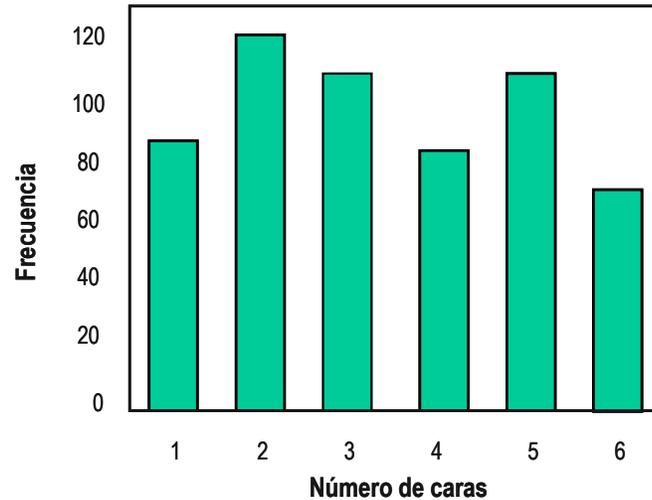
Histograma de frecuencias de distribución del sTfR en una población sana

# Gráficos de histograma



Histograma de distribución de la tensión arterial diastólica en el estudio de Venecia N= 1002

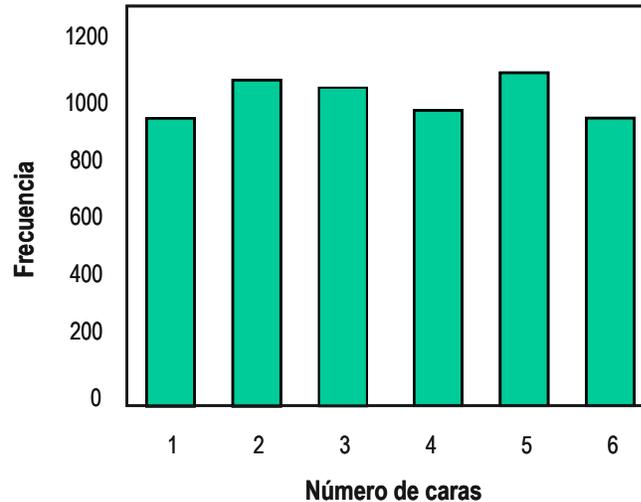
## Distribución uniforme discreta



Resultados empíricos obtenidos en la realidad  
después de lanzar un dado 600 veces

¿Cuál es el número que sale con más frecuencia?

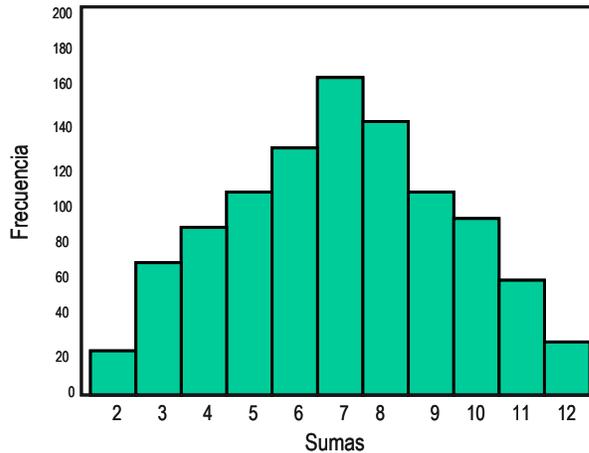
## Distribución uniforme discreta



Resultados empíricos obtenidos en la realidad  
después de lanzar un dado 600 veces

¿Cuál es el número que sale con más frecuencia?

# Distribución normal



$7 = 3 + 4, 4 + 3, 6 + 1, 1 + 6, 2 + 5, 5 + 2$

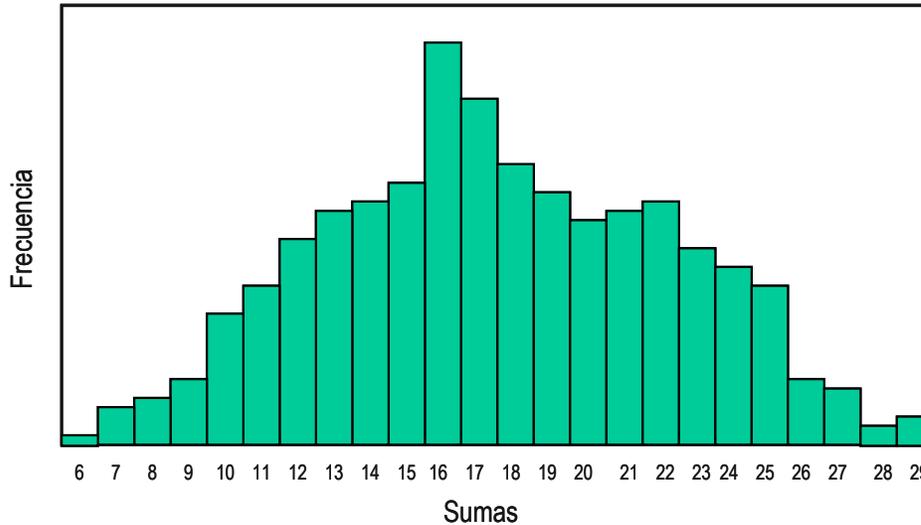
$12 = 6 + 6$

$2 = 1 + 1$

**Resultados empíricos (sumatorio) obtenidos en la realidad después de lanzar dos dados 1000 veces**

¿Cuál es el número que sale con más frecuencia?

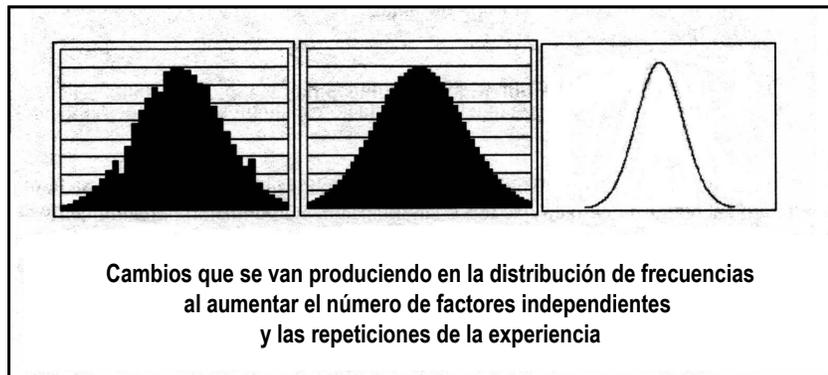
# Distribución normal



Resultados empíricos (sumatorio) obtenidos en la realidad después de lanzar 5 dados 1000 veces

¿Cuál es el número que sale con MENOS frecuencia?

## Distribución normal



Número de factores independientes: Número de dados

Repeticiones de la experiencia: Número de tiradas

Martínez González MA, Jokin de Irala Estévez, Faulin Fajardo FJ. Bioestadística amigable. Madrid: Ediciones Díaz de Santos; 2001.

¿La altura de las personas sigue una distribución normal?

¿De qué depende la altura de una persona?

1. De la altura del padre
2. De la altura de la madre
3. De su nutrición en la infancia y adolescencia
4. De la velocidad de maduración de su esqueleto
5. Del aporte que reciba de vitamina D
6. Del aporte de calcio
7. Del ejercicio físico
8. De sus niveles de hormona del crecimiento
9. Otras variables (algunos de ellas desconocidas)

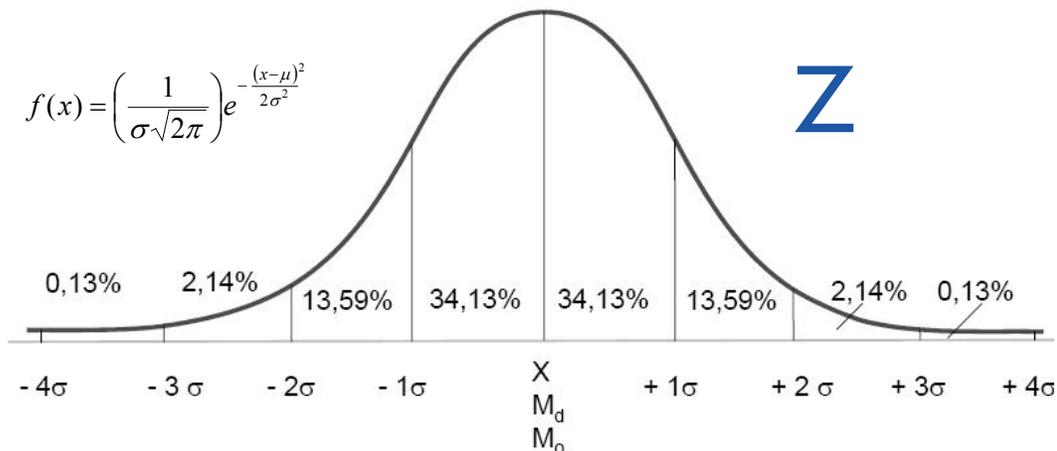
## Características de la distribución normal

1. Tener forma de campana.
2. Ser simétrica.
3. Mesocúrtica (no excesivamente plana ni muy picuda).
4. Coincidir la media, mediana y moda.
5. Tener aproximadamente el 95% de sus valores dentro del intervalo  $\text{media} \pm 2$  desviaciones estándar.

## Distribución normal

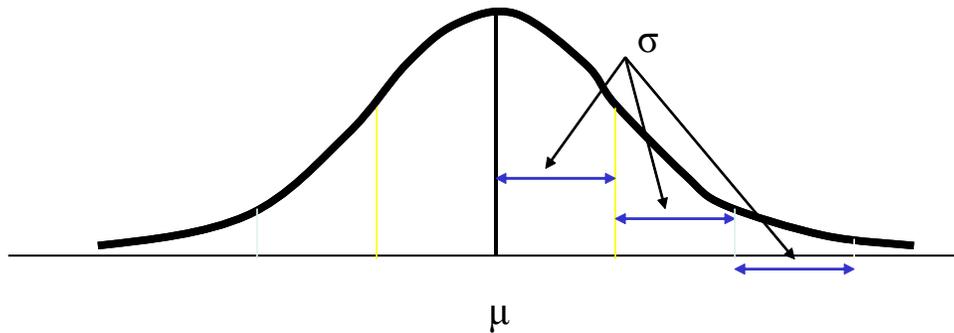
- El 100% de los valores deben estar por encima y por debajo de 3 veces la desviación estándar (s)  
(valores entre  $X \pm 3 s$ )
- La distribución debe ser SIMÉTRICA. La asimetría en valor absoluto  $\gamma_1$  debe ser menor que 2 veces su error estándar  
[Asimetría]  $< 2 EE_{\text{asimetría}}$      $\gamma_1 < 2 EE\gamma_1$
- La distribución debe ser MESOCÚRTICA. La **curtois** en valor absoluto  $\gamma_2$  debe ser menor que 2 veces su error estándar  
[curtois]  $< 2 EE_{\text{curtois}}$      $\gamma_2 < 2 EE\gamma_2$

## Propiedades de la distribución “normal”



- Es simétrica y media, mediana y moda coinciden en el punto central
  - Si añadimos el valor de la desviación típica y lo restamos a la media entre ambos extremos queda comprendido el 68.26% de los sujetos.
  - Si sumamos o restamos  $2\sigma$  a la media el universo comprendido entre estos extremos es de 97.7% . Con  $3\sigma$  será el 99.9% del universo.
- ➡ Con  $\mu \pm 1,96\sigma$  queda comprendido entre ambos valores el 95%

## Propiedades de la curva de Gauss



Media  $\pm$  D.E.  $\alpha$  68% del área total

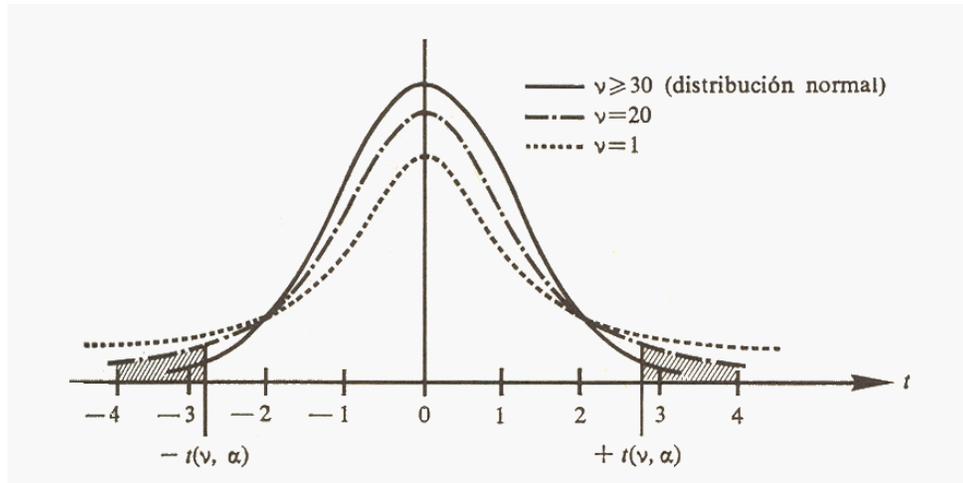
Media  $\pm$  2 D.E.  $\alpha$  95% del área total

Media  $\pm$  3 D.E.  $\alpha$  99% del área total

# Distribución de la t de Student

Grados de libertad  $\nu = n - 1$

$$\nu = n - 1$$



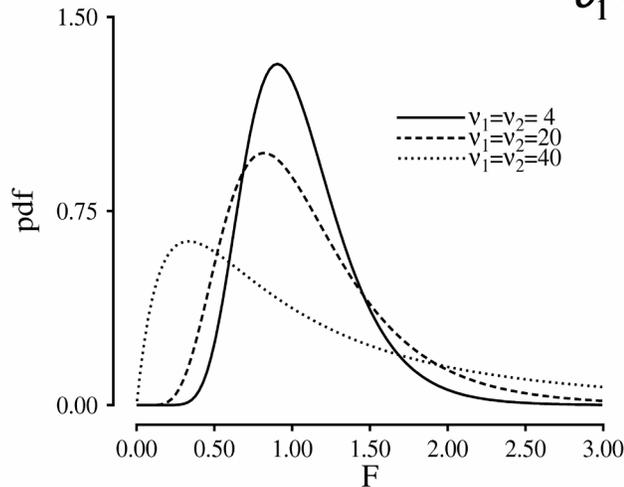
**valores "t":**  $t = \frac{(x - \mu)}{s}$

# Distribución de la F de Snedecor

**valores "F":**  $F = \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2}$

Grados de libertad  $v_1$  y  $v_2$

$v_1 = n_1 - 1$      $v_2 = n_2 - 1$



## ¿Qué variables biológicas siguen una distribución normal? ¿En qué circunstancias y en qué poblaciones?

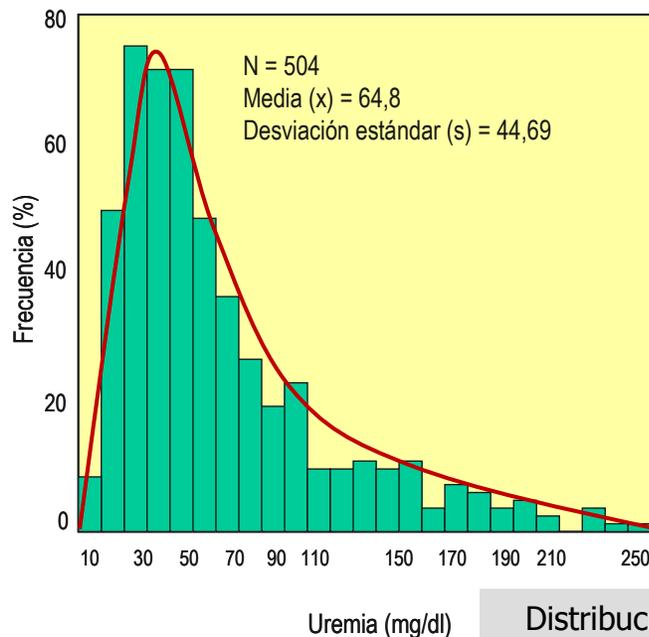
- Colesterol sérico población sana
- Consumo de alcohol
- Peso corporal
- Bilirrubina en pacientes cirróticos
- Urea en sangre en población sana
- Urea en sangre en pacientes con insuficiencia renal
- Troponina I en población sana
- Excreción urinaria de sodio en orina de 24 horas

Asimétrica  $\gamma_1 > 0 (+)$   
Leptocúrtica  $\gamma_2 > 0 (+)$

Distribución ANORMAL

## Histograma de la uremia en 504 individuos

¿Sigue una Distribución Normal?



Criterio de normalidad

$$\frac{\bar{X}}{2} > s$$

$$S \text{ (desviación típica)} < X \text{ (media aritmética)}/2 \text{ (CV}<50\%)$$

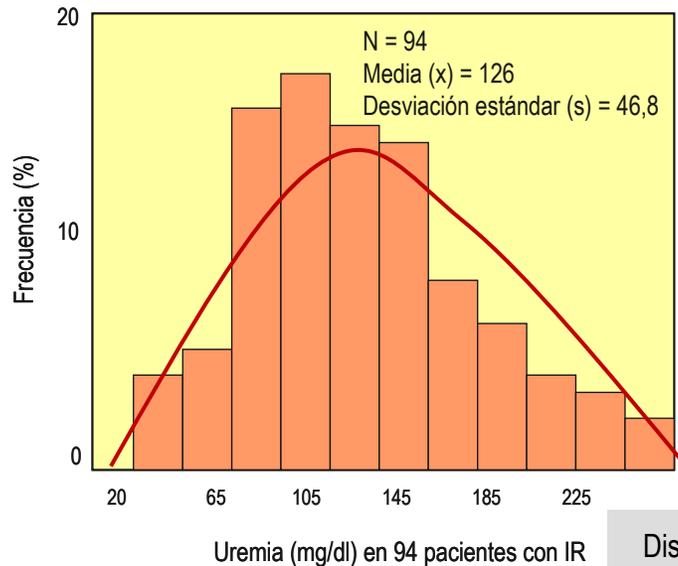
$$S \text{ (desviación típica)} > X \text{ (media aritmética)}/2 \text{ (CV}>50\%)$$

$$44,69 > 64,8/2 \text{ (32,4) (CV=69\%)}$$

Distribución ANORMAL

## Histograma de la urea en sangre en 94 pacientes con insuficiencia renal

¿Sigue una distribución normal?



$$\frac{\bar{X}}{2} > s$$

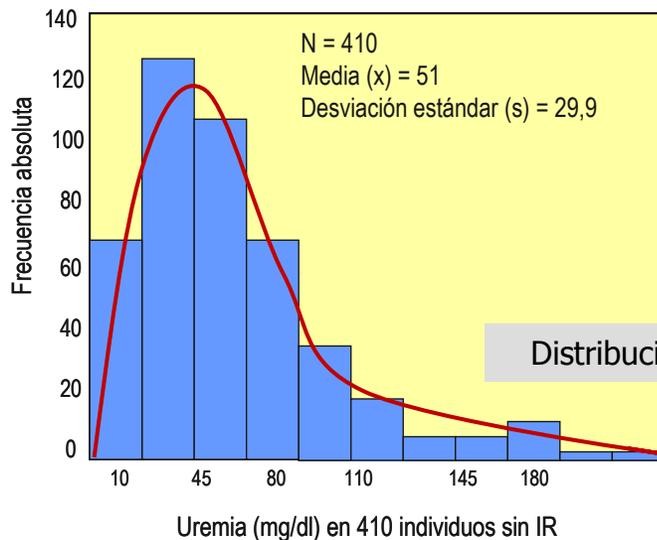
$$S \text{ (desviación típica)} < X \text{ (media aritmética)} / 2 \text{ (CV} < 50\%)$$

$$46,8 < 126 / 2 \text{ (63) (CV} = 37\%)$$

Distribución NORMAL

## Histograma de la urea en sangre en 410 individuos sin insuficiencia renal

¿Sigue una distribución normal?



$$\frac{\bar{X}}{2} > s$$

$$S \text{ (desviación típica)} > X \text{ (media aritmética)}/2 \text{ (CV}>50\%)$$

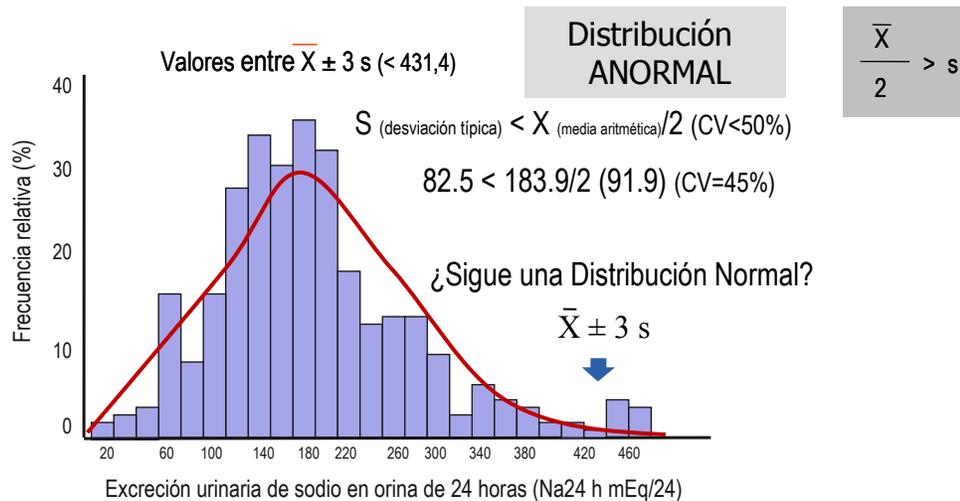
$$29,9 > 51/2 \text{ (25,5) (CV=59\%)}$$

Distribución ANORMAL

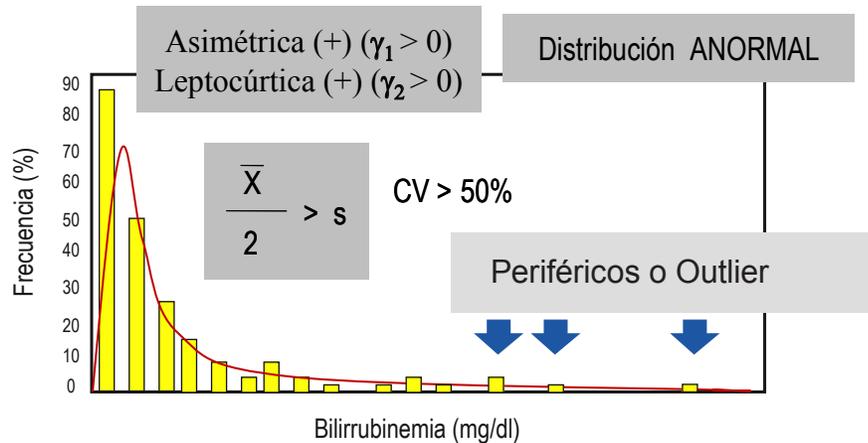
# Medidas de tendencia central

## Medidas de dispersión o variabilidad

Media = 183,9  
Desviación estándar = 82,5  
N = 293



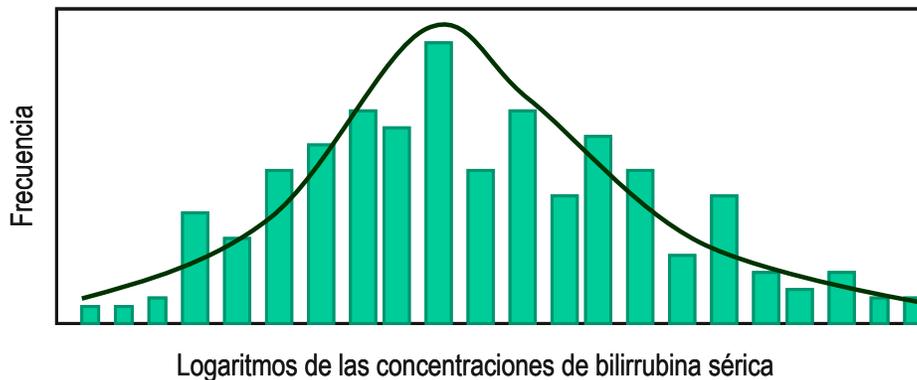
# Condiciones y pruebas de normalidad



**Distribución de frecuencias de los niveles de bilirrubina en pacientes cirróticos.  
(Distribución Anormal con asimetría positiva)**

# Condiciones y pruebas de normalidad

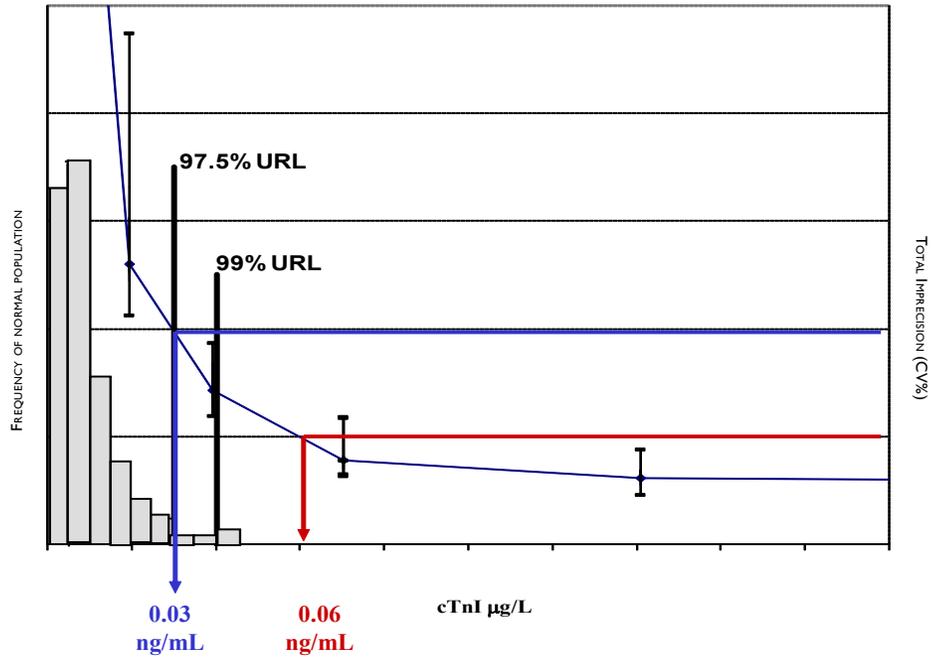
## Transformación logarítmica



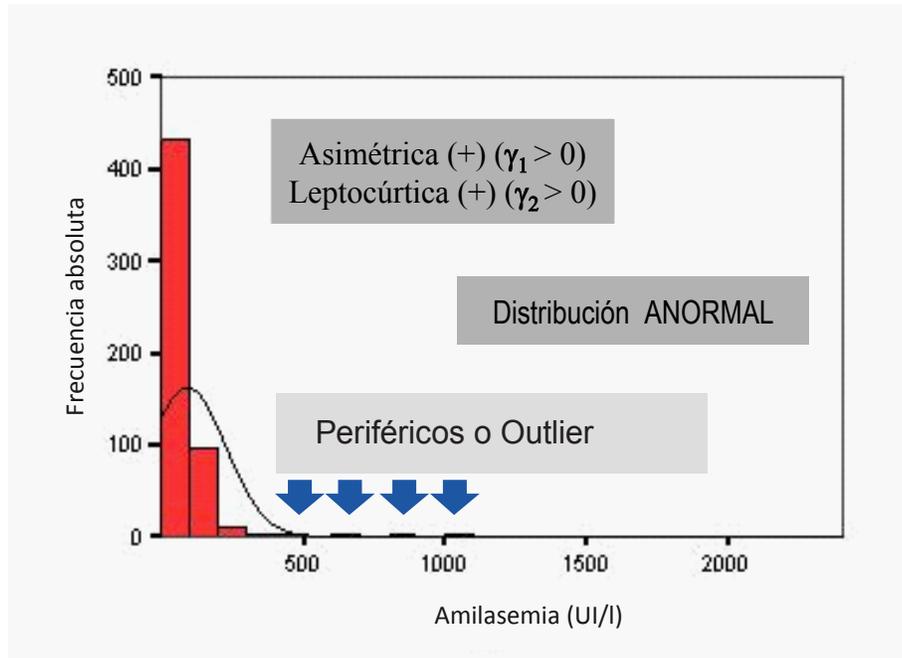
**Distribución de frecuencias de los logaritmos de los niveles de bilirrubina en pacientes cirróticos.**

**Con esta transformación matemática se corrige la asimetría positiva anterior**

# Troponina I en población sana



## ¿La amilasemia sigue una Distribución Normal?



## Estadísticos descriptivos de la amilasemia

	N	Media	Desviación típica	Asimetría		Curtosis	
				$\gamma_1$	$EE\gamma_1$	$\gamma_2$	$EE\gamma_2$
Amilasa	549	90,86	135,043	11,359	0,104	170,89	0,208

	Asimetría			Curtosis		
	$\gamma_1$	$EE\gamma_1$	$\gamma_1/EE\gamma_1$	$\gamma_2$	$EE\gamma_2$	$\gamma_2/EE\gamma_2$
Amilasa	11,359	0,104	109,22	170,89	0,208	821,6

$$[\text{Asimetría}] > 2 EE_{\text{asimetría}} \quad \gamma_2 > 2 EE\gamma_2$$

Distribución ANORMAL

$$[\text{Curtosis}] > 2 EE_{\text{curtois}} \quad \gamma_2 > 2 EE\gamma_2$$

Distribución ANORMAL

## Test de normalidad

- Test de Kolmogoroff-Smirnoff-Lilliefors (test más potente para muestra  $> 30$  casos).
- Test de Shapiro-Wilk (la prueba más potente para muestra  $< 30$  casos).
- Test de D'Agostino.
- Test de Anderson-Darling.
- Test chi-cuadrado de bondad de ajuste.
- Test de Contraste de asimetría: Sirve para contrastar la hipótesis de que el coeficiente de asimetría de Fisher  $\gamma_1$  es cero. Propiedad que verifica la distribución normal ( $\gamma_1 = 0$ ).
- Test de Curtosis: Contraste de apuntamiento, sirve para contrastar la hipótesis de que el coeficiente  $\gamma_2$  curtosis de Fisher (coeficiente de apuntamiento) es cero. Propiedad que verifica la distribución normal ( $\gamma_2 = 0$ ).
- Test de Jarque-Bera.

## Test de Kolmogorov-Smirnov

- Es el test de normalidad más difundido en los programas estadísticos.
- Otorga un MENOR PESO a las observaciones extremas (periféricos o OUTLIER) y por ello es menos tolerante a las desviaciones que normalmente se producen en estos tramos distales.
- Cuando se dispone de un número suficiente de datos (amplio volumen muestral), los demás tests de normalidad serán capaces de detectar diferencias pequeñas aún cuando estas no sean relevantes para la mayor parte de los objetivos, por eso se recomienda el test de Kolmogorov-Smirnov (se considera uno de los tests más potentes para muestras mayores de 30 casos).
- Se basa en la idea de comparar la función de distribución acumulada de los datos observados con la de una DISTRIBUCIÓN NORMAL, midiendo la máxima distancia entre ambas curvas.
- Como en todos los tests de contraste de hipótesis, la hipótesis nula ( $H_0$ ) se rechaza cuando el valor del estadístico calculado supera un cierto valor crítico (tabulado) que se obtiene de una tabla de cálculo de probabilidades.
  - $p > 0,05$  se acepta la  $H_0$  (se cumple la normalidad), se pueden usar test paramétricos.
  - $p < 0,05$  se acepta la  $H_a$  (no se cumple la normalidad), se deben usar test NO paramétricos.

## Pregunta de investigación

¿Los valores analíticos de ambas técnicas siguen una distribución normal?

Objetivo: Evaluación de la amilasa técnica 1 frente a la técnica 2.

Amilasa (sérica)  
Técnica 1



Amilasa (sérica)  
Técnica 2

¿Cuál es el test estadístico más acertado?

Test de Kolmogorov-Smirnov

## Test de Kolmogorov-Smirnov

Resultados del test		Técnica 1	Técnica 2
N		548	548
Parámetros normales	Media	91,03	79,29
	Desviación típica	135,11	120,52
Diferencias más extremas	Absoluta	0,289	0,301
	Positiva	0,289	0,301
	Negativa	- 0,280	- 0,299
Z de Kolmogorov-Smirnov		6,763	7,034
Significación bilateral		0,000	0,000

Como la significación  $p < 0,05$  se rechaza la hipótesis nula  $H_0$  (normalidad) y se acepta la hipótesis alternativa  $H_a$  (no se cumple la normalidad), se deben usar test NO paramétricos.

## Test de Jarque-Bera

- El test de Jarque-Bera se realiza para contrastar:

$H_0: \gamma_1 = \gamma_2 = 0$ , lo que significa simetría y curtosis exactas ( $\gamma = 0$ )

$H_1: \gamma_1 \neq 0$  o  $\gamma_2 \neq 0$  lo que significa que existe curtosis ( $\gamma \neq 0$ ) o asimetría.

- La consecuencia es que si existe asimetría o curtosis falla la normalidad.

$H_0: p > 0,05$   $\longrightarrow$  simetría y curtosis exactas ( $\gamma = 0$ )  $\longrightarrow$  test paramétricos

$H_1: p < 0,05$   $\longrightarrow$  curtosis ( $\gamma \neq 0$ ) o asimetría  $\longrightarrow$  test NO paramétricos

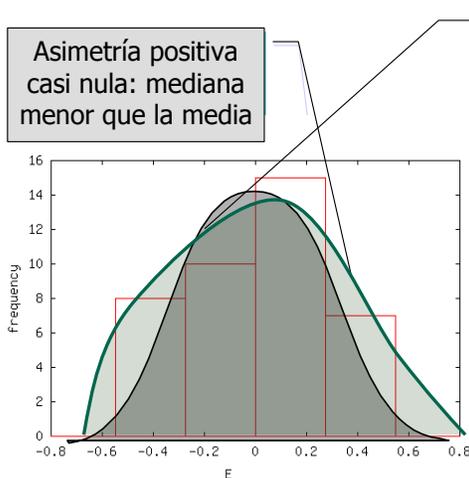
- Al contrastarlo conjuntamente (curtosis y asimetría) exige un fallo mayor de alguna de ellas o de ambas para rechazarse.

$$\gamma_1 = \frac{\sum_{i=1}^n \varepsilon_i^3}{\sigma_R^3}$$

$$\gamma_2 = \frac{\sum_{i=1}^n \varepsilon_i^4}{\sigma_R^4} - 3$$

$$JB = t_1^2 + t_2^2 = T \left[ \frac{\hat{\gamma}_1^2}{6} + \frac{\hat{\gamma}_2^2}{24} \right]$$

# Test de Jarque-Bera



Asimetría positiva casi nula: mediana menor que la media

Forma teórica de la normal

TEST de normalidad de JARQUE-BERA

Estadístico  $\chi^2 = 1.54$

valor de  $p = 0.463$  ( $p > 0,05$ )

**Se acepta la  $H_0$**



$H_0: \gamma_1 = \gamma_2 = 0$ , lo que significa simetría y curtosis exactas.  
DISTRIBUCIÓN NORMAL

Curtosis negativa: mas apuntamiento que la normal

Forma teórica de la distribución empírica

## Estadística descriptiva

TIPO DE DISTRIBUCIÓN	SIMÉTRICA	ASIMÉTRICA
Índice de tendencia central	Media	Mediana
Índice de dispersión	EEM	Rango intercuartil
“Peso” de valores extremos (outlier o periféricos)	Importante	Escaso o nulo
Fiabilidad del estadístico	Mucha	Poca

## Pruebas paramétricas: Requisitos

- Las variables cuantitativas tienen que ser continuas.
- Los datos siguen una distribución normal.
- Los datos numéricos se expresan con la media y la desviación estándar o el EEM
- Homogeneidad de las varianzas muestrales.
- Tamaño muestral  $> 30$ .
- Presentan un mayor poder de contraste que los tests no paramétricos, y pueden analizar interacciones entre variables predictoras.
- Cuando se dan las condiciones de aplicación, las pruebas paramétricas tienen más potencia que las no paramétricas, pero, cuando esto no es así, el riesgo alfa puede ser mayor que el especificado de antemano.

## Pruebas NO paramétricas: Requisitos

- Las variables cuantitativas pueden ser ordinales o discretas.
- Los datos siguen una distribución libre (no necesariamente normal), uno de los grupos puede tener una distribución normal mientras que el otro no.
- Los datos numéricos se expresan con la mediana y los intervalos intercuartílicos.
- Heterogeneidad de las varianzas muestrales (en uno o más de un grupo).
- Tamaño muestral  $< 30$ .
- Presentan un menor poder estadístico que los tests paramétricos.
- Trabajan con la simple ordenación y recuento (asignando rankings) a los valores de la variable sin importar el tipo distribución.

## Principales pruebas NO paramétricas

- Chi cuadrado de Pearson.
- Test exacto de Fisher.
- Chi cuadrado de Mc Nemar.
- Q de Cochran.
- U de Mann Whitney.
- Wilcoxon.
- Kruskal Wallis.
- Friedman.

<http://www.slideshare.net/freddygarcia/pruebas-no-parametricas-presentation>

## Pruebas robustas: Características

- Otra alternativa a los métodos paramétricos; la utilización de los llamados métodos robustos. Estos métodos son menos potentes que los paramétricos, pero se muestran superiores a los no paramétricos clásicos.
- La principal de sus ventajas es que no se afectan por la existencia de datos anómalos.
- Los métodos robustos pueden ser de utilidad para la realización de inferencias sin tener que «depurar» los datos extremos.
- Como parámetros de localización (posición central) se recomienda en primer lugar, el estimador de Huber y, en segundo lugar, la media  $a$ -recortada muestral (con  $a = 0,2$ ).
- Como medida de dispersión se utiliza la desviación absoluta mediana estandarizada.
- Métodos robustos para el contraste de hipótesis con los parámetros antes referidos, se pueden construir intervalos de confianza robustos y realizar contrastes de hipótesis.

# Métodos robustos para el cálculo de medidas de posición:

## Medidas robustas de tendencia central

### Media $\alpha$ -winsorizada muestral

Se sustituye un determinado porcentaje,  $\alpha$ , (20% Generalmente) de valores extremos a cada lado de la muestra por el valor más próximo no sustituido.

### Media $\alpha$ -recortada muestral

Se eliminan las  $k$  observaciones extremas de cada lado en lugar de winsorizarlas, calculando la media aritmética de las observaciones restantes.

### Mediana muestral

Divide la distribución en dos partes con el mismo número de elementos.

### Estimador de Huber

Se encuentra dentro de los denominados M-estimadores, que generalizan al estimador de máxima verosimilitud con buenas propiedades de robustez y eficiencia. En este caso se descartan las observaciones que sean mayores (o menores) a una constante.

# Métodos robustos para el cálculo de medidas de dispersión:

## Medidas robustas de dispersión

Desviación absoluta mediana estandarizada

Es la mediana de las desviaciones absolutas a la mediana.

Cuasi desviación típica  $\alpha$ -winsorizada muestral

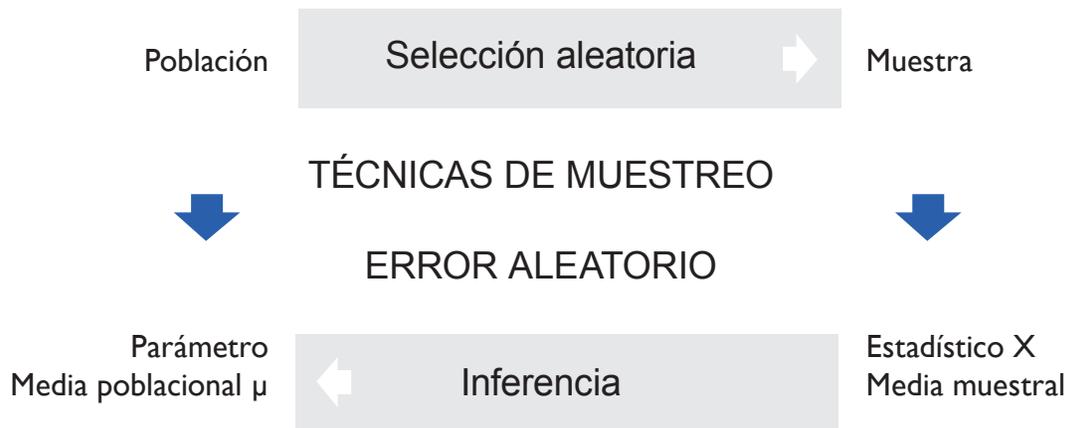
En la que se sustituye un determinado porcentaje de valores extremos a cada lado de la muestra por el valor más próximo no sustituido.

## Pruebas Robustas de contraste de hipótesis

- Métodos robustos para el contraste de hipótesis con los parámetros antes referidos, se pueden construir intervalos de confianza robustos y realizar contrastes de hipótesis.
- El test de Yuen es una alternativa a las pruebas de la t de Student y U de Mann-Whitney para tamaños muestrales pequeños y distribuciones no normales.
- Para comparar más de dos medias, cuando no puede utilizarse el ANOVA clásico, deberíamos recurrir a los clásicos no paramétricos de Kruskal-Wallis o, mejor, a métodos robustos como la generalización robusta del test Welch, que utiliza medias a-recortadas muestrales.
- Deberíamos recurrir a métodos no paramétricos (Spearman) o, mejor, a métodos robustos como el coeficiente de porcentaje ajustado poblacional y el estimador robusto de regresión medio bponderado.

<http://www.slideshare.net/freddygarcia/pruebas-no-parametricas-presentation>

## Selección de la muestra



Estructura del proceso de la INFERENCIA ESTADÍSTICA

## Prueba de rachas

- Sirve para determinar si una muestra de observaciones es o no aleatoria, es decir, si las observaciones de una determinada secuencia son independientes entre si.
- En una serie temporal, las observaciones no son aleatorias: lo que ocurre con una observación cualquiera depende, de las características de una observación anterior.
- En una muestra aleatoria, lo que ocurre con una observación cualquiera sea independiente de las características de la(s) observacion(es) anterior(es).
- Esta prueba permite determinar si el número de rachas ( $R$ ), observado en una determinada muestra de tamaño  $n$  es lo suficientemente grande o lo suficientemente pequeña como para poder rechazar la hipótesis de independencia-aleatoriedad de las observaciones.

## Prueba de rachas

El concepto de RACHA hace referencia a una secuencia de observaciones de un mismo tipo.

Supongamos que lanzamos una moneda al aire 10 veces seguidas, repitiendo la experiencia en tres ocasiones...

¿Cómo garantizamos que el resultado obtenido es ALEATORIO?

## Prueba de rachas

1º EXPERIENCIA: Tenemos 5 RACHAS: CCC, X,CC, XXX, C



2º EXPERIENCIA: Tenemos 2 RACHAS: CCCCC, XXXXX



3º EXPERIENCIA: Tenemos 10 RACHAS: CX, CX, CX, CX, CX



¿En cuál de las 3 experiencias podemos garantizar que el resultado obtenido es ALEATORIO?

## Prueba de rachas

Prueba de RACHAS	EXP 1 <sup>a</sup>	EXP 2 <sup>a</sup>	EXP 3 <sup>a</sup>
Valor de prueba	1	1	1
Casos < Valor de prueba	4	5	5
Casos ≥ Valor de prueba	6	5	5
Casos en total	10	10	10
Número total de rachas	5	2	10
Z	- 0,211	- 2,348	2,348
Significación bilateral	0,833	0,019	0,019

Como el nivel de significación es mayor a 0,05 en la experiencia 1, podemos aceptar la hipótesis de independencia ( $H_0$ ) y concluir que la secuencia de observaciones estudiada es ALEATORIA.

Como el nivel de significación es inferior a 0,05 en la experiencia 2 y 3, podemos rechazar la hipótesis de independencia y concluir que la secuencia de observaciones estudiada NO es aleatoria ( $H_a$ ).

## Pregunta de investigación

¿Los valores analíticos de ambas técnicas son aleatorias?

Objetivo: Evaluación de la amilasa técnica 1 frente a la técnica 2.

Amilasa (sérica)  
Técnica 1



Amilasa (sérica)  
Técnica 2

¿Cuál es el test estadístico más acertado?

Prueba de RACHAS

## Prueba de rachas

	Técnica 1	Técnica 2
Valor de prueba (mediana)	70	61
Casos < Valor de prueba	267	267
Casos $\geq$ Valor de prueba	282	281
Casos en total	549	549
Número total de rachas	259	263
Z	- 1,393	- 1,012
Significación bilateral	0,164	0,312

Como el nivel de significación es mayor a 0,05 en ambas técnicas, podemos aceptar la hipótesis de independencia ( $H_0$ ) y concluir que la secuencia de observaciones estudiada es ALEATORIA.

## Capítulo 3

# Publicación Bioestadística: Tamaño muestral, Error $\alpha$ y $\beta$ .

**Dr. Jacobo Díaz Portillo**  
Coordinador de Formación Continuada  
Hospital Universitario de Ceuta

# Materiales y métodos

## Estructura

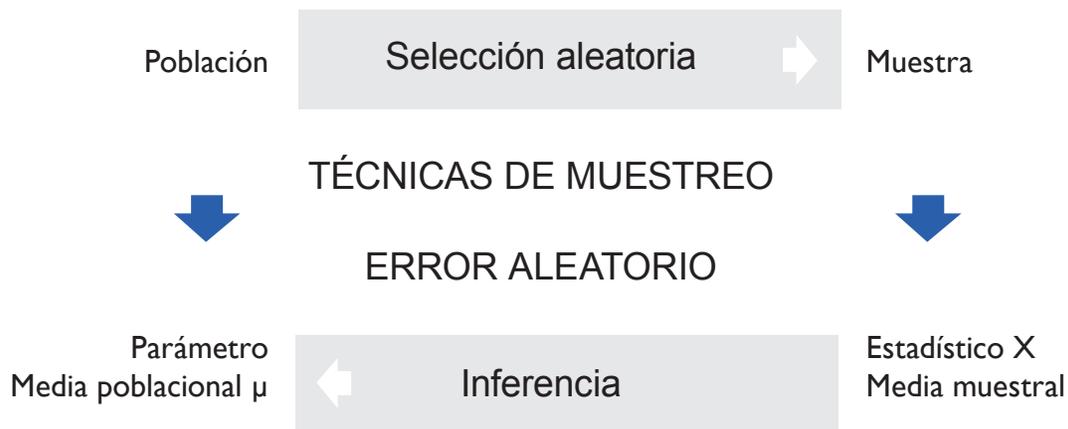
1. Tipo de diseño (estudio).
2. *Sujetos incluidos (criterios de selección de la muestra).*
3. Fuentes de información.
4. Descripción de la intervención y definición del punto final de la investigación (estudios experimentales).
5. Instrumentos y procedimientos de medida (cuestionarios).
6. Definición de las variables y covariables incluidas en el estudio (factor de estudio y criterio de evaluación).
7. Limitaciones y consideraciones éticas de la investigación.
8. Pruebas estadísticas a utilizar (programa estadístico).

# Materiales

## Pacientes

1. *Criterios de selección de la muestra (inclusión y exclusión).*
2. Procedencia.
3. Técnica de reclutamiento utilizada.
4. Tiempo empleado para reunir a todos los participantes.
5. Principales características de los pacientes.
  - Edad.
  - Sexo.
  - Procedencia.
  - Gravedad.
  - Años de evolución de la enfermedad.
  - Patologías asociadas...
6. Cálculo de tamaño muestral.

## Selección de la muestra



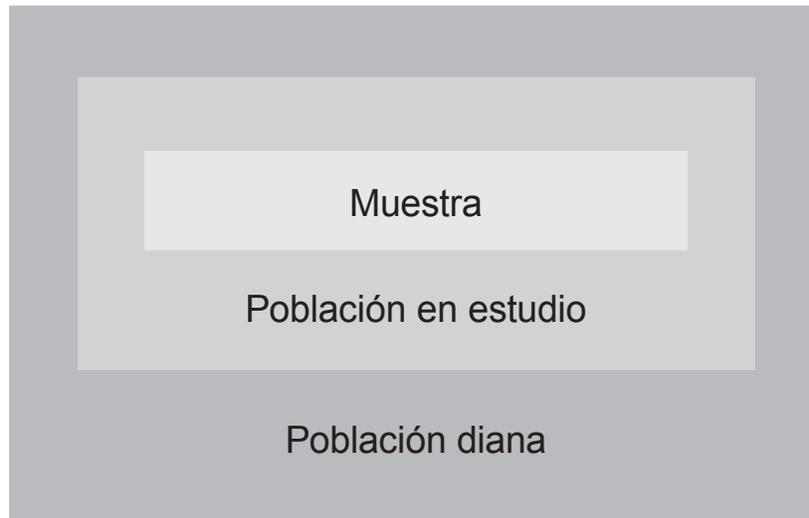
Estructura del proceso de la INFERENCIA ESTADÍSTICA

**INFERENCIA ESTADÍSTICA:** Es el conjunto de procedimientos estadísticos que permiten pasar de lo particular (la muestra) a lo general (la población).

**TÉCNICAS DE MUESTREO:** Es el conjunto de procedimientos que permiten elegir muestras que sean representativas de la población (de todos los individuos de la población).

**ERROR ALEATORIO:** Es el error que se comete al estudiar un problema en base a una muestra elegida al azar (es siempre un error inevitable). La imposibilidad de estudiar a todos los individuos de la población obliga al investigador a trabajar con muestras, lo que supone siempre cometer un error aleatorio.

## Muestra de la población



## Proceso de selección de individuos para el estudio

**Población diana:** Conjunto de individuos al que hace referencia la pregunta principal del objetivo del estudio. Es la población a la que se desearía generalizar los resultados obtenidos. Se define principalmente sobre la base de características demográficas y clínicas generales.

Criterios de selección



Validez externa

**Población de estudio:** Es un subconjunto de la población diana definida por criterios de selección establecidos en un protocolo. Esta población presenta determinadas características geográficas y temporales que la hacen accesible al estudio.

Muestreo



Validez interna

**Muestra:** Es un conjunto de individuos que pueden ser estudiados en el momento de realizar el estudio y que cumplen con los criterios definidos en el protocolo de estudio.

Población diana: Pacientes con HTA esencial del área sanitaria de Ceuta.

Criterios de selección



Validez externa

Población de estudio: Pacientes con HTA esencial leve-moderada de edad comprendida 40-60 años del área sanitaria de Ceuta.

Muestreo



Validez interna

Muestra: Pacientes con HTA esencial leve-moderada de edad comprendida 40-60 años registrados en los ficheros del Centro de Salud 1 que incluyen la edad y el sexo.

Población diana: Población pediátrica de Ceuta.

Criterios de selección



¿Validez externa?

Población de estudio: Niños que acuden a la consulta de Atención al niño sano de Atención Primaria de edad comprendida 12-14 años.

Muestreo



¿Validez interna?

Muestra: Niños que acuden al Programa de Atención al niño sano de los centros de Salud de Atención Primaria de edades comprendida entre 12-14 años que sus madres demandan la realización de una analítica rutinaria completa.

Población diana: Población gestante de Ceuta.

Criterios de selección



Validez externa

Población de estudio: Mujeres gestantes del área sanitaria que acuden a la consulta de Obstetricia del INGESA de Ceuta.

Muestreo



Validez interna

Muestra: Mujeres gestantes del área sanitaria de Ceuta que acuden a la consulta de Obstetricia del INGESA –embarazo controlado- con cartilla de embarazo completa, que incluyen las variables predictoras (etnia, nivel cultural, edad, etc) y la variable resultado (determinación de anticuerpos frente al virus de la rubeola en el primer trimestre) incluidas en el estudio.

## Proceso de selección de individuos para el estudio

Población diana: Población inmigrante de Ceuta.

Criterios de selección



Validez externa

Población de estudio: Individuos inmigrantes residentes en el campamento de Calamocarro de Ceuta atendidos por el personal sanitario.

Muestreo



Validez interna

Muestra: Individuos inmigrantes residentes en el momento del estudio en el campamento de Calamocarro de Ceuta, atendidos por el personal sanitario, y en cuya ficha se habían recogido las variables predictoras (filiación, país de origen, edad), así como el resultado de todas las pruebas complementarias (incluido la Variable resultado: prueba de la tuberculina) incluidas en el estudio.

# Material

## Pacientes

- Criterios de selección de la muestra (inclusión y exclusión).
- Procedencia.
- Técnica de reclutamiento utilizada.
- Tiempo empleado para reunir a todos los participantes.
- Principales características de los pacientes.
  - Edad.
  - Sexo.
  - Procedencia.
  - Gravedad.
  - Años de evolución de la enfermedad.
  - Patologías asociadas...
- *Cálculo de tamaño muestral.*

¿Qué error alfa y que error beta estoy dispuesto a asumir?

¿Cuántos pacientes debo elegir para que mi estudio sea correcto?

## Concepto e importancia y utilidad práctica del error alfa ( $\alpha$ ) y error beta ( $\beta$ )

- Determinación del tamaño muestral
  - Estudios para determinar parámetros poblacionales (sólo error alfa)
    - Estimación de una proporción
    - Estimación de una media
  - Estudios para contraste de hipótesis (error alfa y beta)
    - Comparación de dos proporciones
    - Comparación de dos medias
    - Correlación entre dos variables cuantitativas
- Rentabilidad diagnóstico de una prueba-test
  - Exactitud
    - Sensibilidad
    - Especificidad
  - Seguridad
    - Valor predictivo positivo (VPP)
    - Valor predictivo negativo (VPN).

## Riesgos al tomar decisiones

Ejemplo: Se juzga a un individuo por la presunta comisión de un delito

$H_0$ : Hipótesis nula

- Es inocente (no se encuentran indicios de culpabilidad)



Los datos pueden refutarla

Es la hipótesis que se acepta si las pruebas no indican lo contrario  
Rechazarla por error tiene graves consecuencias (condenar a un inocente)

$H_1$ : Hipótesis alternativa

- Es culpable (se encuentran indicios de culpabilidad)



No debería ser aceptada sin una gran evidencia a favor.  
Rechazarla por error tiene consecuencias consideradas menos graves que la anterior (liberar a un delincuente)

## Tipos de error al tomar una decisión

Tipos de error al tomar una decisión		Realidad	
		Culpable	Inocente
Veredicto Resultado de la prueba Sentencia del juez	Culpable Hipótesis ALTERNATIVA aceptada $H_a$	OK No error VP	Error $\alpha$ Muy grave FP Encarcelar a un inocente
	Inocente Hipótesis NULA aceptada $H_0$	Error $\beta$ FN Menos grave: Liberar un culpable	OK No error VN

Hipótesis nula ( $H_0$ ): No existen diferencias significativas en la eficacia clínica de dos fármacos A y B ( $A=B$ )

Hipótesis alternativa ( $H_a$ ): Existen diferencias en la eficacia clínica de dos fármacos A y B ( $A \neq B$ )

$A > B$  ó  $A < B$

**Tipos de ERROR ALEATORIO en una prueba estadística de contraste de hipótesis**

		<i>Realidad (población)</i>	
		Existe diferencia o asociación (Ho falsa) $A \neq B$ $H_a$	No existe diferencia o asociación (Ho cierta) $A=B$ $H_0$
 Resultado de la prueba (muestra)	Diferencia o asociación significativa $A \neq B$ (Rechazo de $H_0$ ) $H_a$	No error <b>VP</b>	Error tipo I ( $\alpha$ ) <b>FP</b>
	Diferencia o asociación no significativa $A=B$ (No rechazo de $H_0$ ) $H_0$	Error tipo II ( $\beta$ ) <b>FN</b>	No error <b>VN</b>

Ho: hipótesis nula.

Error tipo I (alfa): El investigador rechaza la  $H_0$  siendo esta verdadera en la población (el investigador concluye que existe diferencia cuando en realidad no la hay).

Error tipo II (beta): El investigador no rechaza la  $H_0$  siendo esta falsa en la población (el investigador concluye que ha sido incapaz de encontrar una diferencia que existe en la realidad).

## Tipos de error al contrastar hipótesis

	Realidad	
	$H_0$ Falsa $A \neq B$	$H_0$ Cierta $A=B$
Rechazo $H_0$ Acepto $H_1$ $A \neq B$	Correcto El tratamiento nuevo B tiene mayor efecto que A y el experimento lo confirma.	Error de tipo I El tratamiento nuevo B realmente no es mejor que el antiguo A pero se decide erróneamente que sí. Probabilidad $\alpha$
No Rechazo $H_0$ Acepto $H_0$ $A=B$	Error de tipo II El tratamiento nuevo B realmente si tiene mayor efecto que el antiguo A pero no lo percibimos en nuestro estudio. Probabilidad $\beta$	Correcto El tratamiento nuevo B no tiene mayor efecto que el antiguo A y así se decide. VN
	VP	FN

## Tipos de error al contrastar hipótesis

	Realidad	
	$H_0$ Falsa $A \neq B$	$H_0$ Cierta $A=B$
Rechazo $H_0$ Acepto $H_1$ $A \neq B$ Encuentro pruebas (+) $p < 0,05$	Correcto Se rechaza la hipótesis nula ( $H_0$ ) (se dice que los resultados son estadísticamente significativos y es verdad.	Error de tipo I Se rechaza la $H_0$ (se dice que los resultados son estadísticamente significativos) pero nos equivocamos. Probabilidad $\alpha$
	VP	FP
No Rechazo $H_0$ Acepto $H_0$ $A=B$ No encuentro pruebas (-) $p > 0,05$	Error de tipo II No se rechaza la hipótesis nula ( $H_0$ ) (se dice que no hay diferencias significativas) pero nos equivocamos. Probabilidad $\beta$	Correcto No se rechaza la hipótesis nula $H_0$ (el azar puede explicar las diferencias observadas en los datos), y es verdad.
	FN	VN
	Error BETA	
		Error ALFA

## Tipos de error al contrastar hipótesis

	Realidad	
	$H_a$	$H_0$
$H_a$ Encuentro pruebas (+) $p < 0,05$	Acierto $(1 - \beta)$ <b>POTENCIA ESTADÍSTICA</b>	Error de tipo I $\alpha$ FP
	VP	Error ALFA
$H_0$ No encuentro pruebas (-) $p > 0,05$	Error de tipo II $\beta$ FN	Acierto $(1 - \alpha)$ VN
	Error BETA	

**POTENCIA ESTADÍSTICA:** Es la capacidad del estudio para detectar una diferencia si ésta existe realmente.

## Tipos de error al contrastar hipótesis

	Realidad	
	$H_a$	$H_0$
$H_a$ Encuentro pruebas (+) $p < 0,05$	Acierto $(1 - \beta)$ POTENCIA ESTADÍSTICA	Error de tipo I $\alpha$ FP
	VP	Error ALFA
$H_0$ No encuentro pruebas (-) $p > 0,05$	Error de tipo II $\beta$ FN	Acierto $(1 - \alpha)$ VN
	Error BETA	

**NIVEL DE CONFIANZA ( $1 - \alpha$ ):** Es la probabilidad a priori de que el intervalo de confianza (IC) a calcular contenga al verdadero valor del parámetro

## El error tipo I (error alfa)

- En nuestro estudio se concluye diciendo que hay diferencias cuando realmente no las hay.
- Se detecta significación estadística ( $p < 0,05$ ) y se rechaza  $H_0$ , cuando en realidad ésta es verdadera.
- La probabilidad de cometer este error es lo que mide la “p”.

FALSOS POSITIVOS

## El error tipo II (error beta)

- Cuando los resultados de nuestro estudio NO son significativos ( $p > 0,05$ ) se concluye diciendo que no hay diferencias (“estudio negativo”) cuando realmente SI las hay, pero nosotros no hemos sido capaces de detectarlas...
  - La probabilidad de tomar esta decisión (y aceptar la  $H_0$ ) siendo incorrecta se llama error beta.
- ¡En realidad existen diferencias pero el estudio ha sido incapaz de detectarlas!
  - Son los “falsos negativos”.

FALSOS NEGATIVOS

## Recomendaciones para disminuir el error alfa ( $\alpha$ ) tipo I

- Disponer de una teoría que guíe la investigación, evitando el “salir de pesca” con el ordenador buscando asociaciones entre variables.
- Disminuir el número de test estadísticos llevados a cabo en el estudio.
- Depurar la base de datos para evitar errores de valores extremos que puedan producir hallazgos significativos.
- Utilizar valores de alfa más reducidos (0,01 ó 0,001).
- Reproducir el estudio. Si al reproducir el estudio se obtienen resultados similares, estaremos más seguros de no estar cometiendo el error de tipo I.

## Recomendaciones para disminuir el error beta ( $\beta$ ) tipo II

- Incrementar el tamaño de la muestra.
- Estimar el poder estadístico del estudio ( $1 - \beta$ ).
- Incrementar el tamaño del efecto a detectar.
- Incrementar el valor de alfa.
- Utilizar test paramétricos (más potentes) en lugar de test no paramétricos

## Importancia del error alfa y error beta

- Determinación del tamaño muestral
  - Estudios para determinar parámetros poblacionales (sólo error alfa)
    - Estimación de una proporción.
    - Estimación de una media.
  - Estudios para contraste de hipótesis (error alfa y beta).
    - Comparación de dos proporciones.
    - Comparación de dos medias.
    - Correlación entre dos variables cuantitativas.
  
- Rentabilidad diagnóstico de una prueba-test.
  - Sensibilidad.
  - Especificidad.
  - Valor predictivo positivo (VPP).
  - Valor predictivo negativo (VPN).

# Cálculo del tamaño muestral de un parámetro

## 1. VARIABILIDAD del parámetro que se desea estimar

- Variables cuantitativas: varianza
- Variables cualitativas: 1-P
  1. Variabilidad del parámetro  
Cuando esta variabilidad es desconocida
    - Bibliografía
    - Datos propios previos
    - Estudio piloto

2. PRECISIÓN ( $i$ ) con la que se desea obtener la estimación (amplitud deseada del IC). Se fija  $i$  previo al estudio. Cuanto más precisión más estrecho será el IC, y mayor deberá ser el volumen muestral.

### 2. Precisión

3. NIVEL DE CONFIANZA deseado (generalmente 95%), que corresponde a un valor de alfa de 0,05. Este valor indica el grado de confianza de que el verdadero valor poblacional  $\mu$  desconocido del parámetro en estudio se sitúe en el intervalo obtenido.

### 3. Nivel de confianza 95%

Fórmulas para el cálculo del número de sujetos necesarios para la realización de un estudio cuyo objetivo es la estimación de un parámetro

Estimación de una variable CUALITATIVA (proporción - prevalencia)

$$N = \frac{Z^2_{\alpha} * P * (1-P)}{i^2}$$

Estimación de una variable CUANTITATIVA (MEDIA)

$$N = \frac{Z^2_{\alpha} * s^2}{i^2}$$

N = número de sujetos necesarios

$Z^2_{\alpha}$  = Valor del estadístico Z correspondiente al riesgo alfa fijado ( $\alpha = 0,05$ )

P = Valor de la proporción (prevalencia) que se supone que existe en la población

$S^2$  = Varianza de la distribución de la variable cuantitativa que se supone que existe en la población

i: Precisión con que se desea estimar el parámetro

Influencia del valor supuesto de la proporción (P) o prevalencia (variabilidad del Parámetro que se desea estimar sobre el número de sujetos necesarios N (tamaño muestral mínimo necesario).

### INFLUENCIA DE LA PROPORCIÓN

1. Variabilidad del parámetro	2. Precisión (i = 0,05)	3. Nivel de confianza 95% $\alpha = 0,05$
Prevalencia	Precisión (i = cte)	Tamaño muestral (N)
0,10	0,05 (5%)	138
0,20	0,05 (5%)	246
0,30	0,05 (5%)	323
0,40	0,05 (5%)	369
0,50	0,05 (5%)	384

Postura de máxima indeterminación

$$i = Z_{\alpha} \sqrt{P(1-P)/N}$$

Todos los cálculos han sido realizados con un nivel de confianza del 95%

P = valor supuesto de la proporción que se desea estimar.

i = precisión de la estimación, que corresponde a la mitad del intervalo de confianza ( $P \pm i$ )

N = número de sujetos necesarios para realizar la estimación de P con la precisión deseada y el 95% de confianza

Influencia de la precisión (i) de la estimación sobre el número de sujetos necesarios N (tamaño muestral mínimo necesario).

### INFLUENCIA DE LA PRECISIÓN

1. Variabilidad del parámetro (P=cte)	2. Precisión (i = 0,05)	3. Nivel de confianza 95% $\alpha = 0,05$
Prevalencia	Precisión (i = cte)	Tamaño muestral (N)
0,40	0,1 (10%)	92
0,40	0,075 (7,5%)	164
0,40	0,050 (5%)	369
0,40	0,025 (2,5%)	1475
0,40	0,01 (1%)	9220

$$i = Z_{\alpha} \sqrt{P(1-P)/N}$$

Todos los cálculos han sido realizados con un nivel de confianza del 95% ( $\alpha = 0,05$ )

P = valor supuesto de la proporción que se desea estimar.

i = precisión de la estimación, que corresponde a la mitad del intervalo de confianza ( $P \pm i$ )

N = número de sujetos necesarios para realizar la estimación de P con la precisión deseada y el 95% de confianza

The screenshot shows the STATCALC software window. The title bar reads "STATCALC". The menu bar includes "Auto" and several icons. The main window has a blue background with white text. At the top, it says "EpiInfo Version 6", "Statcalc", and "November 1993". Below this, it reads "Population Survey or Descriptive Study Using Random (Not Cluster) Sampling". The main question is "¿Cuál es el tamaño muestral necesario para conocer la prevalencia de inmunización toxoplásmica en la ciudad de Ceuta?". The input fields are: "Size of population from which the sample will be selected : 70000", "Expected frequency of the factor under study (err toward 50%) : 50.00 %", and "Worst acceptable result : 55.00 %". A callout box on the right says "Postura de máxima indeterminación" and "Precisión 5% i = 0,05". At the bottom, a white box contains the text: "Tamaño muestral necesario (n) para conocer la prevalencia de inmunización frente a Toxoplasma gondii en la población de la ciudad de Ceuta". The taskbar at the bottom shows "Procesador de textos", "EPI INFO 2000", "STATCALC", and "Página WEB de EpiInfo". The system tray shows the time "15:09".

STATCALC

Auto

EpiInfo Version 6 Statcalc November 1993

Population Survey or Descriptive Study Using Random (Not Cluster) Sampling

¿Cuál es el tamaño muestral necesario para conocer la prevalencia de inmunización toxoplásmica en la ciudad de Ceuta?

Size of population from which the sample will be selected : 70000

Expected frequency of the factor under study (err toward 50%) : 50.00 % Postura de máxima indeterminación

If 50.00 % is the true rate in the population, what is the result farthest from the rate that you would accept in your sample (higher or lower)?

Worst acceptable result : 55.00 % Precisión 5% i = 0,05

Tamaño muestral necesario (n) para conocer la prevalencia de inmunización frente a Toxoplasma gondii en la población de la ciudad de Ceuta

Procesador de textos EPI INFO 2000 STATCALC Página WEB de EpiInfo 15:09

STATCALC

Auto

EpiInfo Version 6 Statcalc November 1993

Population Survey or Descriptive Study Using Random (Not Cluster) Sampling

Variabilidad

Precisión 5%  
 $i = 0,05$

Nivel de confianza  
95%  
 $\alpha = 0,05$

Population Size : 70,000

Expected Frequency : 50.00 %

Worst Acceptable : 55.00 %

Confidence Level Sample Size

Confidence Level	Sample Size
80 %	164
90 %	270
95 %	382
99 %	657
99.9 %	1,066
99.99 %	1,482

$i = 0,05$

Precisión 5%

Change value of Population, Frequency, or Worst Acceptable to recalculate.

Tamaño muestral necesario (n) para conocer la prevalencia de inmunización frente a *Toxoplasma gondii* en la población de la ciudad de Ceuta

INICIO EPI INFO 2000 STATCALC 15:55

STATCALC

Auto

EpiInfo Version 6 Statcalc November 1993

Population Survey or Descriptive Study Using Random (Not Cluster) Sampling

¿Cuál es el tamaño muestral necesario para conocer la prevalencia de rasgo beta-talasémico en la ciudad de Ceuta?

Size of population from which the sample will be selected : 70000

Expected frequency of the factor under study (err toward 50%) : 1.00 %  $i = 0,005$   
Precisión 0,5%

If 1.00 % is the true rate in the population, what is the result farthest from the rate that you would accept in your sample (higher or lower)?

Worst acceptable result : 1.50 %

F1-Help F4-Calc F6-Open File F10-Done

Editor de gráficos

Procesador de textos

Página WEB de EpiInfo

Version 1.0.5

INICIO EPI... STA... MICR... STA... 16:07



STATCALC

Auto

EpiInfo Version 6 Statcalc November 1993

Population Survey or Descriptive Study Using Random (Not Cluster) Sampling

Variabilidad

Precisión  $i = 0,5\%$

Nivel de confianza  
95%  
 $\alpha = 0,05$

Population Size : 70,000

Expected Frequency : 1.00 %  $i = 0,005$

Worst Acceptable : 1.50 % Precisión 0,5%

Confidence Level	Sample Size
80 %	644
90 %	1,055
95 %	1,489
99 %	2,532
99.9 %	4,040
99.99 %	5,521

Change value of Population, Frequency, or Worst Acceptable to recalculate.

F1-Help F5-Print F6-Open File F10-Done

Sexo

Haga clic para agregar notas

DIBUJO

Diapositiva 82 de 252 Diseño predeterminado

INICIO EPI I... STA... MICR... STA... 16:07

## ¿Cuántas personas tenemos que estudiar para conocer la prevalencia de diabetes tipo 2 en el distrito de salud X de Ceuta que atiende a una población de 15.000 habitantes?

1. Proporción esperada: asumamos que puede ser próxima al 5%
2. Precisión (i)(d): 3%
3. Nivel del confianza (seguridad): 95%

$$n = \frac{Z_{\alpha}^2 * p * q}{d^2}$$

- $Z_{\alpha}^2 = 1.96^2$  (ya que la seguridad es del 95%)
- $p =$  proporción esperada (en este caso 5% = 0.05)
- $q = 1 - p$  (en este caso  $1 - 0.05 = 0.95$ )
- $d =$  precisión (en este caso deseamos un 3%)

$$n = \frac{1.96^2 * 0.05 * 0.95}{0.03^2} = 203 \quad n = 203 \text{ personas del distrito de salud X}$$

4. Tamaño de la población de origen (población atendida por el distrito de salud X)

$$n = \frac{N * Z_{\alpha}^2 * p * q}{d^2 * (N - 1) + Z_{\alpha}^2 * p * q} \quad n = \frac{15.000 * 1.96^2 * 0.05 * 0.95}{0.03^2 * (15.000 - 1) + 1.96^2 * 0.05 * 0.95} = 200$$

$n = 200$  personas del distrito de salud X

## Cálculo del tamaño muestral para contraste de hipótesis

1. Magnitud de la diferencia a detectar que tenga interés clínico.
  - Comparación de dos medias.
  - Comparación de dos proporciones:  $P_1$  y  $P_2$ .
2. Aproximación de los parámetros de la variable que se estudia (bibliografía o estudios previos).
3. NIVEL DE CONFIANZA deseado o seguridad del estudio (generalmente 95%), que corresponde a un valor de  $\alpha$  de 0,05 (riesgo de cometer un error alfa).
4. PODER ESTADÍSTICO ( $1 - \beta$ )(riesgo de cometer un error  $\beta$ ).
5. Definir si la hipótesis va a ser unilateral o bilateral.
  - Bilateral: cualquiera de los dos parámetros a comparar puede ser mayor o menor que el otro (no hay dirección) ( $A \neq B$ ).
  - Unilateral: cuando se considera que uno debe ser mayor que el otro (dirección de las diferencias) ( $A < B$ ).

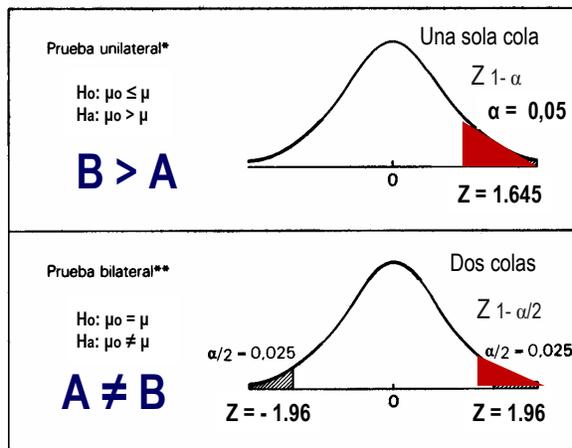
Prueba UNILATERAL: Se utiliza cuando se pretende determinar si el fármaco nuevo B es más eficaz que el clásico A, es decir, tan sólo interesa una de las direcciones de la comparación ( $B > A$ ), interesa demostrar que B sea más efectivo que A.

## DISTRIBUCIÓN NORMAL TIPIFICADA (Z)

Valores de la distribución normal tipificada (Z) correspondientes al valor  $\alpha$  del 5% en las pruebas estadísticas unilaterales y bilaterales.

Las pruebas unilaterales especifica la dirección de la diferencia.  
Las pruebas bilaterales NO especifica la dirección de la diferencia.

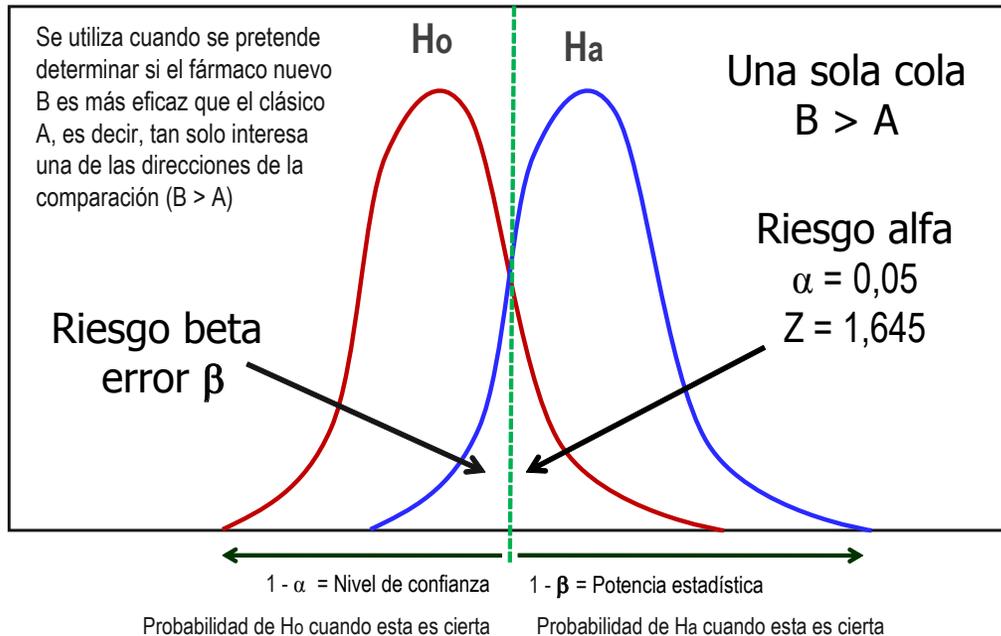
tribuye simétricamente en las dos colas de la distribución.  
Ho: hipótesis nula; Ha: hipótesis alternativa;  $\mu$ : valor de la diferencia en la población;  $\mu_o$ : valor de la diferencia observada en el estudio.



Prueba BILATERAL: Se utiliza cuando se pretende determinar si dos fármacos A y B difieren en su efectividad clínica ( $A \neq B$ ).

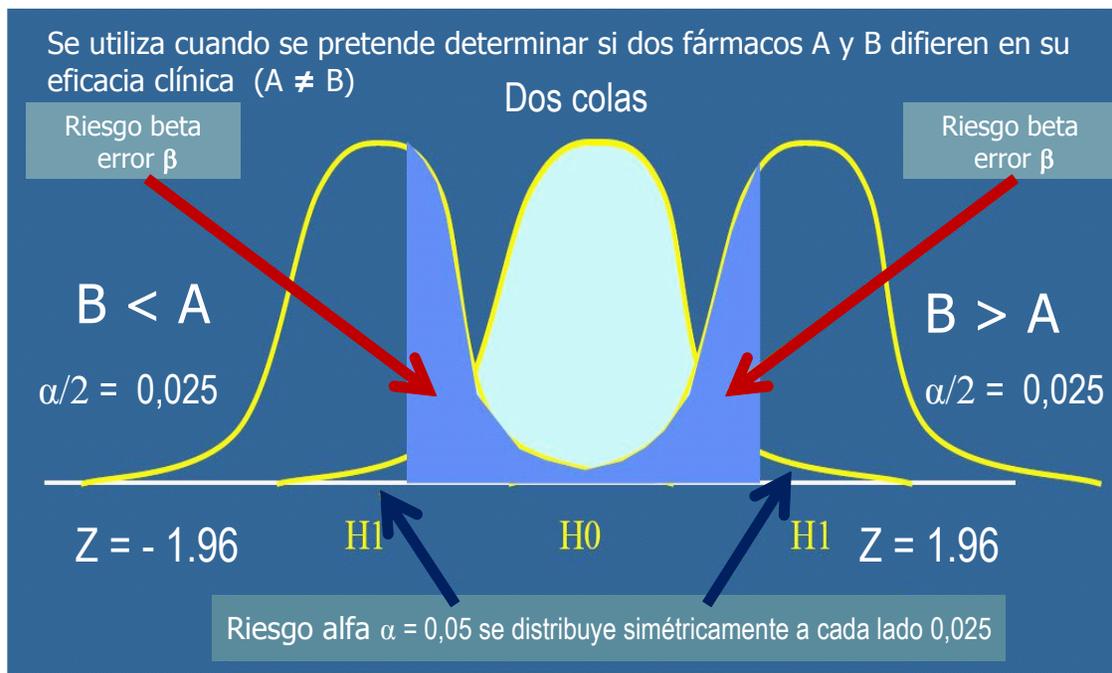
## Pruebas de contraste unilateral:

### Riesgo en la toma de decisiones



## Pruebas de contraste bilateral:

### Riesgo en la toma de decisiones



## Ejemplo de comparación de cálculo del tamaño muestral para comparación de dos proporciones a una sola cola

Deseamos evaluar si el nuevo fármaco B es mejor que el tratamiento clásico A para el alivio del dolor, para lo que diseñamos un ensayo clínico. Sabemos por datos previos (bibliografía) que la eficacia del fármaco A es próxima al 70% y consideramos clínicamente relevante si el nuevo fármaco B alivia el dolor en un 90%. Nuestro nivel de riesgo alfa lo fijamos en 0.05 (error  $\alpha = 5\%$ ) y deseamos un poder estadístico ( $1 - \beta$ ) de un 80% (error  $\beta = 20\%$ ). ¿Cuál es el volumen muestral?

$$n = \frac{\left[ Z_{\alpha} * \sqrt{2p(1-p)} + Z_{\beta} * \sqrt{p_1(1-p_1) + p_2(1-p_2)} \right]^2}{(p_1 - p_2)^2} \quad \begin{array}{l} \mathbf{B > A} \quad \text{Una sola cola} \\ Z = 1.645 \end{array}$$

$$n = \frac{\left[ 1.645 * \sqrt{2 * 0.8(1-0.8)} + 0.842 * \sqrt{0.7(1-0.7) + 0.9(1-0.9)} \right]^2}{(0.7 - 0.9)^2} \quad p = \frac{0.7 + 0.9}{2} = 0.8$$

$n = 48$  pacientes

En cada grupo (A y B) precisamos 48 pacientes.

## Ejemplo de comparación de cálculo del tamaño muestral para comparación de dos proporciones a dos colas.

Sabemos por datos previos (bibliografía) que la mortalidad del fármaco A es próxima al 31% comparada con el fármaco B que tiene una tasa de mortalidad del 40%. Nuestro nivel de riesgo alfa lo fijamos en 0.05 (error  $\alpha = 5\%$ ) y deseamos un poder estadístico ( $1 - \beta$ ) de un 90% (error  $\beta = 10\%$ ). ¿Cuál es el volumen muestral?

### Cálculo del tamaño de la muestra por fórmula:

$$n = 2x \left( \frac{z_{(1-\alpha/2)}x\sqrt{[2\pi_o(1-\pi_o)]} + z_{(1-\beta)}x\sqrt{\pi_1(1-\pi_1) + \pi_2(1-\pi_2)}}{\pi_1 - \pi_2} \right)^2$$

Donde :

$\pi_1$  es la proporción anticipada del evento en el grupo 1 (0.31)

$\pi_2$  es la proporción anticipada del evento en el grupo 2 (0.40)

Dos colas

B  $\neq$  A

$$\pi_o = \left( \frac{\pi_1 + \pi_2}{2} \right) = 0.355$$

$$Z = 1.96$$

$Z_{(1-\alpha/2)}$  es 1.96 si se emplea la *prueba de significancia*  $p = 0.05$  de dos colas

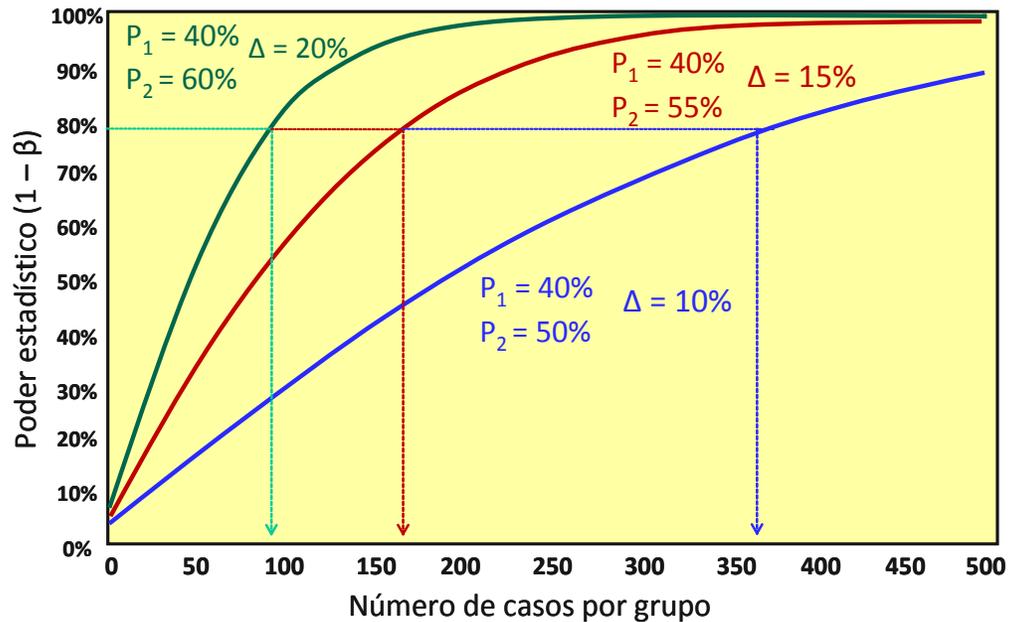
$Z_{(1-\beta)}$  es 0.84 para un poder de 80% y es 1.28 para un poder de 90%

$$n = 2x \left( \frac{1.96\sqrt{[2x0.355(1-0.355)]} + 1.28\sqrt{[0.31(1-0.31) + 0.40(1-0.40)]}}{0.31 - 0.40} \right)^2 = 1184$$

El tamaño muestral de cada grupo entonces será de 592 pacientes (1184÷2)

Necesitamos un total de 1184 pacientes, 592 para cada grupo (A y B).

Influencia de la magnitud del efecto a detectar y del poder estadístico en la determinación del tamaño muestral ( $n_1$  y  $n_2$ ).  
Comparación de dos Prevalencias (proporciones)  $P_1$  y  $P_2$ .



## Cálculo del tamaño de la muestra por software (ejemplo)

El cálculo resulta aún más fácil, cuando se utiliza un software. Por ejemplo utilizando el software gratuito PS: Power and Sample Size Program, tenemos que introducir los valores correspondientes a:

1. Las tasa de mortalidad en los grupos que recibieron el tratamiento A y B (0.31 y 0.40).
2. El nivel de significación estadística (ERROR ALFA del 5%) ( $p < 0.05$ ) .
3. El poder estadístico del estudio (ERROR BETA del 10%) (0.90) asumido para obtener el tamaño muestral por cada grupo.

**Power and Sample Size Program: Main Window**

File Log Help

Survival t-test Regression 1 Regression 2 **Dichotomous** Log

Studies that are analysed by chi-square or Fisher's exact test

**Output**

What do you want to know? Sample size

Case sample size for uncorrected chi-squared test 592

---

**Design**

Matched or Independent? Independent

Case control? Prospective

How is the alternative hypothesis expressed? Two proportions

Uncorrected chi-square or Fisher's exact test? Uncorrected chi-square test

---

**Input**

$\alpha$	0.05	$p_0$	0.31
<u>power</u>	0.90	$p_1$	0.40
		$m$	1

Calculate

Graphs

## Ejemplo del cálculo del tamaño muestral para el cálculo del coeficiente de correlación entre dos variables cuantitativas

Supongamos que se desea estudiar la asociación entre la edad y el nivel de colesterol entre los pacientes de un centro de salud. Se diseña un estudio piloto en el que se determinará mediante una analítica el colesterol en una muestra aleatoria de los pacientes durante un periodo de tiempo prefijado, de los que también se registrará su edad. Se obtiene un valor del coeficiente de correlación lineal de Pearson entre la edad y el colesterol de  $r=0,3$ . Aplicando las fórmulas, con un NIVEL DE CONFIANZA del 95% (error  $\alpha=5\%$ ) y un PODER ESTADÍSTICO del 80% (error  $\beta=20\%$ ), se obtiene:

$$n = \left( \frac{z_{1-\alpha/2} + z_{1-\beta}}{\frac{1}{2} \ln \left( \frac{1+r}{1-r} \right)} \right)^2 + 3 = \left( \frac{1,96 + 0,84}{\frac{1}{2} \ln \left( \frac{1+0,3}{1-0,3} \right)} \right)^2 + 3 \approx 85$$

En este caso, se necesitaría incluir a 85 pacientes para llevar a cabo el estudio.

Si, además, en este último caso se prevé un 20% de posibles pérdidas de información durante la ejecución del estudio, el tamaño muestral se recalcula según la expresión:

$$n' = \frac{n}{1-L} = \frac{85}{1-0,2} = 106,25 \approx 107$$

Es decir, se necesitaría una muestra de 107 pacientes para llevar a cabo la investigación.

Pértegas Díaz S , Pita Fernández S. Determinación del tamaño muestral para calcular la significación del coeficiente de correlación lineal. Cad Aten Primaria 2001; 9: 209-211.

Fórmulas para el cálculo del poder estadístico ( $1-\beta$ ) en función del tipo de diseño del estudio.

Cálculo del poder estadístico de un estudio.

Pértegas Díaz S, Pita Fernández S. Unidad de Epidemiología Clínica y Bioestadística. Complejo Hospitalario-Universitario Juan Canalejo. La Coruña

	Test unilateral	Test bilateral
Comparación de dos proporciones	$z_{1-\beta} = \frac{ p_1 - p_2  \sqrt{n} - z_{1-\alpha} \sqrt{2p(1-p)}}{\sqrt{p_1(1-p_1) + p_2(1-p_2)}}$	$z_{1-\beta} = \frac{ p_1 - p_2  \sqrt{n} - z_{1-\alpha/2} \sqrt{2p(1-p)}}{\sqrt{p_1(1-p_1) + p_2(1-p_2)}}$
Comparación de dos media	$z_{1-\beta} = \sqrt{\frac{n}{2}} \frac{d}{S} - z_{1-\alpha}$	$z_{1-\beta} = \sqrt{\frac{n}{2}} \frac{d}{S} - z_{1-\alpha/2}$
Estimación de un OR en estudios de casos y controles	$p_1 = \frac{OR \cdot p_2}{(1-p_2) + OR \cdot p_2}$ $z_{1-\beta} = \frac{ p_1 - p_2  \sqrt{nc} - z_{1-\alpha} \sqrt{(c+1)p(1-p)}}{\sqrt{c \cdot p_1(1-p_1) + p_2(1-p_2)}}$ $m = c \cdot n$	$p_1 = \frac{OR \cdot p_2}{(1-p_2) + OR \cdot p_2}$ $z_{1-\beta} = \frac{ p_1 - p_2  \sqrt{nc} - z_{1-\alpha/2} \sqrt{(c+1)p(1-p)}}{\sqrt{c \cdot p_1(1-p_1) + p_2(1-p_2)}}$ $m = c \cdot n$
Estimación de un RR	$p_1 = RR \cdot p_2$ $z_{1-\beta} = \frac{ p_1 - p_2  \sqrt{n} - z_{1-\alpha} \sqrt{2p(1-p)}}{\sqrt{p_1(1-p_1) + p_2(1-p_2)}}$	$p_1 = RR \cdot p_2$ $z_{1-\beta} = \frac{ p_1 - p_2  \sqrt{n} - z_{1-\alpha/2} \sqrt{2p(1-p)}}{\sqrt{p_1(1-p_1) + p_2(1-p_2)}}$
Estimación de un coeficiente de correlación lineal	$z_{1-\beta} = \sqrt{n-3} \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right) - z_{1-\alpha}$	$z_{1-\beta} = \sqrt{n-3} \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right) - z_{1-\alpha/2}$

- n = Tamaño muestral. En un estudio de casos y controles, n es el número de casos.
- $p_1$  = En un estudio transversal o de cohortes, proporción de expuestos que desarrollan la enfermedad. En un estudio de casos y controles, proporción de casos expuestos.
- $p_2$  = En un estudio transversal o de cohortes, proporción de no expuestos que desarrollan la enfermedad. En un estudio de casos y controles, proporción de controles expuestos.
- $$p = \frac{p_1 + p_2}{2}$$
- d = Valor mínimo de la diferencia a detectar entre dos medias
- $S^2$  = Varianza en el grupo control o de referencia

## Fórmulas para el cálculo del poder estadístico ( $1-\beta$ ) en función del tipo de diseño del estudio.

	test unilateral	test bilateral
estimación de un OR en estudios de casos y controles	$p_1 = \frac{OR \cdot p_2}{(1-p_2) + OR \cdot p_2}$ $z_{1-\beta} = \frac{ p_1 - p_2  \sqrt{nc} - z_{1-\alpha} \sqrt{(c+1)p(1-p)}}{\sqrt{c p_1(1-p_1) + p_2(1-p_2)}}$ $m = c \times n$	$p_1 = \frac{OR \cdot p_2}{(1-p_2) + OR \cdot p_2}$ $z_{1-\beta} = \frac{ p_1 - p_2  \sqrt{nc} - z_{1-\frac{\alpha}{2}} \sqrt{(c+1)p(1-p)}}{\sqrt{c p_1(1-p_1) + p_2(1-p_2)}}$ $m = c \times n$
estimación de un RR	$p_1 = RR \times p_2$ $z_{1-\beta} = \frac{ p_1 - p_2  \sqrt{n} - z_{1-\alpha} \sqrt{2p(1-p)}}{\sqrt{p_1(1-p_1) + p_2(1-p_2)}}$	$p_1 = RR \times p_2$ $z_{1-\beta} = \frac{ p_1 - p_2  \sqrt{n} - z_{1-\frac{\alpha}{2}} \sqrt{2p(1-p)}}{\sqrt{p_1(1-p_1) + p_2(1-p_2)}}$
estimación de un coeficiente de correlación lineal	$z_{1-\beta} = \sqrt{n-3} \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right) - z_{1-\alpha}$	$z_{1-\beta} = \sqrt{n-3} \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right) - z_{1-\frac{\alpha}{2}}$

### Cálculo del poder estadístico de un estudio.

Pértegas Díaz S, Pita Fernández S.

Unidad de Epidemiología Clínica y Bioestadística. Complejo Hospitalario-Universitario Juan Canalejo. La Coruña

$n$  = Tamaño muestral. En un estudio de casos y controles,  $n$  es el número de casos.

$P_1$  = En un estudio transversal o de cohortes, proporción de expuestos que desarrollan la enfermedad. En un estudio de casos y controles, proporción de casos expuestos

$P_2$  = En un estudio transversal o de cohortes, proporción de NO expuestos que desarrollan la enfermedad. En un estudio de casos y controles, proporción de controles expuestos

$d$  = Valor mínimo de la diferencia a detectar entre dos medias

$s^2$  = Varianza en el grupo control o de referencia

$c$  = Número de controles por caso, en estudios caso-control

$m$  = En un estudio de casos y controles, número de controles

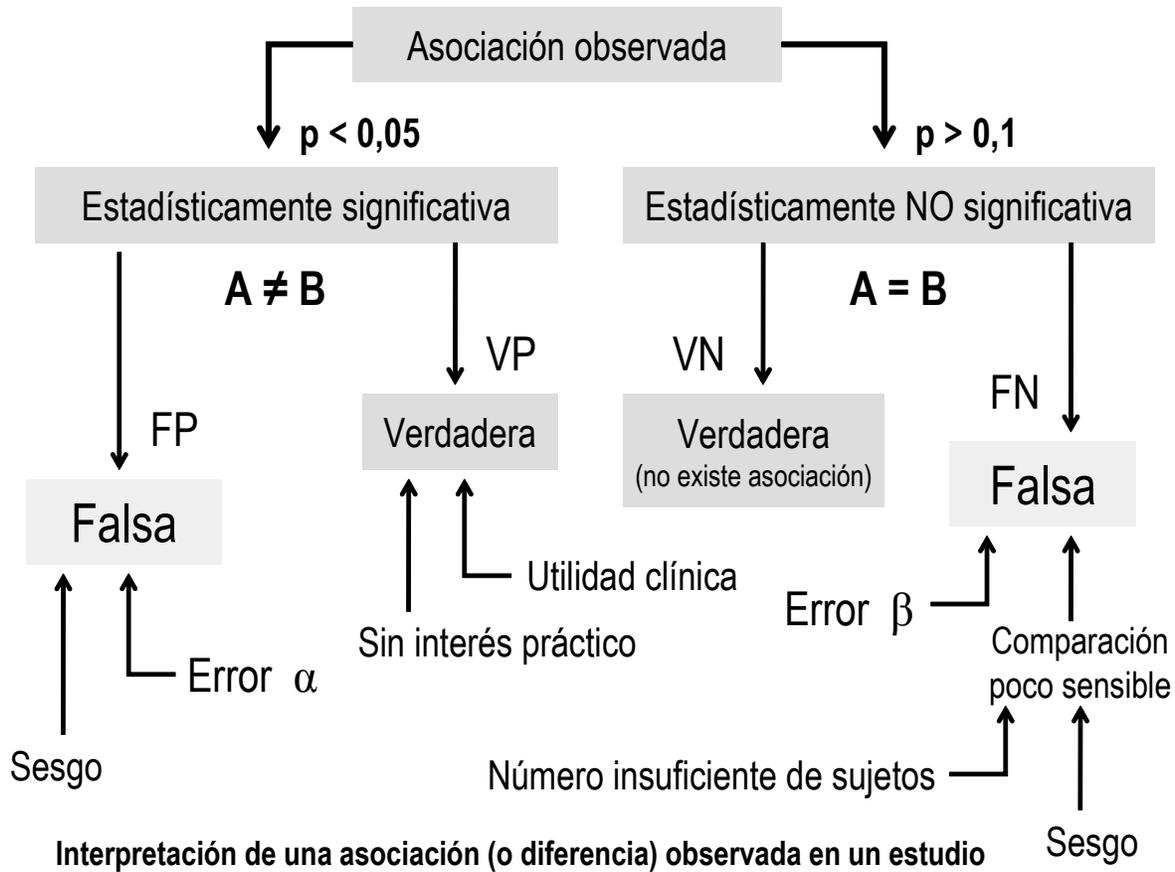
OR = Valor aproximado del *odds ratio* a detectar

RR = Valor aproximado del riesgo relativo a detectar

$r$  = Magnitud del coeficiente de correlación a detectar

Valores de los estadísticos Z más frecuentemente utilizados en el cálculo del tamaño muestral considerando el error  $\alpha$  y  $\beta$

Error $\alpha$	Seguridad (1 - $\alpha$ )	Test unilateral Z 1- $\alpha$	Test bilateral Z 1- $\alpha/2$
0,05	95%	1,645	1,96
0,01	99%	2,326	2,576
Error $\beta$	Poder estadístico (1 - $\beta$ )	$\beta$	Z 1- $\beta$
0,10	90%	0,10	1,282
0,20	80%	0,20	0,842



## Interpretación de una asociación (o diferencia) observada en un estudio ERROR ALFA $\alpha$

### Expedición de pesca

La probabilidad de encontrar resultados FALSOS POSITIVOS FP (error alfa  $\alpha$ ) aumenta notablemente con el número de comparaciones injustificadas realizadas entre las distintas variables durante el análisis (“ir de pesca”).

Estudios que sin apoyarse en ninguna hipótesis a PRIORI (estrategia de análisis), consideran todas las variables de igual importancia (hipótesis a POSTERIORI), analizándolas unos contra otras (sin criterio), esperando a ver que se encuentra...

## Decisiones que se pueden tomar en un test de contraste de hipótesis después de conocer el valor de significación estadística $p$

$p < 0,05$ $A \neq B$	$p \geq 0,10$ $A = B$
Se rechaza la hipótesis nula	No se puede rechazar la hipótesis nula
No parece que el azar pueda explicarlo todo	No se puede descartar que el azar lo explique todo
El EFECTO observado es mayor que el ERROR	El EFECTO observado es similar que el ERROR
Hay diferencias estadísticamente significativas	NO hay diferencias estadísticamente significativas
Existen evidencias a favor de la hipótesis alternativa $H_a$ , y por tanto rechazamos la hipótesis nula $H_0$	NO existen evidencias a favor de la hipótesis alternativa $H_a$ , y por tanto aceptamos la hipótesis nula $H_0$

Los límites de 0,05 y 0,1 pueden ser considerados en cierto modo arbitrarios y aproximados

# Materiales

## Pacientes

1. *Criterios de selección de la muestra (inclusión y exclusión).*
2. Procedencia.
3. Técnica de reclutamiento utilizada.
4. Tiempo empleado para reunir a todos los participantes.
5. Principales características de los pacientes.
  - Edad.
  - Sexo.
  - Procedencia.
  - Gravedad.
  - Años de evolución de la enfermedad.
  - Patologías asociadas...
6. *Cálculo de tamaño muestral.*

## Capítulo 4

# Publicación Bioestadística: Error $\alpha$ y $\beta$ . Evaluación de pruebas diagnósticas.

Dr. Jacobo Díaz Portillo  
Coordinador de Formación Continuada  
Hospital Universitario de Ceuta

## Bioestadística: Error $\alpha$ y $\beta$

Influencia de los errores  $\alpha$  y  $\beta$  en la rentabilidad clínica de las pruebas diagnósticas

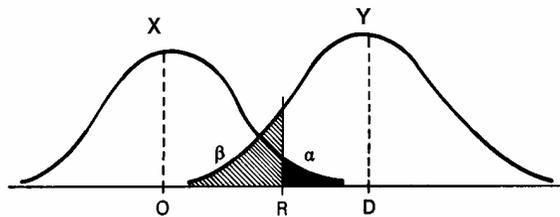
Evaluación de las pruebas diagnósticas

## Evaluación de las pruebas diagnósticas

- Exactitud (validez, efectividad o rentabilidad).
  - Sensibilidad.
  - Especificidad.
- Seguridad: valores predictivos.
  - Valor predictivo positivo (VPP).
  - Valor predictivo negativo (VPN).
- Razón de verosimilitud.
- Curvas roc.
- Reproductibilidad.
- Invasividad.
- Coste económico (URV).

Hipótesis nula ( $H_0$ ): No existen diferencias significativas en la eficacia clínica de dos fármacos A y B ( $A = B$ )

Hipótesis alternativa ( $H_a$ ): Existen diferencias en la eficacia clínica de dos fármacos A y B ( $A \neq B$ )  $A > B$  ó  $A < B$



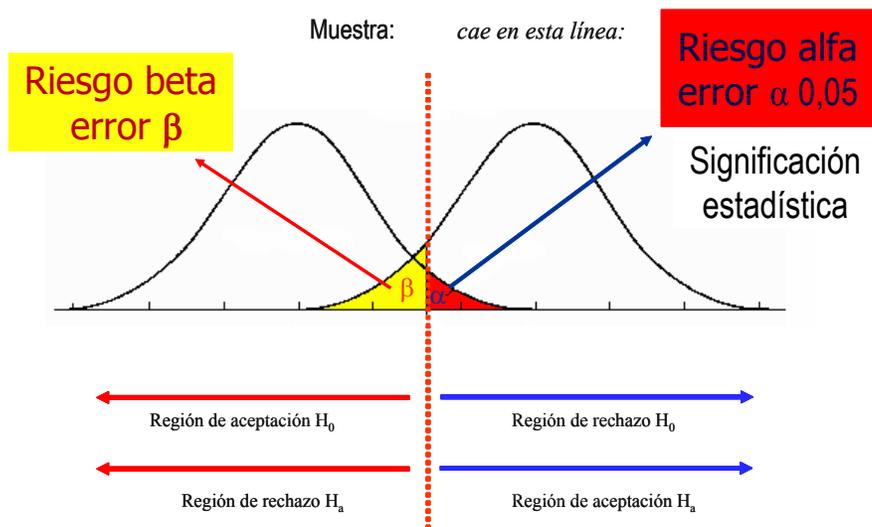
Se produce un cierto solapamiento en las distribuciones de probabilidad de la  $H_0$  y de la  $H_a$ , en este solapamiento siempre aparece el error  $\alpha$  y  $\beta$  asumido, el cual será mayor o menor en función del resultado o diferencia observada en nuestro estudio

		Realidad	
		Existe diferencia ( $H_a$ verdadera)	No existe diferencia ( $H_0$ verdadera)
Resultado de la prueba estadística	$p < 0,05$ Acepto la $H_a$	No error VP ( $1-\beta$ )	Error tipo I FP ( $\alpha$ )
	$p > 0,05$ Acepto la $H_0$	Error tipo II FN ( $\beta$ )	No error VN

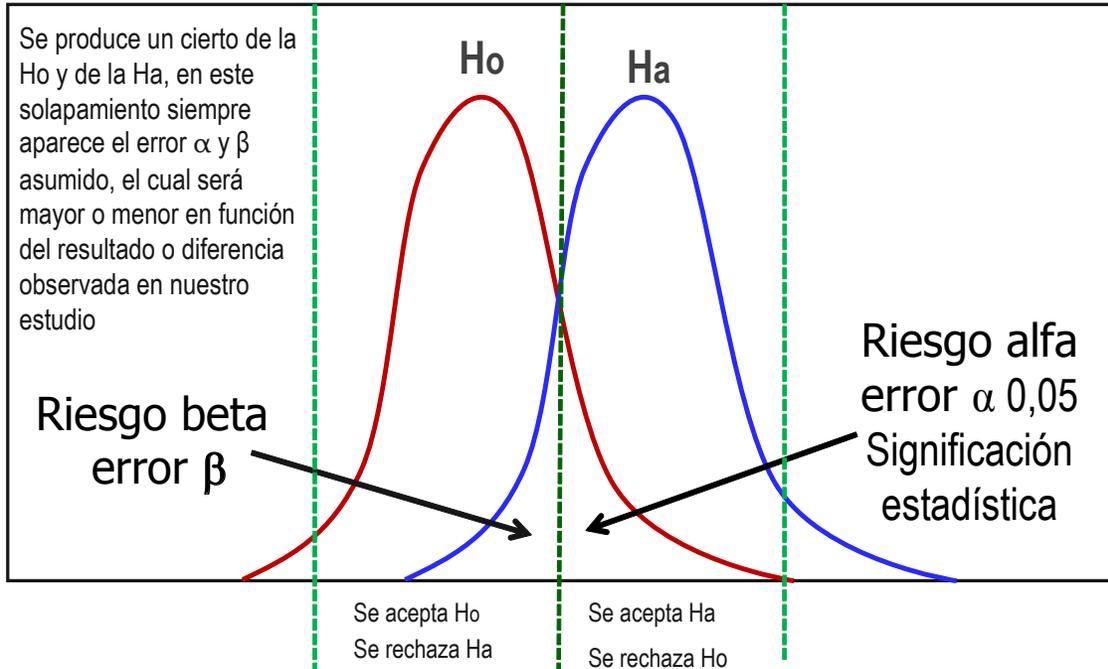
Resultados de una prueba de significación estadística  
 $H_0$ : hipótesis nula (no hay diferencias significativas)  
 X: distribución de la diferencia cuando  $H_0$  es cierta  
 Y: distribución de la diferencia cuando  $H_a$  es cierta  
 R: resultado (diferencia) observado en nuestro estudio

## Relación de los riesgos asociados a un test de hipótesis unilateral (estadístico valor de la media):

Error tipo I (riesgo  $\alpha$ ) y tipo II (riesgo  $\beta$ )

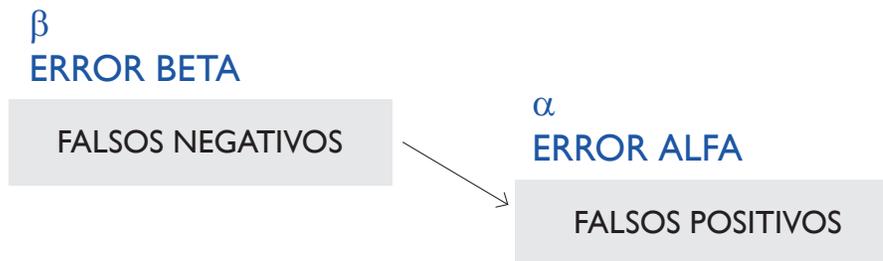


## Riesgo alfa y riesgo beta



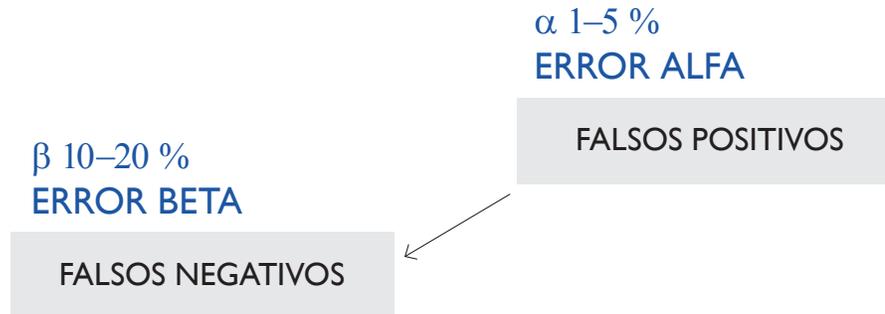
## Error alfa versus error beta:

No se puede tener todo



## Error alfa versus error beta:

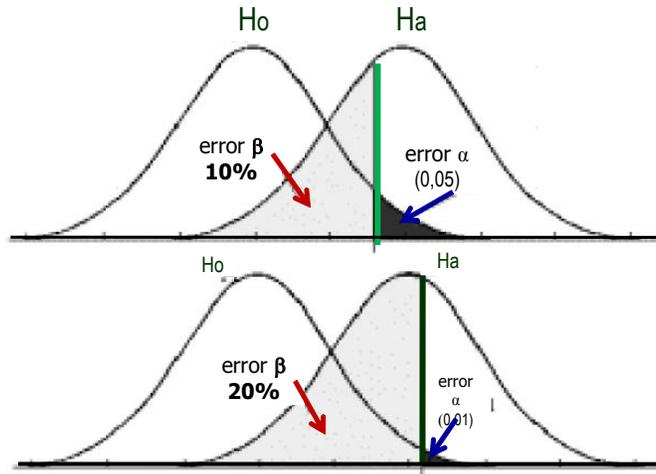
No se puede tener todo



## Error alfa versus error beta:

No se puede tener todo

Hay que buscar un equilibrio entre los FALSOS POSITIVOS (error  $\alpha$ ) y los FALSOS NEGATIVOS (error  $\beta$ ).



Para un tamaño muestral fijo ( $N = \text{cte}$ ), no se pueden reducir a la vez ambos tipos de error. Para reducir  $\beta$ , hay que aumentar el tamaño muestral  $N$ .

## Tipos de error en las pruebas diagnósticas

		Realidad	
		Enfermos	Sanos
Test de laboratorio	Positivo (+) Hipótesis alternativa	VERDADEROS POSITIVOS VP (1- $\beta$ )	FALSO POSITIVO error $\alpha$ Error Muy grave
	Negativo (-) Hipótesis nula	FALSO NEGATIVO error $\beta$ Error Menos grave	VERDADEROS NEGATIVOS VN

## Tipos de error en las pruebas diagnósticas

Tipos de error en las pruebas diagnósticas		Realidad	
		Enfermos	Sanos
Test de laboratorio PSA	Positivo (+)	VERDADEROS POSITIVOS VP	FALSO POSITIVO FP $\alpha$
	Negativo (-)	FALSO NEGATIVO FN $\beta$	VERDADEROS NEGATIVOS VN

## Tipos de error en las pruebas diagnósticas

Tipos de error en las pruebas diagnósticas		Evento o enfermedad estudiada		
		Enfermos	Sanos	Total
Resultados de la prueba diagnóstica	Positivo (+)	VERDADEROS POSITIVOS VP	FALSO POSITIVO FP $\alpha$	VP + FP
	Negativo (-)	FALSO NEGATIVO FN $\beta$	VERDADEROS NEGATIVOS VN	FN + VN
Total		VP + FN	FP + VN	



$$\text{SENSIBILIDAD} = \frac{VP}{VP + FN}$$

$$\text{SENSIBILIDAD} = \frac{\text{Enfermos positivos}}{\text{Total de ENFERMOS}}$$

## La sensibilidad de la prueba diagnóstica

- Es la probabilidad de clasificar correctamente a un individuo enfermo.
- Es la probabilidad de que para un sujeto enfermo se obtenga en la prueba un resultado positivo.
- *Es la capacidad del test para detectar la enfermedad.*
- Probabilidad de obtener un resultado positivo cuando el individuo tiene la enfermedad.
- Mide su capacidad para detectar la enfermedad cuando está presente.
- Su valor depende del error BETA ( $\beta$ ).

## Tipos de error en las pruebas diagnósticas

Tipos de error en las pruebas diagnósticas		Evento o enfermedad estudiada		
		Enfermos	Sanos	Total
Resultados de la prueba diagnóstica	Positivo (+)	VERDADEROS POSITIVOS VP	FALSO POSITIVO FP $\alpha$	VP + FP
	Negativo (-)	FALSO NEGATIVO FN $\beta$	VERDADEROS NEGATIVOS VN	FN + VN
Total		VP + FN	FP + VN	



$$\text{ESPECIFICIDAD} = \frac{\text{Sanos negativos}}{\text{Total de SANOS}}$$

$$\text{ESPECIFICIDAD} = \frac{\text{VN}}{\text{VN} + \text{FP}}$$

## La especificidad de la prueba diagnóstica

- Es la probabilidad de clasificar correctamente a un individuo sano.
- Es la probabilidad de que para un sujeto sano se obtenga un resultado negativo.
- *Es la capacidad del test para detectar a los sanos.*
- Es la probabilidad de obtener un resultado negativo cuando el individuo no tiene la enfermedad.
- Mide la capacidad para descartar la enfermedad cuando no está presente.
- Su valor depende del error ALFA ( $\alpha$ ).

## Exactitud en las pruebas diagnósticas

- Es la principal cualidad clínica de una prueba diagnóstica.
- Es la capacidad de una prueba diagnóstica para clasificar correctamente a los pacientes.
- Es la capacidad de una prueba diagnóstica para clasificar correctamente a los individuos en subgrupos clínicamente relevantes (enfermos y NO enfermos).
- En su forma más simple es la capacidad para distinguir entre dos estados de salud (enfermos y NO enfermos).

$$\text{EXACTITUD} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{FN} + \text{FP} + \text{VN}}$$

## Exactitud en las pruebas diagnósticas

$$\text{EXACTITUD} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{FN} + \text{FP} + \text{VN}}$$



SENSIBILIDAD



ESPECIFICIDAD

$$\text{EXACTITUD} = \frac{\text{ACIERTOS}}{\text{ACIERTOS} + \text{FALLOS}}$$

## Tipos de error en las pruebas diagnósticas

Tipos de error en las pruebas diagnósticas		Evento o enfermedad estudiada		
		Enfermos	Sanos	Total
Resultados de la prueba diagnóstica	Positivo (+)	VERDADEROS POSITIVOS VP	FALSO POSITIVO FP $\alpha$	VP + FP (total de positivos)
	Negativo (-)	FALSO NEGATIVO FN $\beta$	VERDADEROS NEGATIVOS VN	FN + VN
Total		VP + FN	FP + VN	

$$VPP = \frac{VP}{\text{Total de POSITIVOS}}$$

$$VPP = \frac{VP}{VP + FP}$$

## Valor Predictivo Positivo (VPP) de la prueba diagnóstica

- *Es la probabilidad de padecer la enfermedad si se obtiene un resultado positivo en el test.*
- Se puede definir el VPP como la proporción de resultados válidos (VP) entre todos resultados positivos de la prueba (VP + FP).
- Su valor depende del error ALFA ( $\alpha$ ) y de la POTENCIA ESTADÍSTICA.
- Su valor depende de la prevalencia (P) de la enfermedad; cuanto menor sea la prevalencia menor será el VPP y cuanto mayor sea la prevalencia mayor será el VPP.

## Tipos de error en las pruebas diagnósticas

Tipos de error en las pruebas diagnósticas		Evento o enfermedad estudiada		
		Enfermos	Sanos	Total
Resultados de la prueba diagnóstica	Positivo (+)	VERDADEROS POSITIVOS VP	FALSO POSITIVO FP $\alpha$	VP + FP (total de positivos)
	Negativo (-)	FALSO NEGATIVO FN $\beta$	VERDADEROS NEGATIVOS VN	FN + VN (total de negativos)
Total		VP + FN	FP + VN	

$$VPN = \frac{VN}{\text{Total de NEGATIVOS}}$$

$$VPN = \frac{VN}{VN + FN}$$

## Valor Predictivo Negativo (VPN) de la prueba diagnóstica

- *Es la probabilidad de que un sujeto con un resultado negativo en la prueba esté realmente sano.*
- El VPN es la proporción de resultados válidos (VN) entre todos los resultados negativos (VN + FN).
- Su valor depende del error BETA ( $\beta$ ).
- Su valor depende de la prevalencia (P) de la enfermedad en estudio. Cuanto menor sea la prevalencia, mayor será el VPN, y cuanto mayor sea la prevalencia, menor será el VPN.

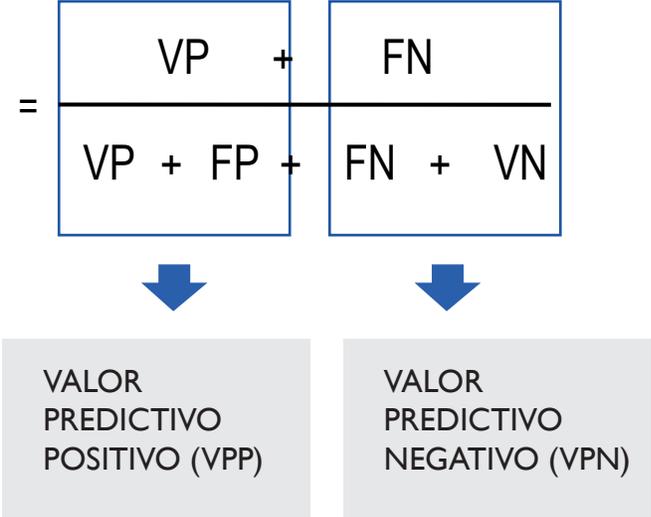
## Prevalencia de la enfermedad

Es la frecuencia con la que se presenta una determina enfermedad en una población definida en un momento dado.

$$\text{PREVALENCIA} = \frac{\text{VP} + \text{FN}}{\text{VP} + \text{FP} + \text{FN} + \text{VN}}$$

$$\text{PREVALENCIA} = \frac{\text{ENFERMOS}}{\text{ENFERMOS} + \text{SANOS}}$$

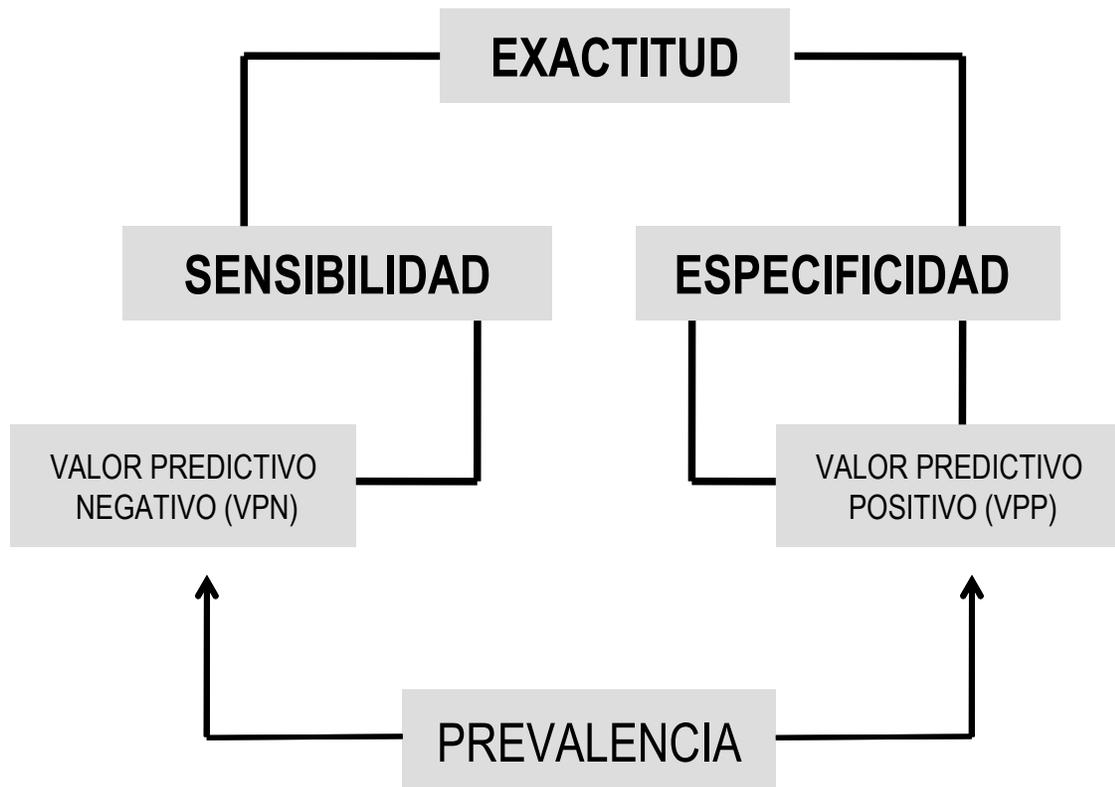
## Prevalencia de la enfermedad

$$\text{PREVALENCIA} = \frac{\text{VP} + \text{FN}}{\text{VP} + \text{FP} + \text{FN} + \text{VN}}$$


VALOR PREDICTIVO POSITIVO (VPP)

VALOR PREDICTIVO NEGATIVO (VPN)

Los valores predictivos (VPP y VPN) dependen de la prevalencia de la enfermedad en estudio.



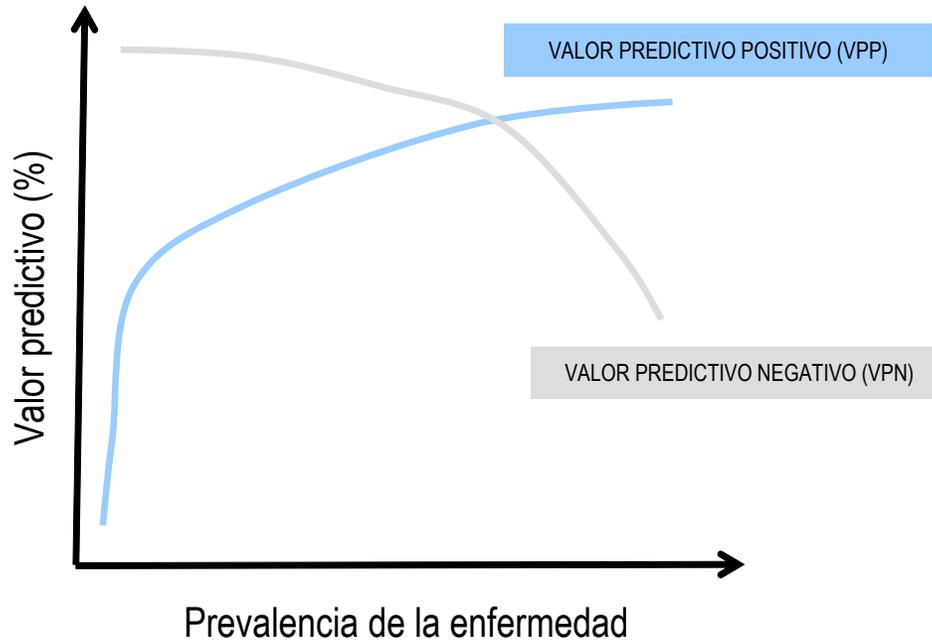
## Relación entre la prevalencia y el valor predictivo positivo (VPP)

Prevalencia de la enfermedad	Resultado prueba diagnóstica	Enfermos	Sanos (No enfermos)	Total	VPP
1%	+	99	495	594	15% (99/595)
	-	1	9404	9406	
	Total	100	9900	10000	
5%	+	495	475	970	51% (495/970)
	-	5	9025	9030	
	Total	500	9500	10000	

SENSIBILIDAD DIAGNÓSTICA = 99%

ESPECIFICIDAD DIAGNÓSTICA = 95%

## Relación entre la prevalencia y el valor predictivo



## Razón de verosimilitud

$$RV + = \frac{\text{SENSIBILIDAD}}{1 - \text{ESPECIFICIDAD}}$$

$$RV + = \frac{\text{Tasa de VP (\%)}}{\text{Tasa de FP (\%)}}$$

¿Cuál es la probabilidad de que un paciente que acude a la consulta tenga un carcinoma de próstata si el PSA (+) ( $[PSA]_s > 4 \text{ ng/ml}$ )?

Variable PREDICTORA

PSA (+)



Variable RESULTADO

Carcinoma de próstata

VALOR PREDICTIVO POSITIVO (VPP)

Es la probabilidad de que un paciente tenga un carcinoma de próstata (CP) una vez que el test (PSA) ha dado positivo ( $[PSA] > 4 \text{ ng/ml}$ ).

¿Cuál es la probabilidad de que un paciente que acude a la consulta NO tenga un carcinoma de próstata si el PSA (-) ( $[PSA] < 4 \text{ ng/ml}$ )?

VALOR PREDICTIVO NEGATIVO (VPN)

Es la probabilidad de que un individuo sano NO tenga un carcinoma de próstata (CP) una vez que el test (PSA) ha dado negativo ( $[PSA] < 4 \text{ ng/ml}$ ).

¿Cuál es la probabilidad de que un paciente diagnosticado de carcinoma de próstata presente un PSA (+) ( $[PSA] > 4$  ng/ml)?

Variable PREDICTORA

Carcinoma de próstata



Variable RESULTADO

PSA (+) ( $[PSA] > 4$  ng/ml)

SENSIBILIDAD DIAGNÓSTICA

Es la probabilidad de que el test (PSA) sea positivo entre los individuos realmente enfermos (carcinoma de próstata).

¿Cuál es la probabilidad de que un paciente sin carcinoma de próstata (CP) presente un PSA (-) ( $[PSA] < 4 \text{ ng/ml}$ )?

### ESPECIFICIDAD DIAGNÓSTICA

Es la probabilidad de que el test (PSA) sea negativo entre los individuos realmente no enfermos (sin cáncer de próstata).

¿Cuál es la probabilidad de que un valor de PSA patológico ( $[PSA] > 4 \text{ ng/ml}$ )? sea  $X$  veces más probable en un paciente con carcinoma de próstata (CP) que en otro sujeto sin cáncer?

RAZÓN DE VEROSIMILITUD (RV+)

Es la probabilidad de que un resultado elevado del test (PSA) aparezca en los pacientes con carcinoma de próstata frente a los individuos realmente no enfermos (sin cáncer de próstata).

## Pregunta de investigación

¿Qué utilidad clínica tiene el tacto rectal en el diagnóstico del carcinoma de próstata (CP) en una muestra de pacientes con sospecha de CP que acuden a una consulta de Urología?

Objetivo: Evaluación del tacto rectal en el diagnóstico del CP en una consulta de urología.

Técnica analítica a valorar

TACTO RECTAL



Técnica de referencia  
(patrón de oro)

BIOPSIA PROSTÁTICA mediante  
ecografía transrectal

¿Cuál es el test estadístico más acertado?

## Tipos de error en las pruebas diagnósticas

Tipos de error en las pruebas diagnósticas		Técnica de referencia (BIOPSIA transrectal ecodirigida)	
		POSITIVA	NEGATIVA
TACTO RECTAL	Positivo (+)	VERDADEROS POSITIVOS VP	FALSO POSITIVO FP $\alpha$
	Negativo (-)	FALSO NEGATIVO FN $\beta$	VERDADEROS NEGATIVOS VN

Técnica de referencia (BIOPSIA prostática mediante ecografía transrectal)

TACTO RECTAL	positiva (+)	negativa (-)	Total
Positivo (+)	634	269	903
Negativo (-)	487	1251	1738
Total	1121	1520	2641

$$\text{SENSIBILIDAD} = \frac{\text{VP}}{\text{VP} + \text{FN}}$$

$$\text{SENSIBILIDAD} = \frac{634}{634 + 487} = 56,5\%$$

## Técnica de referencia (BIOPSIA prostática mediante ecografía transrectal)

TACTO RECTAL	positiva (+)	negativa (-)	Total
Positivo (+)	634	269	903
Negativo (-)	487	1251	1738
Total	1121	1520	2641

$$\text{ESPECIFICIDAD} = \frac{\text{VN}}{\text{VN} + \text{FP}}$$

$$\text{ESPECIFICIDAD} = \frac{1251}{1251 + 269} = 82,3\%$$

## Técnica de referencia (BIOPSIA prostática mediante ecografía transrectal)

TACTO RECTAL	positiva (+)	negativa (-)	Total
Positivo (+)	634	269	903
Negativo (-)	487	1251	1738
Total	1121	1520	2641

$$\text{EXACTITUD} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{FN} + \text{FP} + \text{VN}}$$

$$\text{EXACTITUD} = \frac{634 + 1251}{634 + 487 + 269 + 1251} = 71,4 \%$$

Técnica de referencia (BIOPSIA prostática mediante ecografía transrectal)			
TACTO RECTAL	positiva (+)	negativa (-)	Total
Positivo (+)	634	269	903
Negativo (-)	487	1251	1738
Total	1121	1520	2641

$$VPP = \frac{VP}{VP + FP} \quad VPP = \frac{634}{634 + 269} = 70,2\%$$

La probabilidad de tener un CP es del 70% cuando el tacto rectal es positivo.

Técnica de referencia (BIOPSIA prostática mediante ecografía transrectal)			
TACTO RECTAL	positiva (+)	negativa (-)	Total
Positivo (+)	634	269	903
Negativo (-)	487	1251	1738
Total	1121	1520	2641

$$VPN = \frac{VN}{VN + FN} \quad VPN = \frac{1251}{1251 + 487} = 72,0\%$$

La probabilidad de NO tener un CP es de tan sólo el 72% cuando el tacto rectal es negativo. Como la prueba tiene un VPN bajo, un resultado negativo del tacto rectal sólo permite descartar el diagnóstico de CP con una probabilidad del 72%.

## Técnica de referencia (BIOPSIA prostática mediante ecografía transrectal)

TACTO RECTAL	positiva (+)	negativa (-)	Total
Positivo (+)	634	269	903
Negativo (-)	487	1251	1738
Total	1121	1520	2641

$$\text{PREVALENCIA} = \frac{\text{VP} + \text{FN}}{\text{VP} + \text{FP} + \text{FN} + \text{VN}}$$

$$\text{PREVALENCIA} = \frac{634 + 487}{634 + 269 + 487 + 1251} = 42,4 \%$$

Técnica de referencia (BIOPSIA prostática mediante ecografía transrectal)			
TACTO RECTAL	positiva (+)	negativa (-)	Total
Positivo (+)	634	269	903
Negativo (-)	487	1251	1738
Total	1121	1520	2641

$$RV + = \frac{\text{SENSIBILIDAD}}{1 - \text{ESPECIFICIDAD}} \quad RV + = \frac{0,565}{1 - 0,823} = 3,2$$

Un tacto rectal anormal (+) es 3 veces más probable en un paciente con CP que en otro sujeto sin cáncer.

## Pregunta de investigación

¿Qué utilidad clínica tiene la tira reactiva en orina en el diagnóstico de la infección del tracto urinario (ITU)?

Objetivo: Evaluación de la TIRA REACTIVA en el diagnóstico de la ITU.

Técnica analítica a valorar

TIRA REACTIVA  
(orina)



Técnica de referencia  
(patrón de oro)

CULTIVO  
(orina)

¿Cuál es el test estadístico más acertado?

Latorre C, Noguero M, Mira A. Evaluación de la tira reactiva para el diagnóstico de infección urinaria en niños y adultos. Med Clin (Barc) 2001; 116: 286-289.

## Tipos de error en las pruebas diagnósticas

Tipos de error en las pruebas diagnósticas		Técnica de referencia (CULTIVO)	
		POSITIVO	NEGATIVO
Test de laboratorio TIRA REACTIVA orina	Positiva (+)	VERDADEROS POSITIVOS VP	FALSO POSITIVO FP $\alpha$
	Negativa (-)	FALSO NEGATIVO FN $\beta$	VERDADEROS NEGATIVOS VN

## INFECCIÓN DEL TRACTO URINARIO (ITU)

Tira reactiva	SI	NO	Total
Positiva (+)	86	433	519
Negativa (-)	20	548	568
Total	106	981	1087

$$\text{SENSIBILIDAD} = \frac{\text{VP}}{\text{VP} + \text{FN}}$$

$$\text{SENSIBILIDAD} = \frac{86}{86 + 20} = 81,1\%$$

## INFECCIÓN DEL TRACTO URINARIO (ITU)

Tira reactiva	SI	NO	Total
Positiva (+)	86	433	519
Negativa (-)	20	548	568
Total	106	981	1087

$$\text{ESPECIFICIDAD} = \frac{\text{VN}}{\text{VN} + \text{FP}}$$

$$\text{ESPECIFICIDAD} = \frac{548}{548 + 433} = 55,9\%$$

## INFECCIÓN DEL TRACTO URINARIO (ITU)

Tira reactiva	SI	NO	Total
Positiva (+)	86	433	519
Negativa (-)	20	548	568
Total	106	981	1087

$$\text{EXACTITUD} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{FN} + \text{FP} + \text{VN}}$$

$$\text{EXACTITUD} = \frac{86 + 548}{86 + 20 + 433 + 548} = 58,3\%$$

INFECCIÓN DEL TRACTO URINARIO (ITU)			
Tira reactiva	SI	NO	Total
Positiva (+)	86	433	519
Negativa (-)	20	548	568
Total	106	981	1087

$$VPP = \frac{VP}{VP + FP} \quad VPP = \frac{86}{86 + 433} = 16,6\%$$

La probabilidad de tener una ITU es de tan sólo el 17% cuando la tira reactiva es positiva.

INFECCIÓN DEL TRACTO URINARIO (ITU)			
Tira reactiva	SI	NO	Total
Positiva (+)	86	433	519
Negativa (-)	20	548	568
Total	106	981	1087

$$VPN = \frac{VN}{VN + FN} \quad VPN = \frac{548}{548 + 20} = 96,5\%$$

La probabilidad de NO tener una ITU es del 96,5% cuando la tira reactiva es negativa.

Como la prueba tiene un VPN alto, un resultado negativo de la tira urinaria descarta el diagnóstico de ITU con una probabilidad del 96,5%.

INFECCIÓN DEL TRACTO URINARIO (ITU)			
Tira reactiva	SI	NO	Total
Positiva (+)	86	433	519
Negativa (-)	20	548	568
Total	106	981	1087

$$\text{PREVALENCIA} = \frac{\text{VP} + \text{FN}}{\text{VP} + \text{FP} + \text{FN} + \text{VN}}$$

$$\text{PREVALENCIA} = \frac{86 + 20}{86 + 433 + 20 + 548} = 9,7 \%$$

INFECCIÓN DEL TRACTO URINARIO (ITU)			
Tira reactiva	SI	NO	Total
Positiva (+)	86	433	519
Negativa (-)	20	548	568
Total	106	981	1087

$$RV + = \frac{\text{SENSIBILIDAD}}{1 - \text{ESPECIFICIDAD}}$$

$$RV + = \frac{0,811}{1 - 0,559} = 1,84$$

Una tira reactiva anormal (+) es casi 2 veces más probable en un paciente con ITU que en otro sujeto sin infección.

## Pregunta de investigación

¿Qué utilidad clínica tienen los niveles de ferritina sérica en el diagnóstico de la anemia ferropénica?

Objetivo: Evaluación de la FERRITINA en el diagnóstico de la anemia ferropénica.

Técnica analítica a evaluar

FERRITINA  
(sérica)



Técnica de referencia  
(patrón de oro)

ANEMIA FERROPÉNICA

¿Cuál es el test estadístico más acertado?

## Tipos de error en las pruebas diagnósticas

Tipos de error en las pruebas diagnósticas		Realidad	
		Anemia ferropénica	Sanos
Test de laboratorio FERRITINA sérica	Positivo (+) FERRITINA BAJA	VERDADEROS POSITIVOS VP	FALSO POSITIVO FP $\alpha$
	Negativo (-) FERRITINA NORMAL	FALSO NEGATIVO FN $\beta$	VERDADEROS NEGATIVOS VN

ANEMIA FERROPÉNICA			
FERRITINA	SI	NO	Total
BAJA (+)	731	270	1001
NORMAL (-)	78	1500	1578
Total	809	1770	2579

$$\text{SENSIBILIDAD} = \frac{\text{VP}}{\text{VP} + \text{FN}}$$

$$\text{SENSIBILIDAD} = \frac{731}{731 + 78} = 90,3\%$$

Si una prueba tiene una sensibilidad alta, un resultado negativo descarta el diagnóstico.

ANEMIA FERROPÉNICA			
FERRITINA	SI	NO	Total
BAJA (+)	731	270	1001
NORMAL (-)	78	1500	1578
Total	809	1770	2579

$$\text{ESPECIFICIDAD} = \frac{\text{VN}}{\text{VN} + \text{FP}}$$

$$\text{ESPECIFICIDAD} = \frac{1500}{1500 + 270} = 84,7\%$$

Si una prueba tiene una especificidad alta, un resultado positivo confirma el diagnóstico.

ANEMIA FERROPÉNICA			
FERRITINA	SI	NO	Total
BAJA (+)	731	270	1001
NORMAL (-)	78	1500	1578
Total	809	1770	2579

$$\text{EXACTITUD} = \frac{VP + VN}{VP + FN + FP + VN}$$

$$\text{EXACTITUD} = \frac{731 + 1500}{731 + 78 + 270 + 1500} = 86,5\%$$

Una exactitud elevada permite distinguir con un mínimo error entre dos estados de salud.

ANEMIA FERROPÉNICA			
FERRITINA	SI	NO	Total
BAJA (+)	731	270	1001
NORMAL (-)	78	1500	1578
Total	809	1770	2579

$$VPP = \frac{VP}{VP + FP} \quad VPP = \frac{731}{731 + 270} = 73\%$$

La probabilidad de tener una anemia ferropénica es del 73% cuando la ferritina sérica es baja.

ANEMIA FERROPÉNICA			
FERRITINA	SI	NO	Total
BAJA (+)	731	270	1001
NORMAL (-)	78	1500	1578
Total	809	1770	2579

$$VPN = \frac{VN}{VN + FN} \quad VPN = \frac{1500}{1500 + 78} = 95,1\%$$

La probabilidad de NO tener una anemia ferropénica es del 95,1% cuando la ferritina sérica es normal. Como la prueba tiene un VPN alto, una ferritina normal descarta el diagnóstico de anemia ferropénica con una probabilidad del 95,1%.

## ANEMIA FERROPÉNICA

FERRITINA	SI	NO	Total
BAJA (+)	731	270	1001
NORMAL (-)	78	1500	1578
Total	809	1770	2579

$$\text{PREVALENCIA} = \frac{\text{VP} + \text{FN}}{\text{VP} + \text{FP} + \text{FN} + \text{VN}}$$

$$\text{PREVALENCIA} = \frac{731 + 78}{731 + 270 + 78 + 1500} = 31\%$$

ANEMIA FERROPÉNICA			
FERRITINA	SI	NO	Total
BAJA (+)	731	270	1001
NORMAL (-)	78	1500	1578
Total	809	1770	2579

$$RV + = \frac{\text{SENSIBILIDAD}}{1 - \text{ESPECIFICIDAD}}$$

$$RV + = \frac{0,903}{1 - 0,847} = 5,9$$

Una ferritina baja es casi 6 veces más probable en un paciente con anemia ferropénica que en otro sujeto sin anemia ferropénica.

## Pregunta de investigación

¿Qué utilidad clínica tiene el título sérico de anticuerpos antinucleares en el diagnóstico de enfermedades autoinmunes?

Objetivo: Evaluación del título de ANA en el diagnóstico de diferentes enfermedades autoinmunes.

Técnica analítica a valorar

ANA (IFI)



Técnica de referencia  
(patrón de oro)

ENFERMEDADES  
AUTOINMUNES

¿Cuál es el test estadístico más acertado?

## Tipos de error en las pruebas diagnósticas

Tipos de error en las pruebas diagnósticas		Realidad	
		Enfermedad autoinmune	Sanos
Test de laboratorio ANA (IFI)	Positivo (+)	VERDADEROS POSITIVOS VP	FALSO POSITIVO FP $\alpha$
	Negativo (-)	FALSO NEGATIVO FN $\beta$	VERDADEROS NEGATIVOS VN

## Razón de verosimilitud

Cocientes de probabilidades

$$RV + = \frac{\text{SENSIBILIDAD}}{1 - \text{ESPECIFICIDAD}}$$

$$RV + = \frac{\text{Tasa de VP (\%)}}{\text{Tasa de FP (\%)}}$$

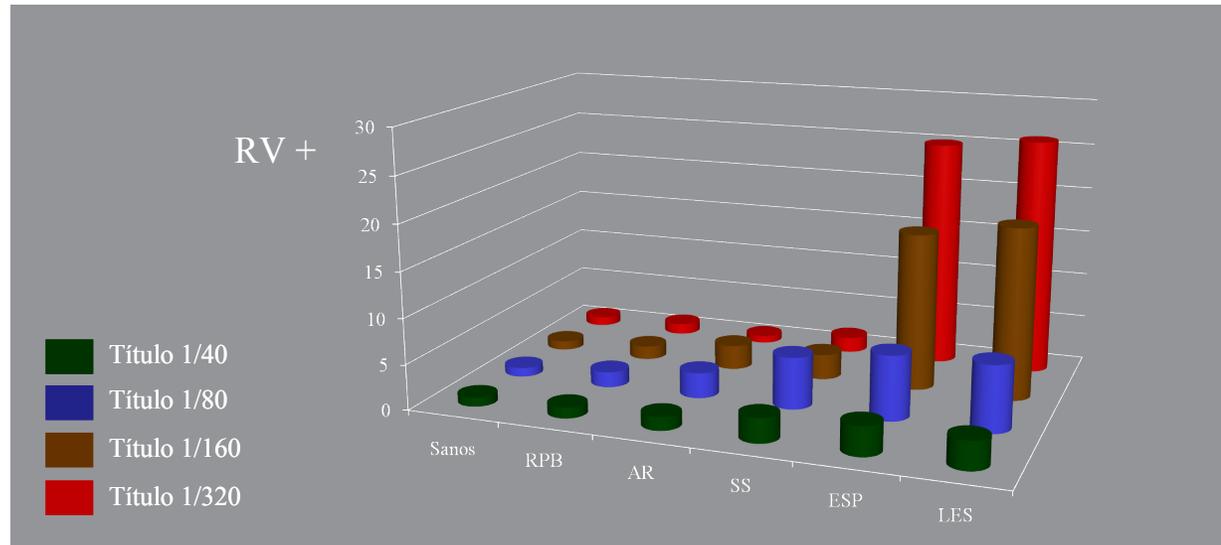
Grupos de estudio	Razón de verosimilitud (+) (cociente de probabilidades positivo)			
	Título 1/40	Título 1/80	Título 1/160	Título 1/320
LES (n= 38)	3,1	7,3	18,9	26,1
ESP (n= 37)	3,2	7,1	17,3	25,1
SS (n= 37)	2,7	5,7	14,7	21,3
AR (n= 38)	1,5	2,8	2,7	0,81
RPB (n= 37)	1,2	1,7	1,5	1,2
Sanos (n=120)	1,0	1,0	1,0	1,0

La razón de verosimilitud positivo (RV+) o cociente de probabilidad (CP) positivo (sensibilidad dividida por el complementario de la especificidad) es una medida que calcula cuánto más probable es que un resultado positivo de una prueba (ANA) se presente en un sujeto enfermo (EAI) frente a uno sano.

La razón de verosimilitud negativo (RV-) o cociente de probabilidad (CP) negativo (complementario de la sensibilidad dividida por la especificidad) relaciona la probabilidad de que un resultado negativo de una prueba se presente en un enfermo frente a un sano.

## Razón de verosimilitud

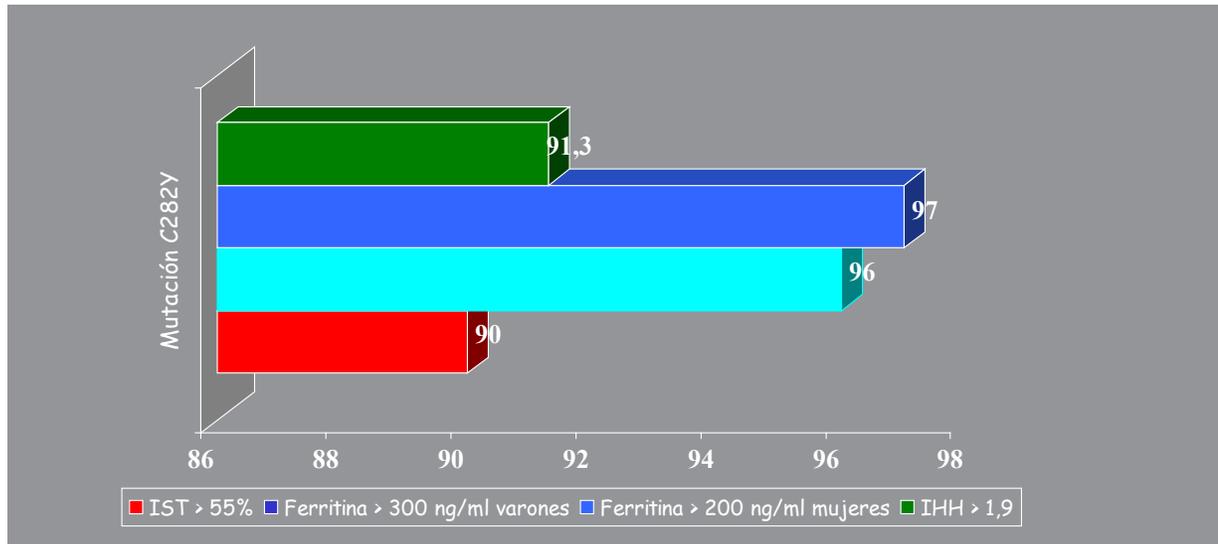
### Cocientes de probabilidades



Camps García MT, Muñoz Rocan N, Calvo de Mora A, De Ramón Garrido E. Los cocientes de probabilidad como medida de la utilidad de la determinación de anticuerpos antinucleares por inmunofluorescencia indirecta en la clínica. Aplicación de los resultados de la investigación clínica. Med Clin (Barc) 1999;113: 476.

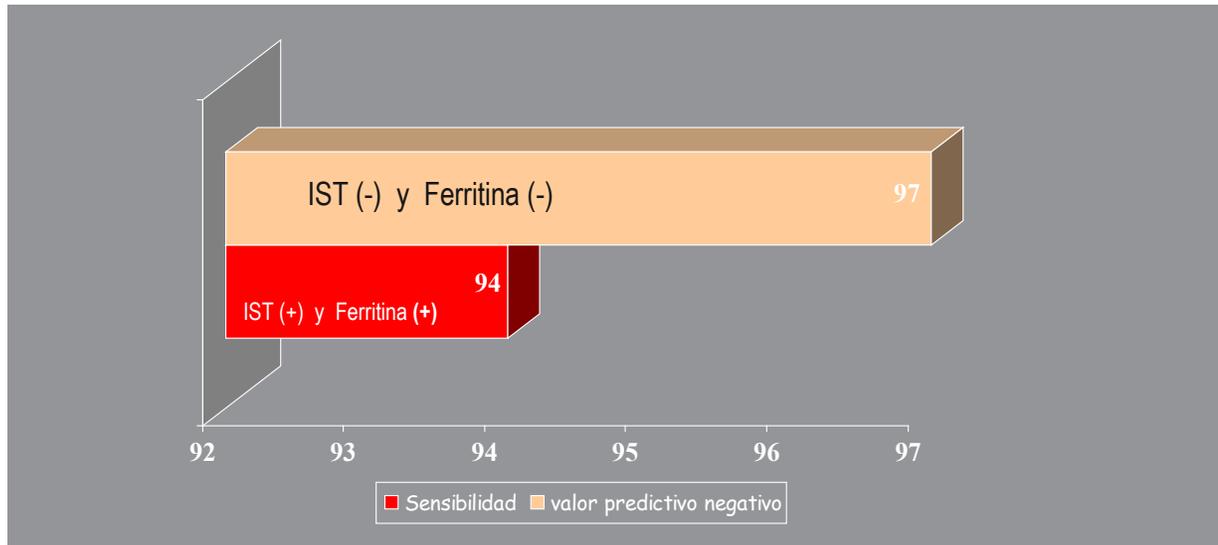
## Tipos de error en las pruebas diagnósticas

Tipos de error en las pruebas diagnósticas		Realidad	
		Hemocromatosis	Sanos
Test de laboratorio FERRITINA IST	Positivo (+) FERRITINA alta IST alto	VERDADEROS POSITIVOS VP	FALSO POSITIVO FP $\alpha$
	Negativo (-) FERRITINA normal IST normal	FALSO NEGATIVO FN $\beta$	VERDADEROS NEGATIVOS VN



Genotypic/phenotypic correlations in genetic hemochromatosis: evolution of diagnostic criteria. Adams PC, Chakrabarti S. Gastroenterology. 1998 Feb;114(2):319-23.

## Valoración conjunta del IST y ferritina en el screening de la hemocromatosis



Holman JR. Hereditary hemochromatosis. J Fam Pract 1997; 44: 304-308.

## Pruebas diagnósticas en el cáncer de próstata

FAP

PSA

Cociente

PSA libre/ PSA total

Tacto rectal

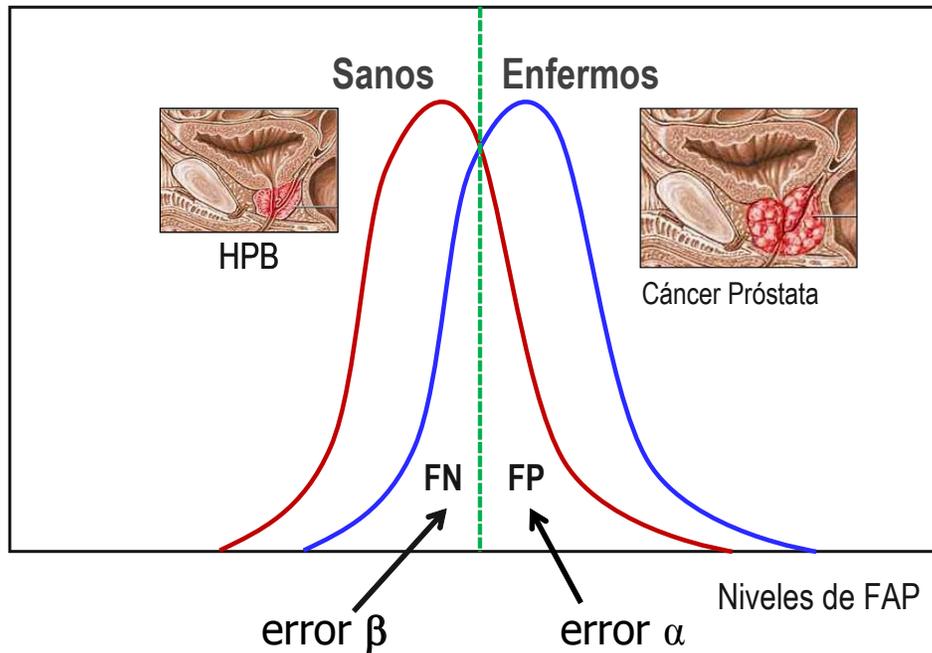
Citología aspiración

Biopsia prostática mediante  
ecografía transrectal

## Pruebas diagnósticas:

Mala rentabilidad diagnóstica.

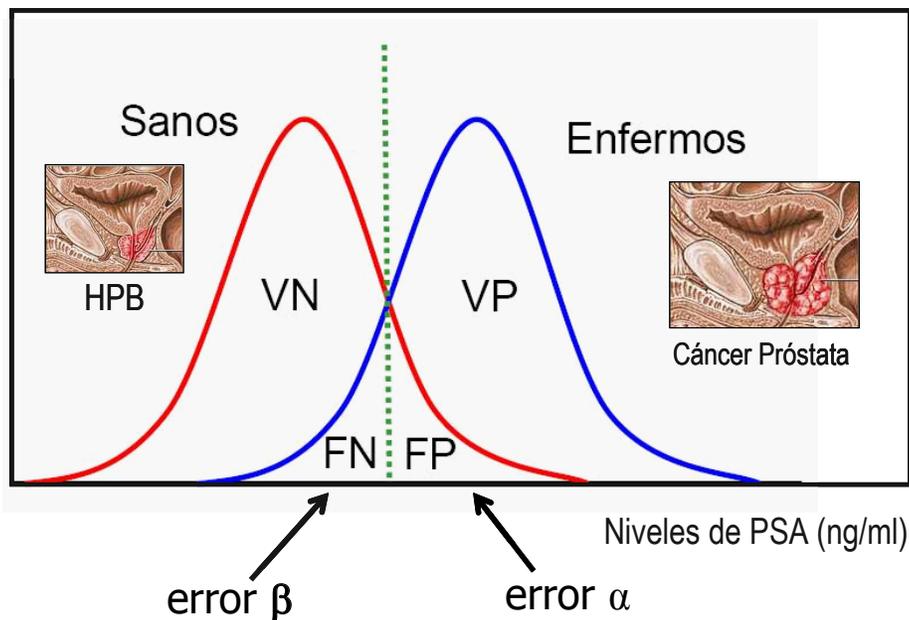
Baja sensibilidad y especificidad (FAP)



## Pruebas diagnósticas:

Buena rentabilidad diagnóstica.

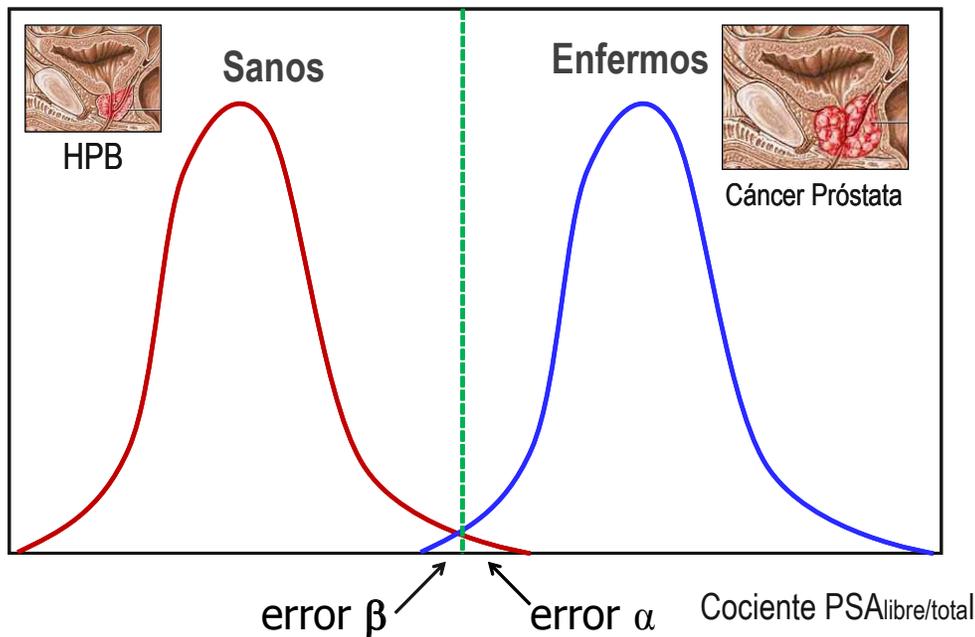
Equilibrio entre sensibilidad y especificidad (PSA)



## Pruebas diagnósticas:

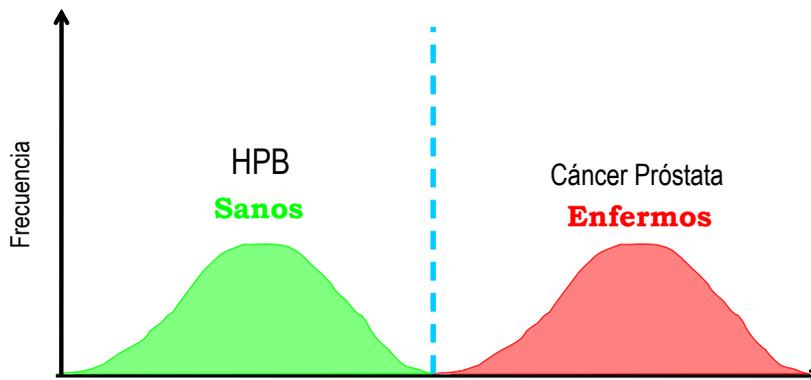
Excelente rentabilidad diagnóstica.

Máxima sensibilidad y especificidad (Cociente PSA<sub>libre/total</sub>)



## Situación ideal de las pruebas diagnósticas:

Máxima sensibilidad y especificidad



Su Negatividad indican  
siempre la ausencia de la  
enfermedad

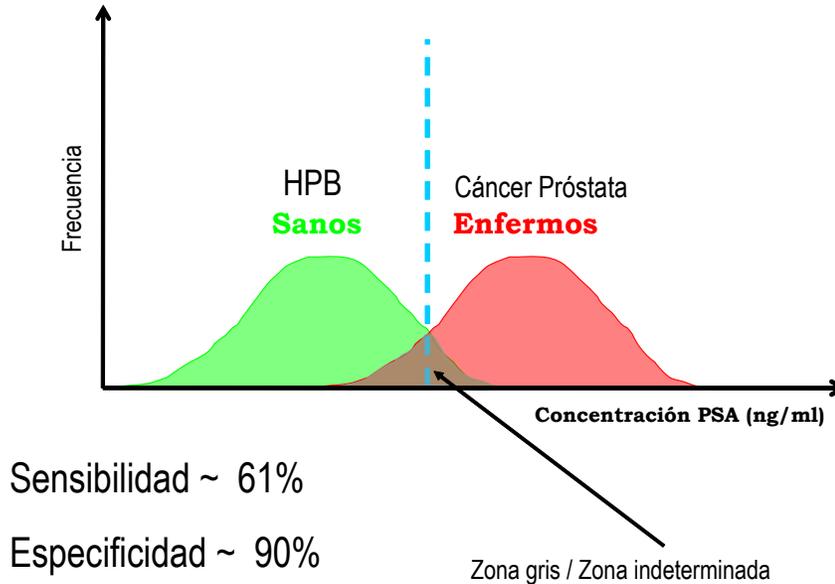
Especificidad ~ 100%

Su Positividad indican  
siempre la presencia de la  
enfermedad

Sensibilidad ~ 100%

## Situación real de las pruebas diagnósticas:

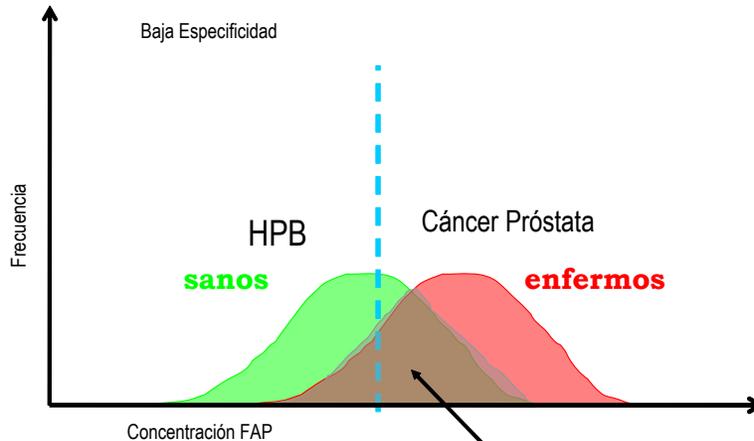
PSA - Buen equilibrio entre sensibilidad y especificidad



## Pruebas diagnósticas a evitar:

Fosfatasa ácida prostática

Mal equilibrio entre sensibilidad y especificidad

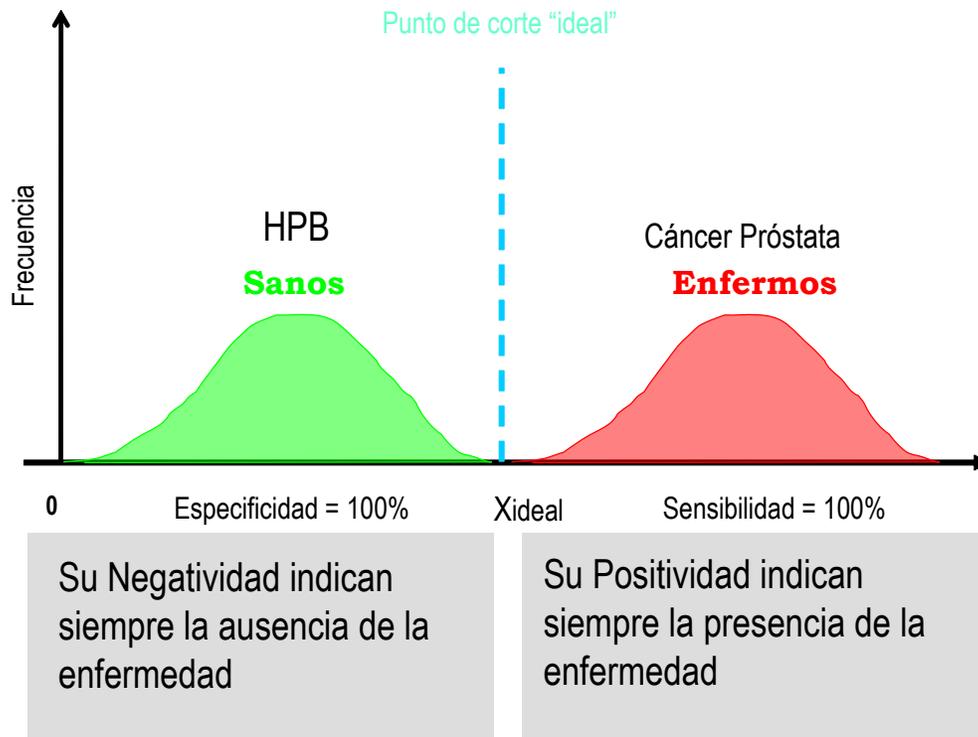


Sensibilidad ~ 60-70%

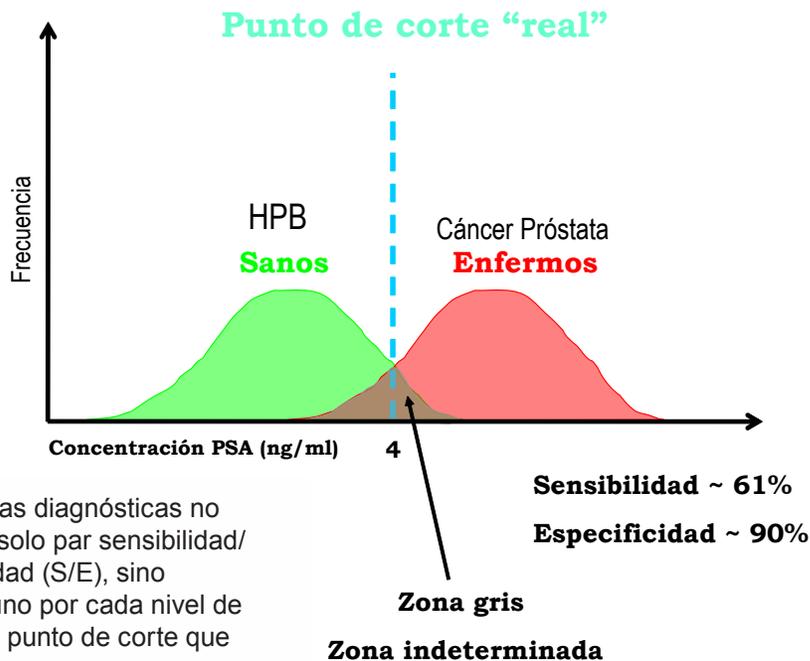
Especificidad ~ 40-50%

Zona gris / Zona indeterminada

## Elección del punto de corte

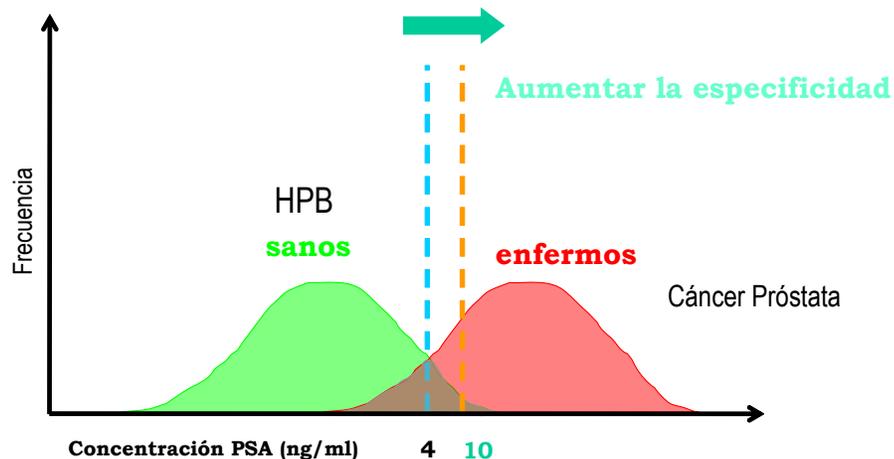


## Elección del punto de corte



Las pruebas diagnósticas no tienen un solo par sensibilidad/especificidad (S/E), sino muchos, uno por cada nivel de decisión o punto de corte que se adopte.

## Aumento del valor del punto de corte

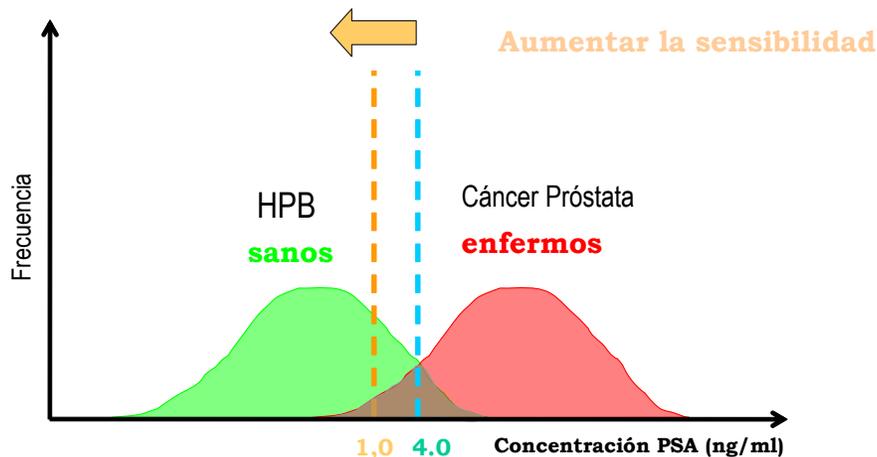


Sensibilidad ~ 30%

Especificidad  $\approx$  99%

Para cualquier prueba en la que la distribución de los resultados de los grupos sano y enfermo se solapen, existe un equilibrio entre sensibilidad y especificidad. Ambas varían en direcciones opuestas: si se AUMENTA el punto de corte disminuye la sensibilidad, y aumenta la especificidad.

## Disminución del valor del punto de corte

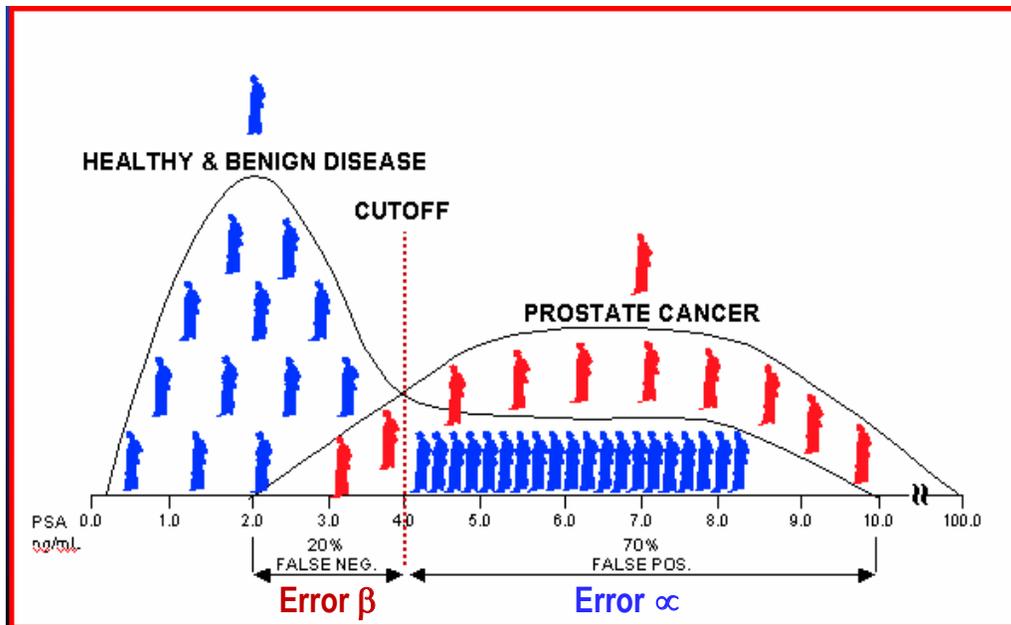


Sensibilidad  $\approx 92\%$

Especificidad  $\sim 35\%$

Para cualquier prueba en la que la distribución de los resultados de los grupos sano y enfermo se solapan, existe un equilibrio entre sensibilidad y especificidad. Ambas varían en direcciones opuestas: si se DISMINUYE el punto de corte aumenta la sensibilidad, y disminuye la especificidad.

## PSA: ejemplo de error alfa (FP) y error beta (FN)



PUNTO DE CORTE = 4 ng/ml

- Buena sensibilidad a la hora de detectar los CP
- Pero existe un número elevado de falsos positivos (HBP) > baja especificidad

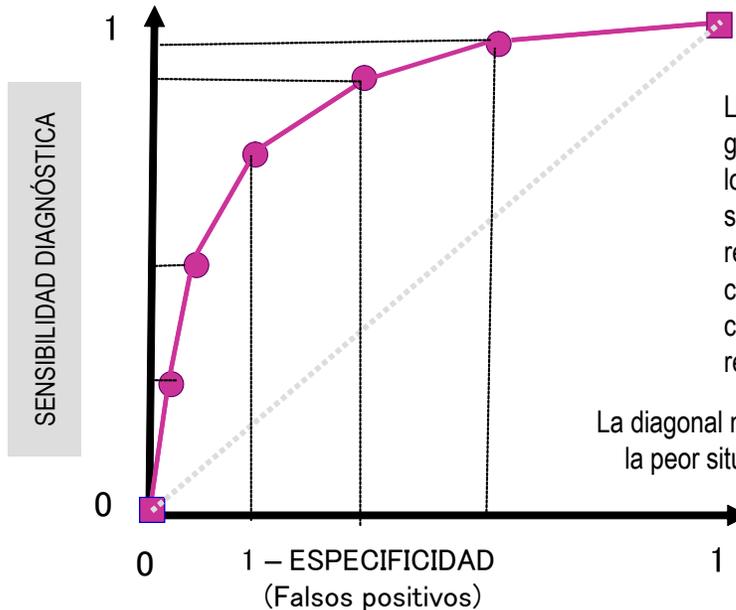
## Sensibilidad y especificidad diagnóstica:

### Pruebas en el Cáncer de Próstata

Prueba	S	E	Patrón de oro (técnica de referencia)
FAP	56%	94%	Biopsia de la próstata mediante Ecografía transrectal
PSA	61%	90%	
PSA libre/PSA total	69%	95%	
Examen rectal	56%	82%	
Citología aspiración	55%	91%	

# Construcción de la curva ROC para cada prueba diagnóstica

En el eje Y de coordenadas se sitúa la sensibilidad o fracción de verdaderos positivos, calculada en los enfermos

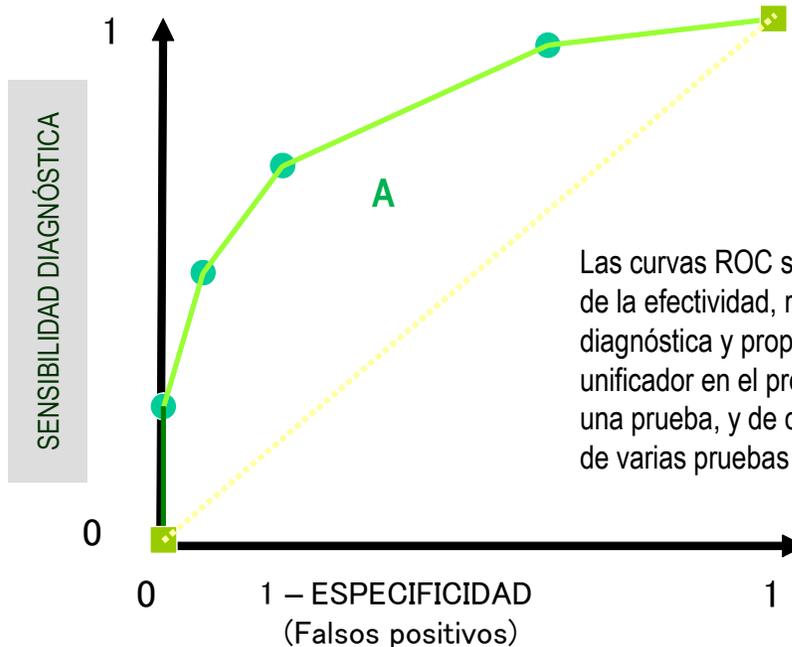


La curva ROC es un gráfico que muestra todos los pares posibles sensibilidad/especificidad resultantes de la variación continua de los puntos de corte en todo el rango de resultados observados.

La diagonal muestra por tanto la peor situación posible.

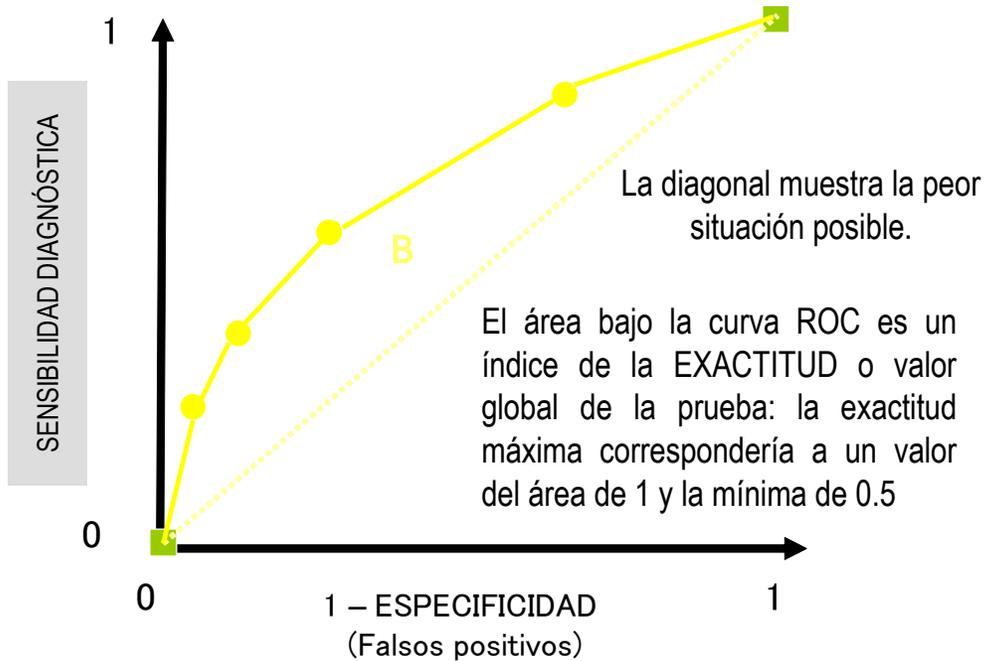
En el eje X se sitúa la fracción de FP o 1 - especificidad, definida como  $1 - E = FP / VN + FP$  y calculada en los no enfermos (sanos).

## Diagrama de la curva ROC de la prueba diagnóstica A

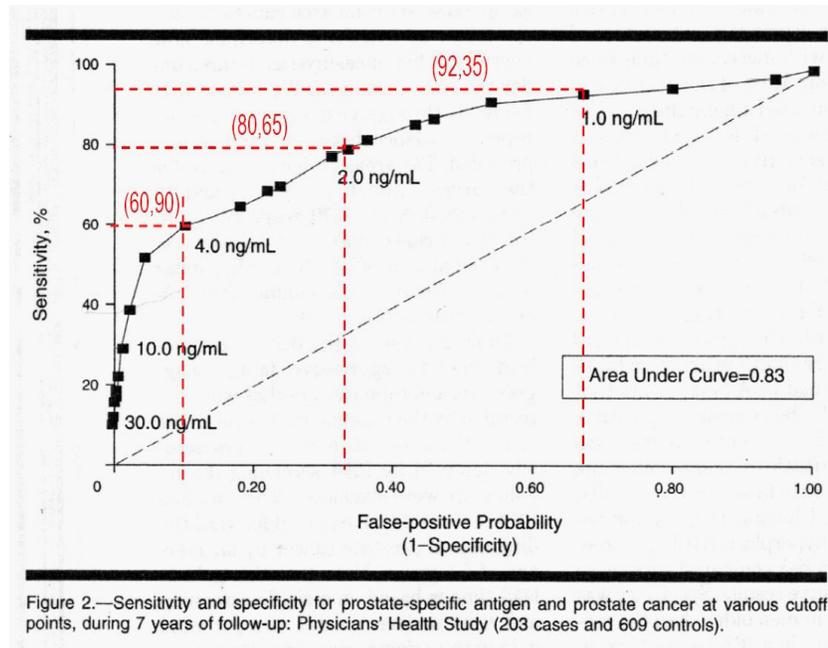


Las curvas ROC son los mejores índices de la efectividad, rentabilidad o exactitud diagnóstica y proporcionan un criterio unificador en el proceso de evaluación de una prueba, y de comparación simultánea de varias pruebas diagnósticas.

## Diagrama de la curva ROC de la prueba diagnóstica B



## Diagrama de la curva ROC del PSA



Gann PH, Hennekens CH, Stampfer MJ. A prospective evaluation of plasma prostate-specific antigen for detection of prostatic cancer. JAMA. 1995; 273: 289-294.

## **ORIGINAL**

# **Precisión diagnóstica de la calprotectina fecal para predecir una colonoscopia patológica**

María del Valle García Sánchez<sup>a</sup>, Raúl González<sup>b</sup>, Eva Iglesias Flores<sup>a</sup>, Federico Gómez Camacho<sup>a</sup>, Luis Casais Juanena<sup>a</sup>, Antonio Cerezo Ruiz<sup>a</sup>, Manuel Montero Pérez-Barquero<sup>c</sup>, Jordi Muntané<sup>b</sup> y Juan Francisco de Dios Vega<sup>a</sup>

<sup>a</sup>Unidad Clínica Aparato Digestivo. Hospital Universitario Reina Sofía. Córdoba.

<sup>b</sup>Unidad de Investigación. Hospital Universitario Reina Sofía. Córdoba.

<sup>c</sup>Medicina Interna. Hospital Universitario Reina Sofía. Córdoba. España.

**FUNDAMENTO Y OBJETIVO:** El dolor abdominal y la diarrea son síntomas comunes en la población general que no permiten discriminar enfermedad intestinal orgánica. La colonoscopia es la prueba de referencia, pero es cara, invasiva, no puede repetirse frecuentemente y supone una gran carga asistencial. La calprotectina fecal (CPF) es un marcador fácil y rápido de medir, no invasivo y barato, que podría detectar la enfermedad orgánica en la zona del colon. El objetivo de nuestro estudio es valorar la utilidad de este marcador para predecir una colonoscopia patológica y correlacionar sus valores con el grado de actividad en la enfermedad inflamatoria intestinal (EII).

**PACIENTES Y MÉTODO:** Se determinó la calprotectina en muestras fecales (mediante enzimoanálisis) en 190 personas sometidas a una colonoscopia. Los diagnósticos endoscópicos fueron: 117 normales, 28 pólipos colónicos, 20 cáncer colorrectal (CCR) y 25 EII. El análisis de los datos se realizó con el programa SPSS 11.0.

**RESULTADOS:** La media (DE) de CPF estaba aumentada en CCR (726,6 [533] mg/kg) y EII (2.171,1 [2.133,6] mg/kg). No hubo diferencias en pacientes con pólipos (158,3 [15,8] mg/kg). En pacientes con EII, los valores se correlacionaron con el grado de actividad de la enfermedad. El valor de corte fue 217 mg/kg, y se obtuvo una sensibilidad de la CPF del 85% y un valor predictivo negativo del 93%. En el grupo control, la única variable que se relacionó con un aumento de CPF fue la medicación previa con ácido acetilsalicílico (AAS) y/o antiinflamatorios no esteroideos (AINE).

**CONCLUSIONES:** La CPF es una prueba no invasiva, barata y sensible para detectar enfermedad orgánica en la zona del colon, que puede ser útil para seleccionar a pacientes que deben someterse a una colonoscopia. Se correlaciona con el grado de actividad en la EII. La toma de AAS y/o AINE podría aumentar la tasa de falsos positivos.

*Palabras clave:* Calprotectina fecal. Colonoscopia. Inflamación intestinal.

## Introducción (fragmento del texto)

### Objetivos

“El objetivo fundamental de este estudio es valorar la utilidad de la calprotectina fecal (CPF) para predecir la presencia de una colonoscopia patológica y, secundariamente, analizar su uso para discriminar entre diferentes enfermedades orgánicas y correlacionar sus valores con el grado de actividad clínica, endoscópica e histológica en pacientes con enfermedad inflamatoria intestinal (EII)”.

## Pregunta de investigación

¿Qué utilidad clínica tienen los niveles de calprotectina fecal de la actividad de la enfermedad inflamatoria intestinal?

### Técnica analítica a valorar

Calprotectina fecal



### Técnica de referencia (patrón de oro)

Colonoscopia: Índice de actividad de la enfermedad inflamatoria intestinal

## Resultados (fragmento del texto)

“La curva ROC mostró que el valor de CPF de 217 mg/kg fue el mejor punto de corte para discriminar a los pacientes con enfermedad orgánica en colon. A partir de estos datos, obtuvimos una sensibilidad de la medida de CPF del 85%, una especificidad del 81%, un valor predictivo positivo del 64% y un valor predictivo negativo del 93%, respectivamente ( $p < 0,001$ )”.

	Valor
Sensibilidad	85%
Especificidad	81%
Valor predictivo positivo	64%
Valor predictivo negativo	93%

## Discusión (fragmento del texto)

“En el presente estudio, la CPF ha mostrado ser una prueba altamente sensible para detectar a pacientes con CCR y EII utilizando un punto de corte de 217  $\mu\text{g/g}$ ”.

“La elección de un punto de corte óptimo debería realizarse a partir de un procedimiento de referencia estándar, lo que en pacientes con dolor abdominal y diarrea implicará un estudio gastrointestinal completo. De todas formas, siempre es aconsejable utilizar puntos de corte más bajos porque lo que debemos pretender con esta prueba es seleccionar a los pacientes que requieren un estudio endoscópico posterior”.

## MITO I: “La sonrisa de la esfinge”

Me parece acertada la elección de la curva ROC. Sin embargo...

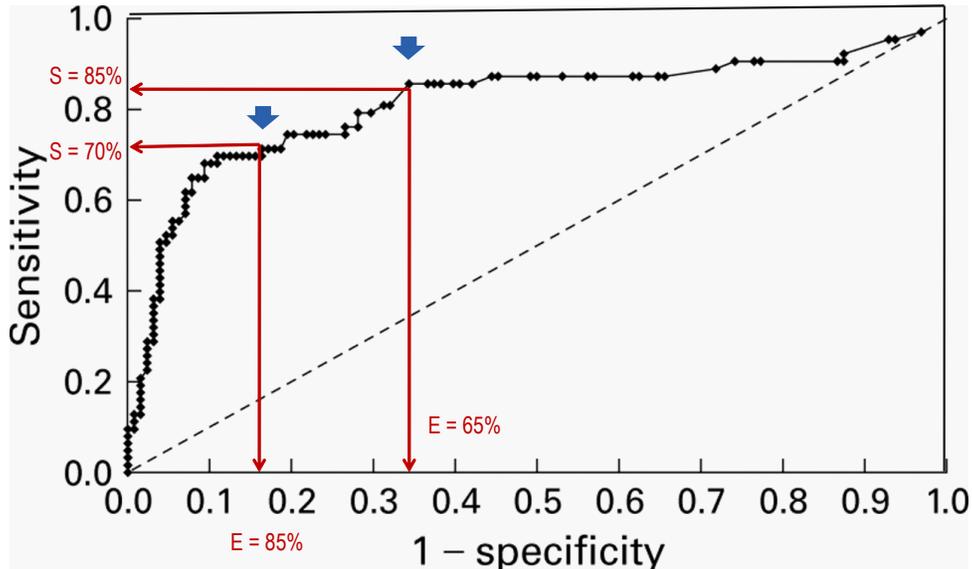
*¿Por qué no has incluido el gráfico de la curva?*

*¿Cuál es el área de la curva ROC y su IC95%?*

En la bibliografía existen datos de curvas ROC que dan datos inferiores de rentabilidad diagnóstica, que no coinciden con lo presentados por García Sánchez et al.

*¿Cómo justificas esas diferencias?*

## Diagrama de la curva ROC de la calprotectina fecal



Tibblea J et al. Faecal calprotectin and faecal occult blood tests in the diagnosis of colorectal carcinoma and adenoma. Gut 2001;49:402-408.

## Interpretación de Hanley y McNeil del Área bajo la curva ROC

- El área bajo la curva (ABC) ROC es una medida global de la exactitud de una prueba diagnóstica.
- Se define como la probabilidad de clasificar correctamente un par de individuos sano y enfermo, seleccionados al azar de la población, mediante los resultados obtenidos al aplicarles la prueba diagnóstica.
- Es la probabilidad de que el resultado de la prueba se presente más elevado en el paciente enfermo

Área bajo la curva ROC = Probabilidad ( $X_{\text{enfermo}} > X_{\text{sano}}$ )

- *En el ejemplo del PSA anterior un área de 0,83 significa que un individuo seleccionado al azar del grupo de enfermos tendrá el 83% de las veces un valor de PSA cuantitativamente mayor que el PSA de un individuo elegido al azar del grupo no enfermo (sanos).*

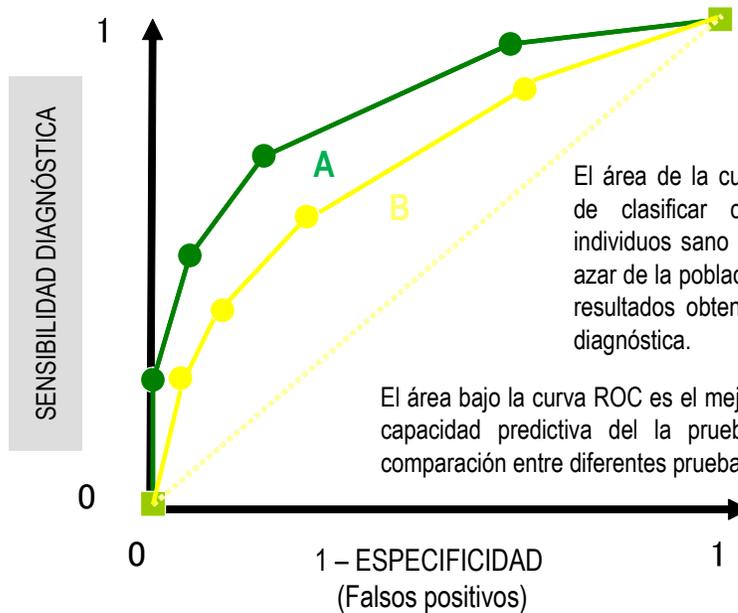
Hanley JA, McNeil BJ. The meaning and use of the area under a Receiver Operating Characteristic (ROC) Curve. Radiology 1982; 143: 29-36.

## Interpretación de Swets del área bajo la curva ROC

- El área bajo la curva ROC es siempre mayor o igual que 0,5. Toma valores comprendidos entre 0,5, si no existen diferencias en la distribución de resultados de la prueba entre los grupos enfermo y sano, y 1,0, cuando existe separación perfecta entre las dos distribuciones.
- Swets interpreta el valor del área de la siguiente manera:
  - Entre 0,5 y 0,7 indican baja exactitud
  - Entre 0,7 y 0,9 pueden ser útiles para algunos propósitos
  - *Valores mayor de 0,9 indica alta exactitud.*
- La capacidad de discriminación de la prueba diagnóstica puede evaluarse estimando el intervalo de confianza del área de la curva ROC. Si el intervalo no incluye el valor 0,5 la prueba es capaz de discernir entre enfermos y sanos.
- En el ejemplo del PSA anterior el área de 0,83 debe ser interpretado que la prueba puede ser útil en algunas situaciones clínicas, pero no es útil por ejemplo, para realizar programas de screening de la población en búsqueda del cáncer de próstata en la población general.

## Comparación del área bajo la curva:

### Prueba A versus prueba B

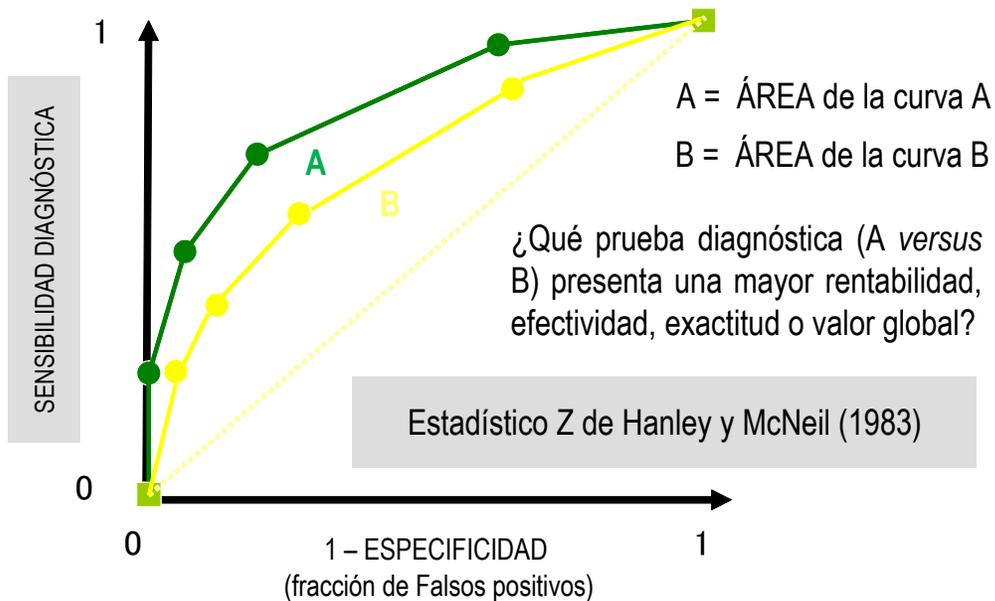


El área de la curva ROC es la probabilidad de clasificar correctamente un par de individuos sano y enfermo, seleccionados al azar de la población en estudio, mediante los resultados obtenidos al aplicarles la prueba diagnóstica.

El área bajo la curva ROC es el mejor indicador de la capacidad predictiva de la prueba, y permite la comparación entre diferentes pruebas diagnósticas.

## Comparación del área bajo la curva:

### Prueba A versus prueba B



## Comparación del área de las curvas ROC en muestras independientes

$$Z = \frac{(A_1 - A_2)}{\sqrt{(EE_{A_1})^2 + (EE_{A_2})^2}}$$

$A_1$  = área bajo la curva ROC de la técnica 1

$A_2$  = área bajo la curva ROC de la técnica 2

$EE_{A_1}$  = error estándar del área bajo la curva ROC de la técnica 1

$EE_{A_2}$  = error estándar del área bajo la curva ROC de la técnica 2

Si el estudio se realiza en grupos INDEPENDIENTES, la comparación de áreas mediante métodos estadísticos es directa. Si se utilizan datos APAREADOS o dependientes es necesario calcular previamente la correlación  $r$  entre las pruebas coeficiente de correlación entre ambas técnicas para muestras independientes vale cero ( $r = 0$ ).

## Comparación del área de las curvas ROC en muestras apareadas

$$Z = \frac{(A_1 - A_2)}{\sqrt{(EE_{A_1})^2 + (EE_{A_2})^2 - 2r * EE_{A_1} * EE_{A_2}}}$$

$A_1$  = área bajo la curva ROC de la técnica 1

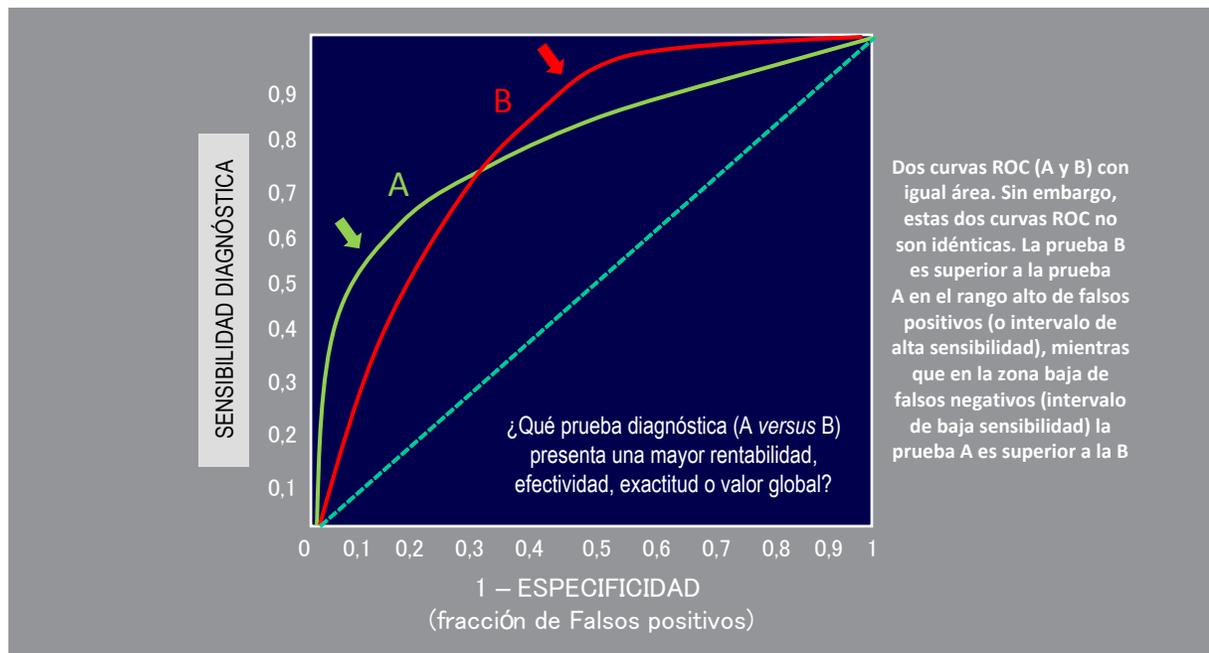
$A_2$  = área bajo la curva ROC de la técnica 2

$EE_{A_1}$  = error estándar del área bajo la curva ROC de la técnica 1

$EE_{A_2}$  = error estándar del área bajo la curva ROC de la técnica 2

$r$  = coeficiente de correlación entre las áreas de ambas técnicas para muestras dependientes

## Comparación del área bajo la curva: prueba A versus prueba B



## Pregunta de investigación

¿Qué utilidad clínica tienen los niveles de urea en sangre en el diagnóstico de la insuficiencia renal?

Objetivo: Evaluación de la uremia en el diagnóstico de la insuficiencia renal.

Técnica analítica a valorar

Uremia (mg/dl)



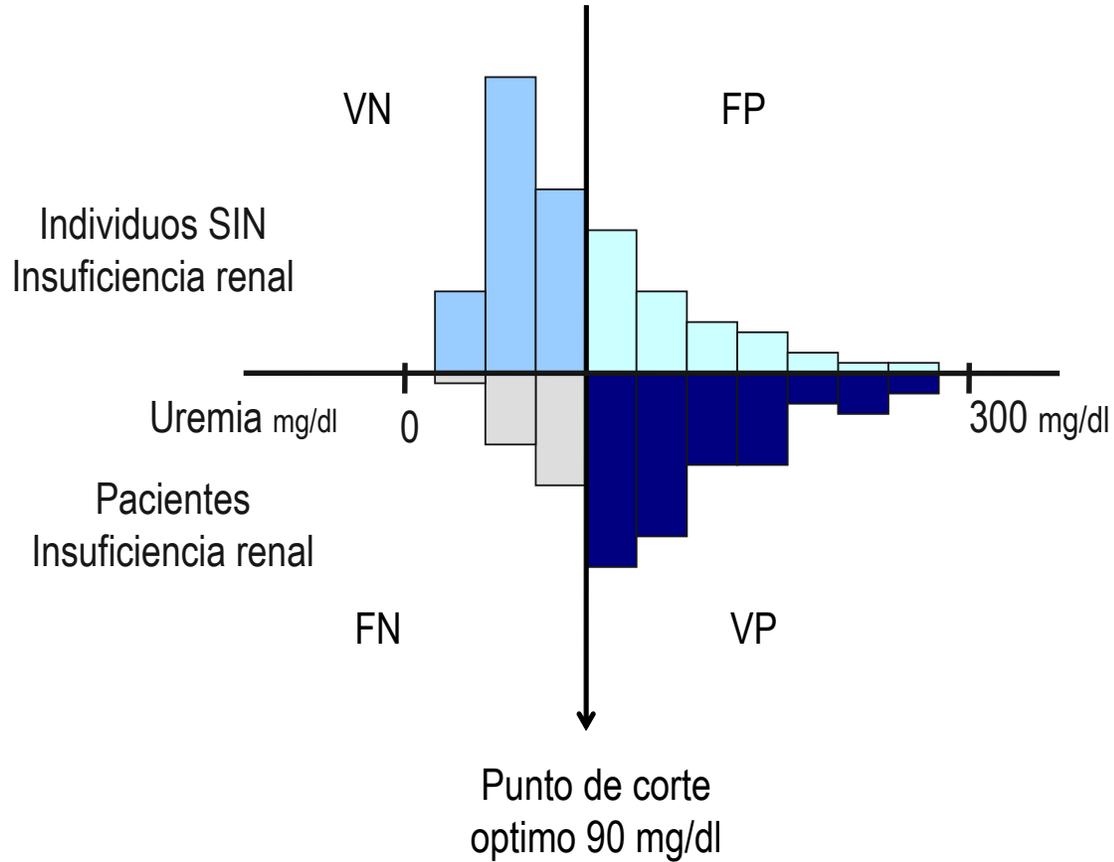
Técnica de referencia  
(patrón de oro)

Insuficiencia renal

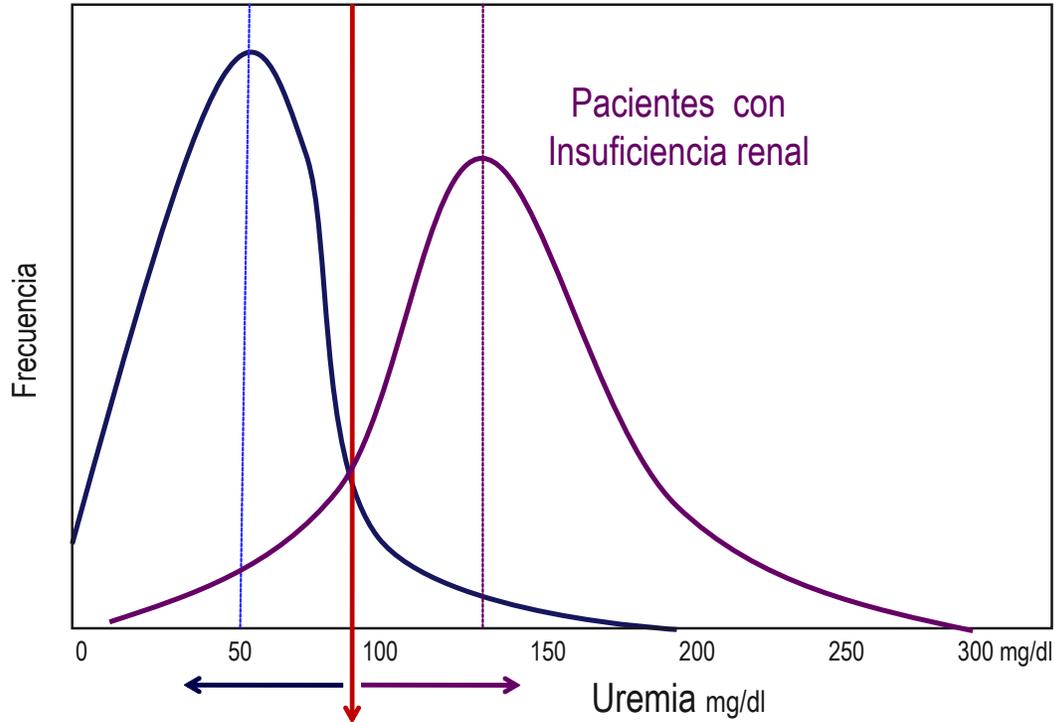
¿Cuál es el test estadístico más acertado?

## Tipos de error en las pruebas diagnósticas

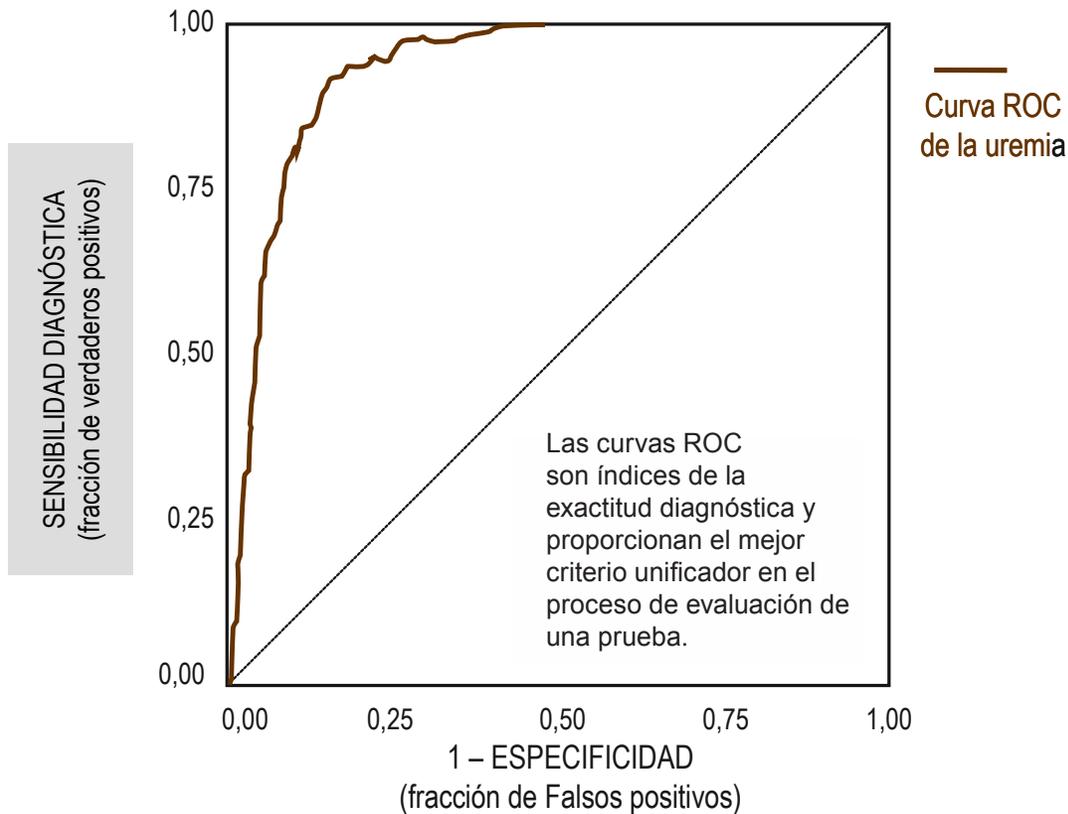
Tipos de error en las pruebas diagnósticas		Enfermedad en estudio	
		Insuficiencia renal	No enfermos
Valor del punto de corte 90 mg/dl			
Test de laboratorio uremia	Positivo (+) Uremia > 90 mg/dl	VERDADEROS POSITIVOS VP	FALSO POSITIVO FP $\alpha$
	Negativo (-) Uremia $\leq$ 90 mg/dl	FALSO NEGATIVO FN $\beta$	VERDADEROS NEGATIVOS VN



Individuos SIN  
Insuficiencia renal



**Punto de corte optimo 90 mg/dl**



## Pregunta de investigación

¿Qué técnica analítica presenta una mayor exactitud, valor global, rentabilidad, o efectividad clínica la urea o la creatinina sérica en el diagnóstico de la insuficiencia renal?

Objetivo: Evaluación de la urea frente a la creatinina en el diagnóstico de la insuficiencia renal.

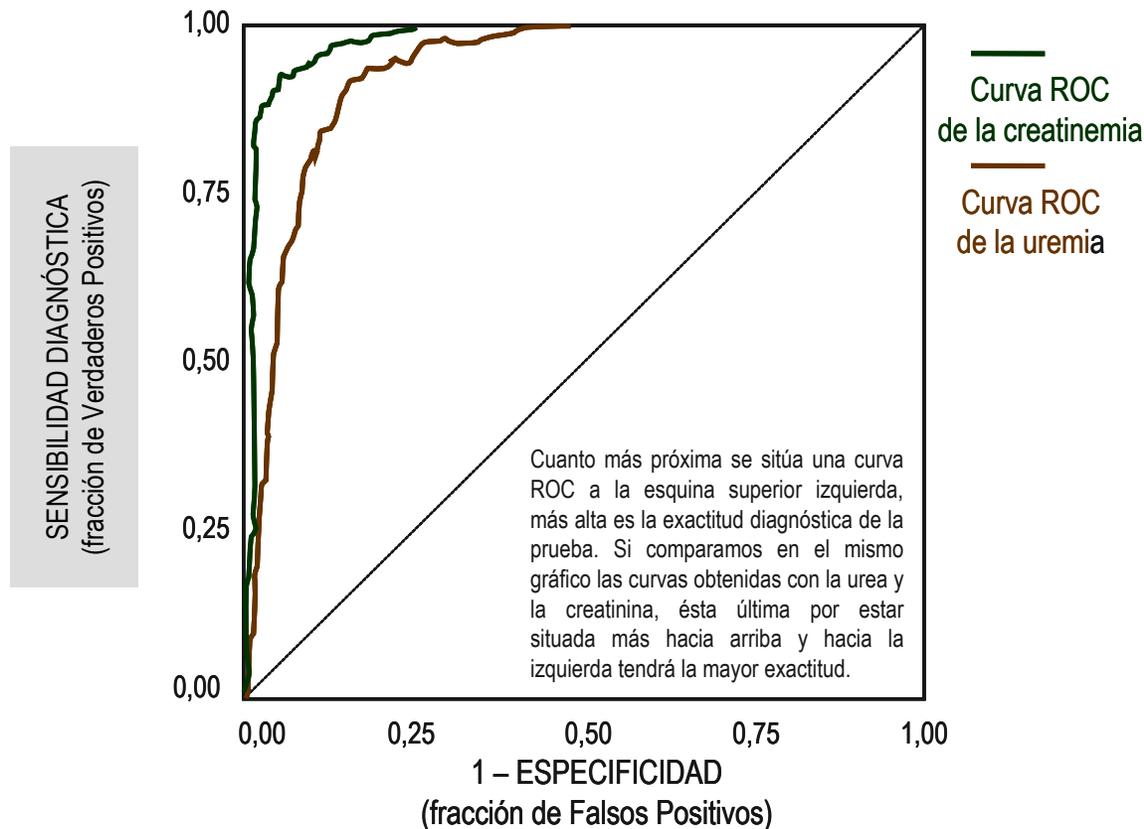
Urea (sérica)



Creatinina (sérica)

¿Cuál es el test estadístico más acertado?

Comparación del área bajo la curva ROC



## Pregunta de investigación

¿Qué prueba diagnóstica (uremia versus creatinemia) presenta una mayor rentabilidad, exactitud o valor global en el diagnóstico de la insuficiencia renal?

Variables contrastadas	Área bajo la curva ROC	Error estándar del área	Límite inferior del intervalo de confianza al 95%	Límite superior del intervalo de confianza al 95%
Uremia	0,930	0,012	0,906	0,955
Creatinemia	0,984	0,005	0,974	0,994

¿Cuál es el test estadístico más acertado?

## Comparación del área de las curvas ROC en muestras dependientes

$$Z = \frac{(A_1 - A_2)}{\sqrt{(EE_{A_1})^2 + (EE_{A_2})^2 - 2r * EE_{A_1} * EE_{A_2}}}$$

$A_1$  = área bajo la curva ROC de la técnica 1

$A_2$  = área bajo la curva ROC de la técnica 2

$EE_{A_1}$  = error estándar del área bajo la curva ROC de la técnica 1

$EE_{A_2}$  = error estándar del área bajo la curva ROC de la técnica 2

$r$  = coeficiente de correlación entre las áreas de ambas técnicas para muestras dependientes

$$Z = \frac{(A_1 - A_2)}{\sqrt{(EE_{A_1})^2 + (EE_{A_2})^2 - 2r * EE_{A_1} * EE_{A_2}}}$$

$$A_1 \text{ (creatinina)} = 0,984 \quad EE_{A1} \text{ (creatinina)} = 0,005.$$

$$A_2 \text{ (urea)} = 0,930 \quad EE_{A2} \text{ (urea)} = 0,012.$$

Hipótesis nula  $H_0: A_1 - A_2 = 0$  .

Hipótesis alternativa  $H_a: A_1 - A_2 \neq 0$  .

Resultado: El estadístico para la comparación de ambas curvas Z de Hanley y McNeil vale 4,15 que como es mayor que  $Z_{95\%} = 1,96$  .

Conclusión: Ambas pruebas tienen un rendimiento significativamente distinto.

## Pregunta de investigación

¿Qué utilidad clínica tienen los niveles de VCM en el diagnóstico de la anemia ferropénica?

Objetivo: Evaluación del VCM en el diagnóstico de la anemia ferropénica.

### Técnica analítica a valorar

VCM  
(hemograma)



### Técnica de referencia (patrón de oro)

Depósitos de Fe en médula ósea

¿Cuál es el test estadístico más acertado?

## Tipos de error en las pruebas diagnósticas

Tipos de error en las pruebas diagnósticas		Examen de los depósitos de hierro en médula ósea	
		Déficit de Fe	Presencia de Fe
Test de laboratorio VCM (hemograma)	Positivo (+) VCM BAJO (microcitosis)	VERDADEROS POSITIVOS VP	FALSO POSITIVO FP $\alpha$
	Negativo (-) VCM NORMAL	FALSO NEGATIVO FN $\beta$	VERDADEROS NEGATIVOS VN

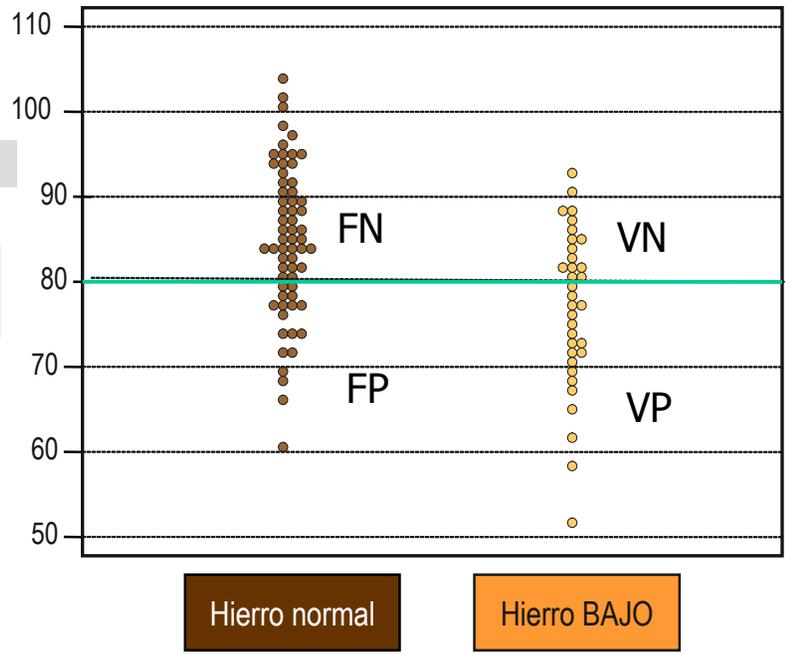
## VCM en 100 pacientes con diagnóstico de anemia ferropénica

Técnica de referencia: Examen de los depósitos de hierro en médula ósea	Volumen corpuscular medio (VCM)
Ausencia de Hierro (n= 33)	52, 58, 62, 65, 67, 68, 70, 72, 72, 73, 73, 74,75, 77, 77, 77, 78, 78, 80, 80, 81, 81, 81, 81, 84, 84, 85, 85, 87, 88, 88, 92, 92
Presencia de hierro (n=67)	60, 66, 68, 69, 71, 71, 71, 73, 74, 74, 74, 76,76, 77, 77, 77, 78, 79, 79, 79, 80, 80, 81, 81,82, 82, 82, 83, 83, 83, 83, 83, 83, 83, 83, 83, 84,84, 84, 85, 85, 86, 86, 86, 86, 88, 88, 88, 88, 89,89, 89, 90, 90, 90, 91, 91, 93, 93, 93, 94, 94,94, 94, 96, 97, 98, 100, 103

Beck JR, Shultz EK. The use of relative operating characteristic (ROC) curves in test performance evaluation. Arch Pathol Lab Med 1986; 110:13-20.

Técnica analítica a evaluar

**VCM  
(hemograma)**



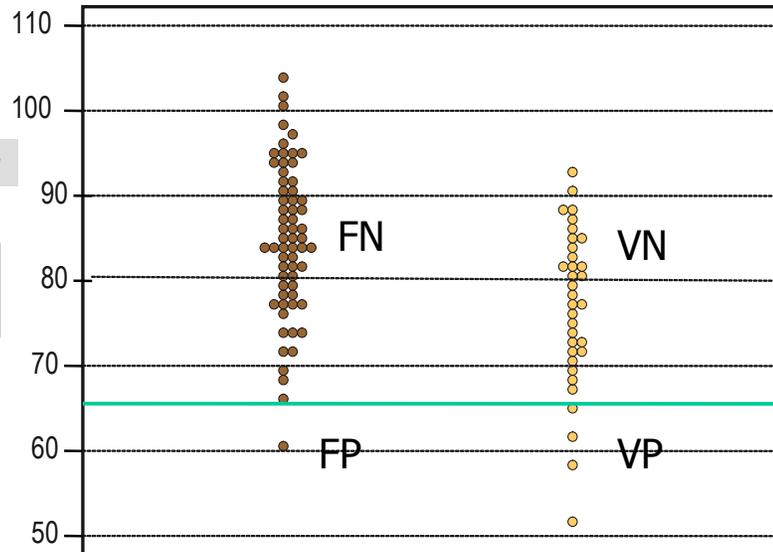
VCM < 80  
Sensibilidad = 55,9%  
Especificidad = 71,2%

Examen de los depósitos de hierro en médula ósea

Técnica de referencia (patrón de oro)

Técnica analítica a evaluar

**VCM**  
(hemograma)



Hierro normal

Hierro BAJO

VCM < 65

Sensibilidad = 8,8%

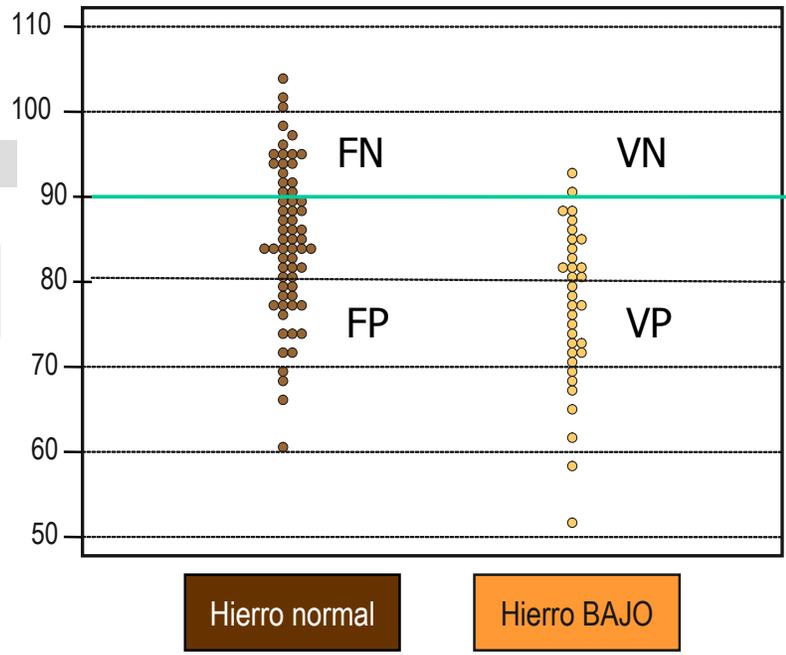
Especificidad = 98,5%

Examen de los depósitos de hierro en médula ósea

Técnica de referencia (patrón de oro)

Técnica analítica a evaluar

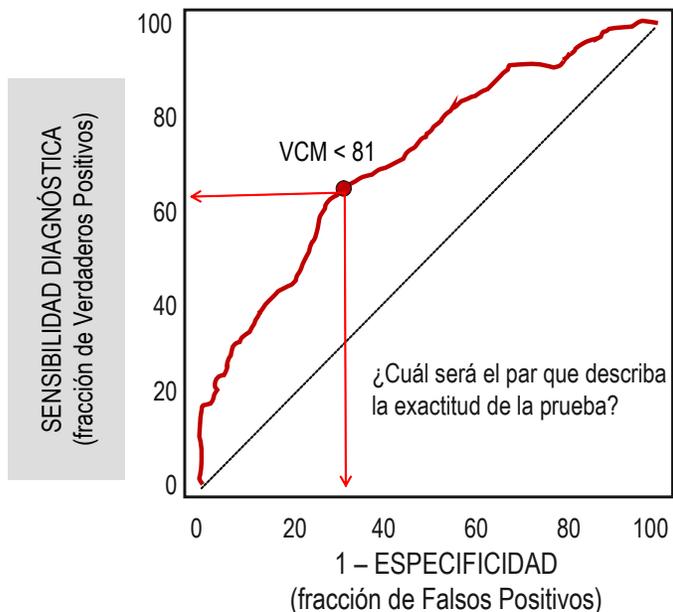
**VCM**  
(hemograma)



VCM < 90  
Sensibilidad = 94,1%  
Especificidad = 25,8%

Examen de los depósitos de hierro en médula ósea

Técnica de referencia (patrón de oro)



Las pruebas diagnósticas no tienen un solo par sensibilidad/especificidad (S/E), sino muchos, uno por cada nivel de decisión o punto de corte que se adopte.

**Punto de corte ideal para el VCM** es de  $< 81$ , valor límite que permite resumir los resultados de una escala continua en dos categorías: positivo y negativo con una mayor exactitud o rentabilidad diagnóstica.

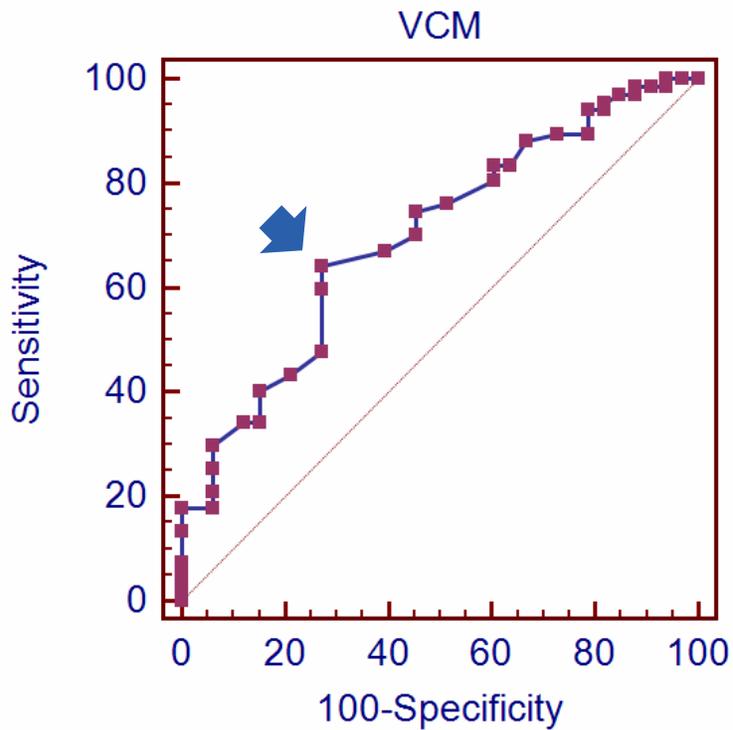
VCM  $< 81$

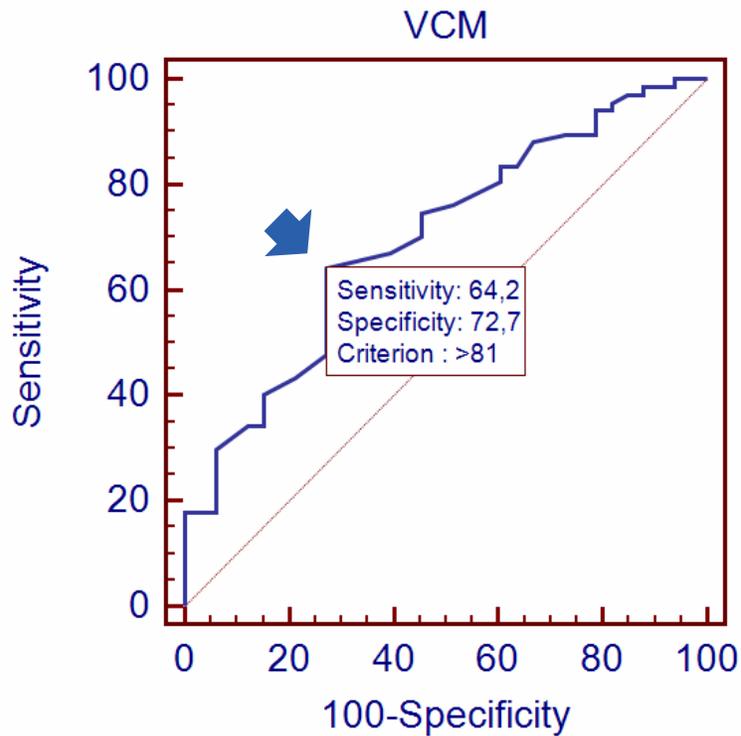
Sensibilidad = 61%

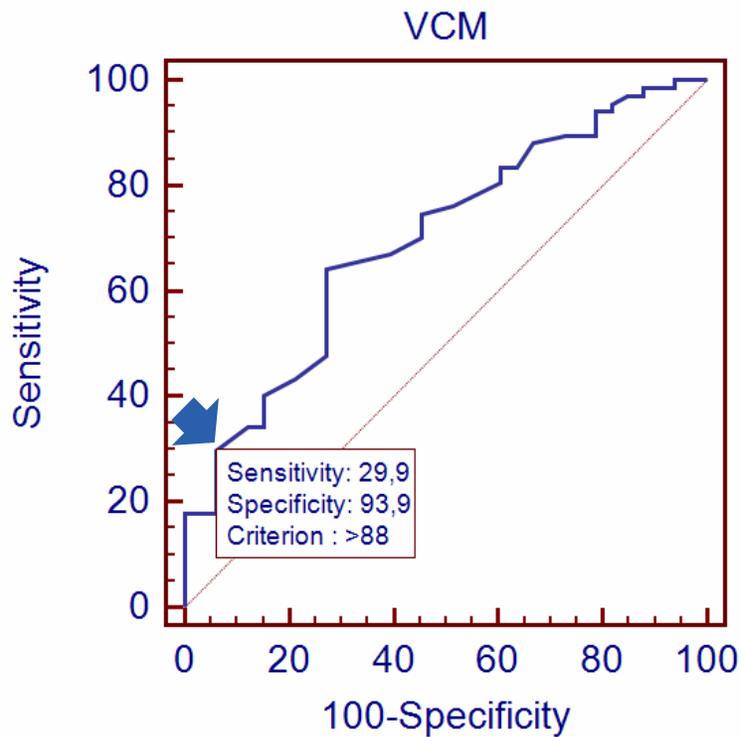
Fracción de FP = 30%

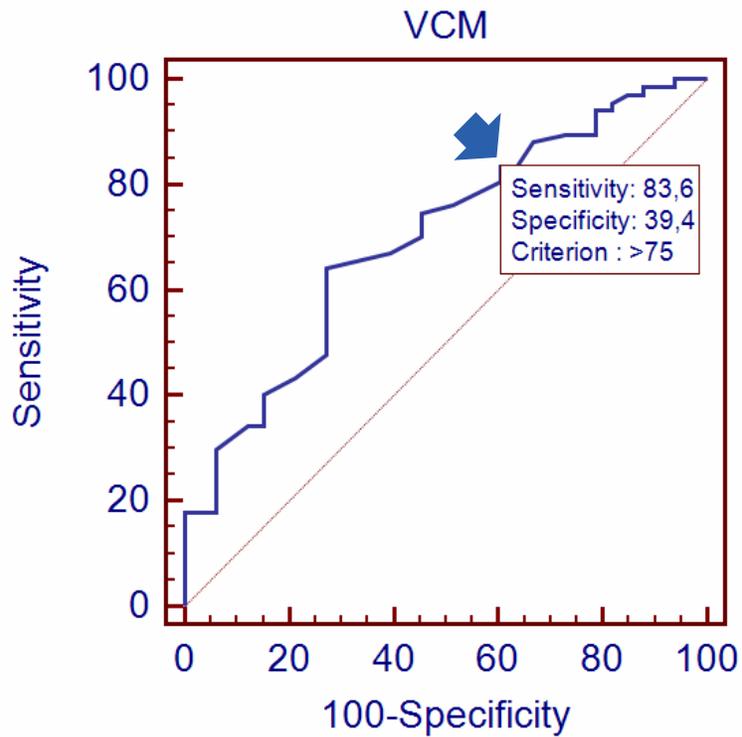
Especificidad = 70%

El cálculo del área bajo la curva es de 0,717 con un EE de 0,05, es decir, el VCM es una prueba diagnóstica de baja rentabilidad, y por tanto no se aconseja, como prueba única, en el diagnóstico de la anemia ferropénica.









## Pregunta de investigación

¿Qué técnica analítica presenta una mayor rentabilidad, el VCM o la ferritina en el diagnóstico de la anemia ferropénica?

Objetivo: Evaluación del VCM frente a la ferritina en el diagnóstico de la anemia ferropénica.

VCM  
(hemograma)



Ferritina (sérica)

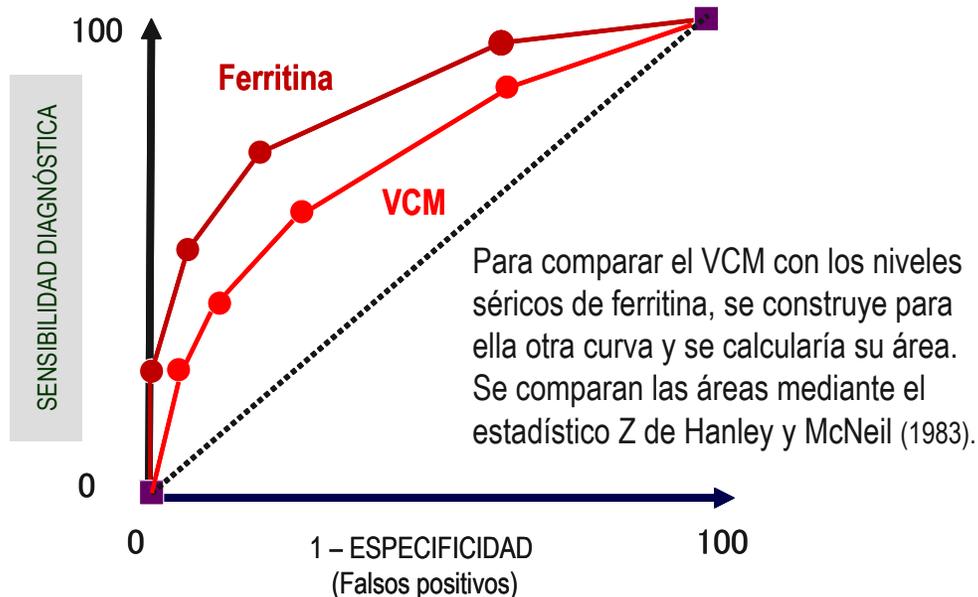
¿Cuál es el test estadístico más acertado?

Comparación del área bajo la curva ROC

Punto Corte VCM	Sensibilidad (%)	Especificidad (%)	100 – especificidad (%)
65	8,8	98,5	1,5
70	20,6	93,9	6,1
75	38,2	84,8	15,2
80	55,9	71,2	28,8
85	79,4	43,9	56,1
90	94,1	25,8	74,2
92	97,1	19,7	80,3

Construcción de la curva ROC: Es un gráfico que muestra todos los pares sensibilidad/especificidad resultantes de la variación continua de los puntos de corte en todo el rango de resultados observados.

## Comparación del área bajo la curva: ferritina sérica versus VCM



Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983; 148: 839-843.

## Comparación del Área de las curvas ROC en muestras dependientes

$$Z = \frac{(A_1 - A_2)}{\sqrt{(EE_{A_1})^2 + (EE_{A_2})^2 - 2r * EE_{A_1} * EE_{A_2}}}$$

$A_1$  = área bajo la curva ROC de la técnica 1

$A_2$  = área bajo la curva ROC de la técnica 2

$EE_{A_1}$  = error estándar del área bajo la curva ROC de la técnica 1

$EE_{A_2}$  = error estándar del área bajo la curva ROC de la técnica 2

$r$  = coeficiente de correlación entre las áreas de ambas técnicas para muestras dependientes

Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983; 148: 839-843.

$$Z = \frac{(A_1 - A_2)}{\sqrt{(EE_{A_1})^2 + (EE_{A_2})^2 - 2r * EE_{A_1} * EE_{A_2}}}$$

$$A_1 \text{ (creatinina)} = 0,984 \quad EE_{A1} \text{ (creatinina)} = 0,005.$$

$$A_2 \text{ (urea)} = 0,930 \quad EE_{A2} \text{ (urea)} = 0,012.$$

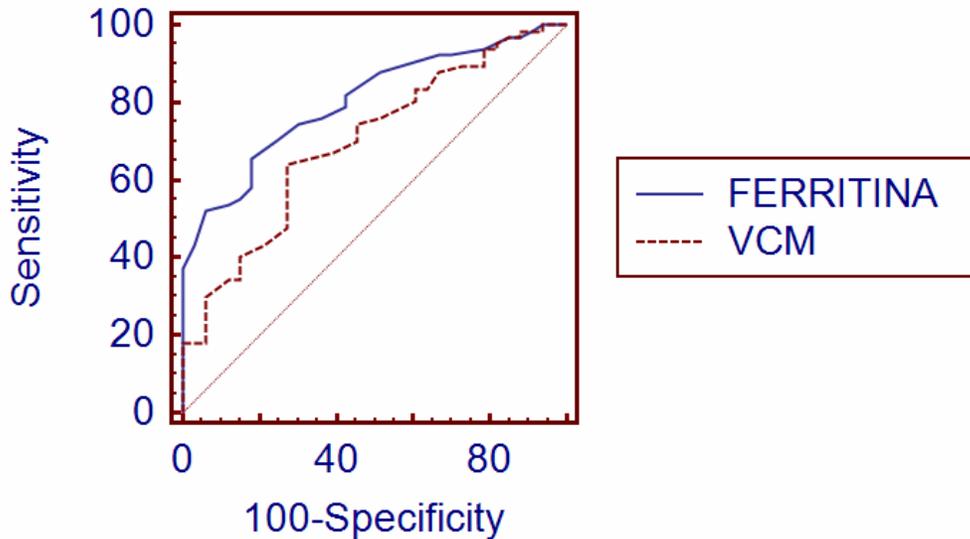
Hipótesis nula  $H_0: A_1 - A_2 = 0$  .

Hipótesis alternativa  $H_a: A_1 - A_2 \neq 0$  .

Resultado: El estadístico para la comparación de ambas curvas Z de Hanley y McNeil vale 4,15 que como es mayor que  $Z_{95\%} = 1,96$  .

Conclusión: Ambas pruebas tienen un rendimiento significativamente distinto.

## Comparación del área bajo la curva: ferritina sérica versus VCM



## Elección del punto de corte:

### Criterio basado en el equilibrio de riesgos y beneficios de las decisiones médicas

Es un enfoque práctico que se basa en el cociente entre los costes de un resultado falso positivo (FP) frente a un falso negativo (FN), en función de la prevalencia (P) de la enfermedad. El valor de corte óptimo se determina hallando el punto de la curva ROC cuya tangente a la curva tiene de pendiente el valor de m.

$$m = \frac{\text{Coste de FP}}{\text{Coste de FN}} * \frac{1 - P}{P}$$

Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 1993; 39: 561-577.

McNeil BJ, Keeler E, Adelstein SJ. Primer on certain elements of medical decision making. N Engl J Med 1975; 293: 211-215.

# Elección del punto de corte:

## Postulados de Galen y Gambino

### I. Criterio de máxima SENSIBILIDAD DIAGNÓSTICA

Se elige la mayor sensibilidad posible cuando:

1. La enfermedad sea grave y no pueda pasar inadvertida.
2. La enfermedad sea tratable, y
3. Los resultados falsos positivos no supongan un traumatismo psicológico o económico en los individuos examinados.

¿Podrías poner algún ejemplo?

## Cribado poblacional de cáncer de mama (CaM)

Programa preventivo destinado a la detección precoz de CaM mediante la realización de mamografía bilateral con periodicidad bienal en mujeres de 45 a 70 años. La realización del screening permite la detección precoz de CaM en un estadio menos avanzado que el que correspondería a un diagnóstico clínico, lo que supone una considerable mejoría en cuanto a la afectación, pronóstico de la enfermedad y reducción de años de vida perdidos.

Variable **PREDICTORA**  
Cualitativa dicotómica

Mamografía bilateral



Variable **RESULTADO**  
Cualitativa dicotómica

Cáncer de mama

## ¿Cumple los postulados de Galen y Gambino?

En España se ha descrito que la tasa acumulada de falsos positivos en un programa de cribado de mama es del 30% tras 10 determinaciones bianuales (1). De hecho, diversos estudios han demostrado que, por cada muerte prevenida a través de un programa de cribado, han aparecido al menos 200 resultados falsos positivos (2). En respuesta a este hecho, la OMS ha incidido en nuevos criterios para los programas de cribado, centrándose en aspectos éticos, en las consecuencias psicológicas de resultados falsos positivos y en la necesidad del consentimiento informado (3).

1. Castells X, Molins E, Macia F. Cumulative false positive recall rate and association with participant related factors in a population based breast cancer screening programme. *J Epidemiol Community Health*. 2006;60:316-21.
2. Raffle AE, Alden B, Quinn M, Babb PJ, Brett MT. Outcomes of screening to prevent cancer: analysis of cumulative incidence of cervical abnormality and modelling of cases and deaths prevented. *BMJ*. 2003;326:901.
3. Brodersen J, McKenna SP, Doward LC, Thorsen H. Measuring the psychosocial consequences of screening. *Health Qual Life Outcomes*. 2007;5:3.

## Cribado poblacional del cáncer de pulmón (CP) mediante TC

Variable PREDICTORA  
Cualitativa dicotómica

Tomografía computarizada (TC)



Variable RESULTADO  
Cualitativa dicotómica

Cáncer de pulmón

¿Cumple los postulados de Galen y Gambino?

Un estudio publicado por Bach et al (2007) acerca del cribado de cáncer de pulmón señalaba que el uso de la tomografía computarizada (TC) puede aumentar la tasa de diagnóstico y tratamiento de dicho cáncer sin reducir el riesgo de avance de la enfermedad o la mortalidad. Por lo tanto, concluían la inutilidad del cribado en la prevención del cáncer de pulmón.

Bach PB, Jett JR, Pastorino U, Tockman MS, Swensen SJ, Begg CB. Computed tomography screening and lung cancer outcomes. *JAMA*. 2007;297:953-61.

## Cribado poblacional de cáncer de colon

Programa preventivo destinado a la detección precoz de colon

Variable PREDICTORA  
Cualitativa dicotómica

Detección de SANGRE OCULTA en  
heces (guayacol)



Variable RESULTADO  
Cualitativa dicotómica

Cáncer de colon

Sensibilidad = 40-80%

Especificidad = 90-99%

La sensibilidad de la prueba es tan sólo de 50–60% cuando se lo utiliza una vez, pero puede ascender a 90% cuando se la utiliza cada 1–2 años en un período de tiempo prolongado (sensibilidad programática).

## Cribado poblacional de cáncer de colon

### Programa preventivo destinado a la detección precoz de colon

El cáncer de colon es el tipo de cáncer con mayor posibilidad de detección precoz y que responde mejor al tratamiento en estadios iniciales, por ello se propone una prueba genética a la población diana: hombres y mujeres a partir de los 50 años..

**Variable PREDICTORA**  
Cualitativa dicotómica

Detección de la forma metilada del gen Septina 9 (SEPT9).



**Variable RESULTADO**  
Cualitativa dicotómica

Cáncer de colon

Sensibilidad = 67%

Especificidad = 88%

## ¿Cumple los postulados de Galen y Gambino?

Los resultados FALSOS POSITIVOS causan ansiedad y pueden conducir a un aumento de pruebas invasivas para determinar si hay enfermedad realmente, en este caso una prueba positiva a la detección de sangre oculta en heces será seguida por una colonoscopia.

## Falsos positivos versus Falsos negativos

1. Gravedad o importancia en salud pública. Si la enfermedad puede ser tratada y detectada evitando complicaciones posteriores han de utilizarse pruebas con alta SENSIBILIDAD, es decir con baja tasa de FALSOS NEGATIVOS.
2. Si el coste del proceso asistencial (posibles pruebas invasivas complementarias, exámenes y cuidados médicos y de enfermería requeridos ) del enfermo es alto se deben elegir pruebas con alta ESPECIFICIDAD, para reducir los falsos positivos.
3. Si la enfermedad diagnosticada conlleva consecuencias psicosociales graves, como la infección por VIH, han de minimizarse los falsos positivos, eligiendo pruebas de alta ESPECIFICIDAD.
4. Baja prevalencia de la enfermedad. En los programas de screening se eligen pruebas de alta ESPECIFICIDAD, para reducir la tasa de falsos positivos y así garantizar la eficiencia del programa.

## ORIGINALES

---

### Iatrogenia atribuible a las pruebas diagnósticas

**Blanca Lumbreras e Idefonso Hernández-Aguado**

*CIBER en Epidemiología y Salud Pública (CIBERESP). Departamento de Salud Pública, Historia de la Ciencia y Ginecología. Universidad Miguel Hernández. Elche. Alicante. España.*

Correspondencia: Dra. B. Lumbreras.

Departamento de Salud Pública, Historia de la Ciencia y Ginecología. Facultad de Medicina. Universidad Miguel Hernández.

Elche. Alicante. España.

Correo electrónico: blumbreras@umh.es

“Las carencias metodológicas en la investigación evaluativa de pruebas diagnósticas explican en parte la inadecuación del uso de pruebas diagnósticas. Por otra parte, no se ha considerado la potencial iatrogenia atribuible al uso de pruebas diagnósticas, y se hace necesaria una monitorización de los posibles efectos adversos derivados de cualquier tecnología diagnóstica, así como la información detallada a los pacientes de sus beneficios y riesgos, tanto en el campo del cribado como en el del diagnóstico clínico”.

## Elección del punto de corte:

### Postulados de Galen y Gambino

#### 2. Criterio de máxima ESPECIFICIDAD DIAGNÓSTICA

Se elige la mayor especificidad posible cuando:

1. La enfermedad sea importante pero difícil de curar o incurable, y
2. El hecho de conocer que no se padezca la enfermedad tiene una importante carga sanitaria y psicológica.

¿Podrías poner algún ejemplo?

## Elección del punto de corte:

### Postulados de Galen y Gambino

#### 3.- Criterio de alto VALOR PREDICTIVO POSITIVO

Se elige un alto VPP cuando:

El tratamiento de los falsos positivos (FP) pueda tener consecuencias graves en personas que NO tienen la enfermedad.

$$\text{VPP} = \frac{\text{VP}}{\text{VP} + \text{FP}}$$

Galen RS, Gambino SR. Beyond normality: The predictive value and efficiency of medical diagnoses. Nueva York: J Wiley and Sons, Inc., 1975.

## Elección del punto de corte:

### Postulados de Galen y Gambino

#### 4. Criterio de máxima EXACTITUD DIAGNÓSTICA

Se desea una elevada exactitud (efectividad, valor global o rentabilidad diagnóstica) cuando:

1. La enfermedad sea importante pero curable, y
2. Tanto los falsos positivos (FP) como los falsos negativos (FN) supongan un traumatismo y conlleven consecuencias graves.

## Ventajas de las curvas ROC en la evaluación de pruebas diagnósticas

1. Son una representación gráfica fácilmente comprensible de la capacidad de discriminación de la prueba diagnóstica en todo el rango de puntos de corte.
2. Son simples, gráficas y fáciles de interpretar visualmente.
3. No requieren un nivel de decisión particular porque está incluido todo el espectro de puntos de corte.
4. Son independientes de la prevalencia, ya que ni la sensibilidad ni la especificidad dependen de ella.
5. Proporcionan una comparación visual directa entre distintas pruebas en una escala común, mientras que otro tipo de gráficos, como los diagramas de puntos o los histogramas de frecuencias requieren diferentes gráficos cuando difieren las escalas.
6. La especificidad y la sensibilidad son fácilmente accesibles en el gráfico, en contraste con los diagramas de puntos y los histogramas.

Zweig MH, Campbell G. Receiver-Operating Characteristic (ROC) Plots: fundamental evaluation tool in clinical medicine. Clin Chem 1993; 39: 561-577.

Burgueño MJ, J.L. García-Bastos JL, González-Buitrago JM. Las curvas ROC en la evaluación de las pruebas diagnósticas. Med Clin (Barc) 1995; 104: 661-670.

## Capítulo 5

Publicación Bioestadística: Análisis bivariante.  
Test de contraste de hipótesis.  
Cualitativa versus cualitativa.

Dr. Jacobo Díaz Portillo

Coordinador de Formación Continuada

Hospital Universitario de Ceuta

# Materiales y métodos

## Estructura

1. Tipo de diseño (estudio)
2. Sujetos incluidos (criterios de selección de la muestra)
3. Fuentes de información
4. Descripción de la *intervención* y definición del punto final de la investigación (estudios experimentales)
5. *Instrumentos* y procedimientos de medida (cuestionarios)
6. Definición de las *variables* y covariables incluidas en el estudio (factor de estudio y criterio de evaluación)
7. Limitaciones y consideraciones éticas de la investigación
8. *Pruebas estadísticas a utilizar (programa estadístico)*

## Fases del análisis estadístico

1. ANÁLISIS DESCRIPTIVO



2. ANÁLISIS BIVARIANTE



3. ANÁLISIS MULTIVARIANTE

## Decisiones que se pueden tomar en un test de contraste de hipótesis después de conocer el valor de significación estadística $p$

$p < 0,05$ $A \neq B$	$p \geq 0,10$ $A = B$
Se rechaza la hipótesis nula	No se puede rechazar la hipótesis nula
No parece que el azar pueda explicarlo todo	No se puede descartar que el azar lo explique todo
El EFECTO observado es mayor que el ERROR	El EFECTO observado es similar que el ERROR
Hay diferencias estadísticamente significativas	NO hay diferencias estadísticamente significativas
Existen evidencias a favor de la hipótesis alternativa $H_a$ , y por tanto rechazamos la hipótesis nula $H_0$	NO existen evidencias a favor de la hipótesis alternativa $H_a$ , y por tanto aceptamos la hipótesis nula $H_0$

Los límites de 0,05 y 0,1 pueden ser considerados en cierto modo arbitrarios y aproximados

## Análisis estadístico

Tipos de test analíticos estadísticos de contraste de hipótesis según el tipo de variables implicadas en el estudio

Variable dependiente

Variable resultado

		Dicotómica	Continua
Variable Predictora	Dicotómica	Test de $\chi^2$	t de Student
Variable Independiente	Continua	Regresión logística	Regresión lineal

## Análisis estadístico

ANÁLISIS ESTADÍSTICO		Variable resultado	
		Dicotómica	Continua
Variable Predictora	Dicotómica	Test de $\chi^2$	t de Student
Variable Independiente	Continua	Regresión logística	Regresión lineal

Test de  $\chi^2$

# Tipos de test analíticos estadísticos bivariantes de contraste de hipótesis más frecuentes

Variable PREDICTORA o Independiente (factor de estudio)	Variable RESULTADO o Dependiente (Criterio de evaluación)			
	Cualitativa nominal dos categorías	Cualitativa nominal > 2 categorías	Cualitativa ordinal o Cuantitativa NO paramétrica	Cuantitativa paramétrica
Cualitativa dos grupos independientes	Chi cuadrado Z comparación de proporciones Test exacto de Fisher	Chi cuadrado	U de Mann-Whitney	T de Student (prueba de Welch)
Cualitativa dos grupos apareados	Test de McNemar	Q de Cochran	Prueba de los rangos de Wilcoxon	T de Student para datos emparejados
Cualitativa > 2 grupos independientes	Chi cuadrado	Chi cuadrado	Prueba de Kruskal-Wallis	Análisis de la Varianza (ANOVA)
Cualitativa > 2 grupos apareados	Q de Cochran	Q de Cochran	Prueba de Friedman	Análisis de la Varianza de dos vías
Cuantitativa	Regresión logística	Regresión logística	Correlación de Spearman Tau de Kendall	Regresión lineal Correlación de Pearson

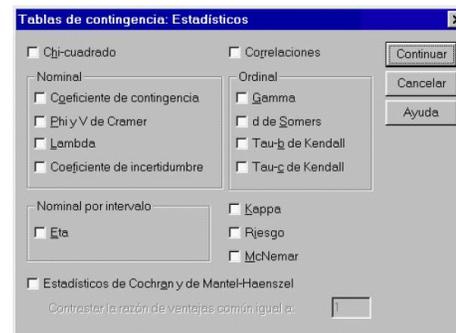
## Pruebas estadísticas de contraste de hipótesis más aplicables a cada situación

Variable Independiente o PREDICTORA	Variable Dependiente o RESULTADO	Pruebas de contraste de hipótesis	Observaciones y/o limitaciones del test
Cualitativa	Cualitativa	Ji cuadrado Prueba exacta de Fisher Test de McNemar Q de Cochran REGRESIÓN LOGÍSTICA	Si el tamaño muestral es grande Si el tamaño muestral es pequeño Dos grupos de muestras repetidas o apareadas Más de dos grupos de muestras repetidas o apareadas Varias variables predictoras cualitativas
Cualitativa	Cuantitativa	t de Student Análisis de la Varianza Test de Mann-Whitney Test de Wilcoxon Test de Kruskal-Wallis Test de Friedman	Compara medias entre dos grupos independientes Compara medias entre más de dos grupos independientes Dos grupos independientes No paramétricos Dos grupos medidas repetidas No paramétrico Más de dos grupos No paramétrico, medidas repetidas Más de dos grupos No paramétrico medidas repetidas
Cuantitativa	Cualitativa	Regresión logística simple Regresión logística multivariante	Una sola variable predictora cuantitativa Varias variable predictoras cuantitativa o cualitativas
Cuantitativa	Cuantitativa	Regresión lineal simple Correlación de Pearson Correlación de Sperman REGRESIÓN MÚLTIPLE	Paramétrico Paramétrico No paramétrico Varias variables predictoras cuantitativas
Cualitativa	Supervivencia	Kaplan-Meier Long-Rank Regresión de COX	Curvas de supervivencia Compara curvas de supervivencia Análisis Multivariante

## Datos categóricos:

### Cualitativa versus cualitativa. Nominales

- Test de  $\chi^2$  de Pearson
  - Relación entre dos variables cualitativas nominales o policotómicas
  - Comparación de proporciones
- Corrección de Yates
- Test exacto de Fisher
  - Valores esperados menores de 5
- Razón de verosimilitud
  - Evaluación de pruebas diagnósticas
- Test de  $\chi^2$  de McNemar
  - Datos apareados o emparejados
- Coeficiente PHI ( $\Phi$ )
- V de Cramer
- Coeficiente de contingencia
- Coeficiente de incertidumbre
- Tau de Goodman y Kruskal
- Lambda de Goodman y Kruskal (coeficiente de predictibilidad de Guttman)

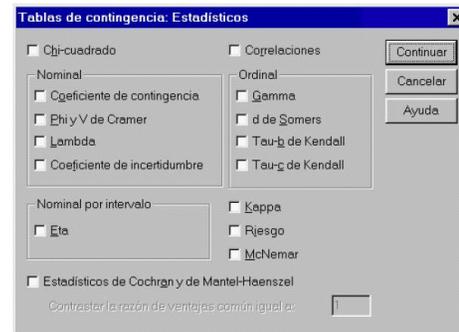


## Datos categóricos:

### Cualitativa versus cualitativa. Ordinales

- Test de  $\chi^2$  de tendencia lineal
- Tau-B de Kendall o Stuart
- Tau-C de Kendall
- Coeficiente Gamma Y de Goodman y Kruskal
- D de Somers
- Correlación de Spearman

Estos estadísticos NO se pueden utilizar cuando alguna de las variables sea NOMINAL



## Test de $\chi^2$ de Pearson

Variable RESULTADO			
Variable PREDICTORA	(+)	(-)	Total
Si	a	b	a + b
No	c	d	c + d
Total	a + c	b + d	n

$$\chi^2 = \frac{n (a d - b c)^2}{(a + b) (c + d) (a + c) (b + d)}$$

Grados de libertad = [ K - 1 ] [ F - 1 ]

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_o}$$

## Test de $\chi^2$ de Pearson

$$\text{Grados de libertad} = [K - 1] [F - 1]$$

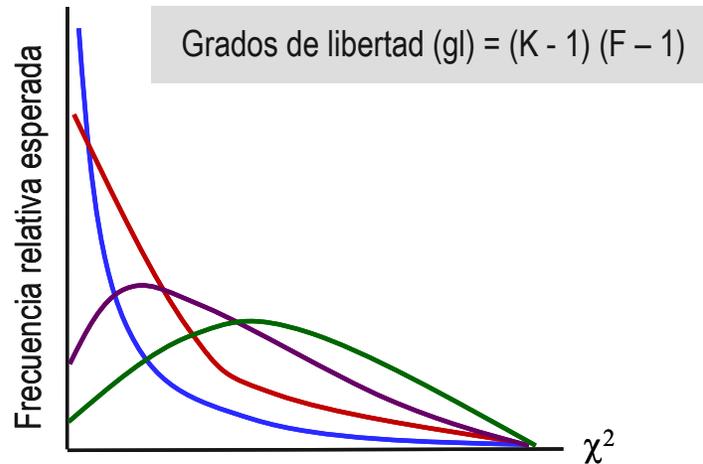
Tabla de valores críticos (tabulados) de  $\chi^2$

Grados de libertad	Tabla 2 x F	Valores críticos con un nivel de confianza del 95%
1	2	3,84
2	3	5,99
3	4	7,81
4	5	9,49
5	6	11,07
6	7	12,59
100	99	124,34

K = nº de columnas

F = nº de filas

## Influencia de los Grados de libertad $k$ en la distribución de probabilidades de $\chi^2$ de Pearson



$\chi^2$  de Pearson para  $gl=1$

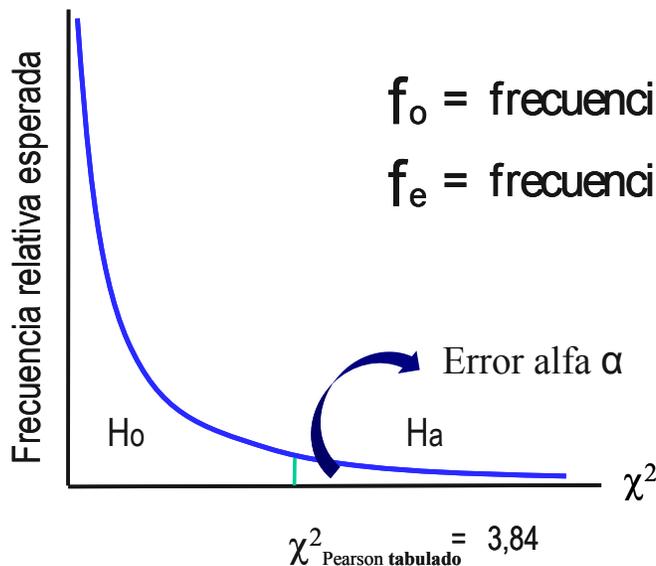
$\chi^2$  de Pearson para  $gl=3$

$\chi^2$  de Pearson para  $gl=2$

$\chi^2$  de Pearson para  $gl=6$

$$\chi^2 = \sum_{i=1}^n \frac{(Obs_i - Esp_i)^2}{Esp_i}$$

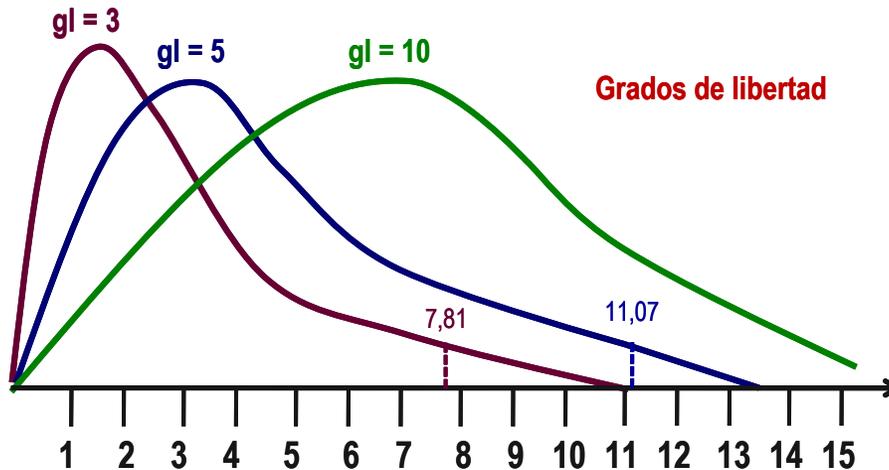
$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$



# Influencia de los Grados de libertad $k$ en la distribución de probabilidades de $\chi^2$ de Pearson

Grados de libertad (gl) =  $(K - 1) (F - 1)$

$K = n^\circ$  de columnas  
 $F = n^\circ$  de filas



## Test de $\chi^2$ de Pearson

### Condiciones necesarias y suficientes de aplicación

1. Ambas variables deben ser cualitativas en escala NOMINAL.
2. Número de sujetos estudiados igual o superior a 40 ( $n > 40$ )
3. Ninguno de los valores esperados en cada casilla debe ser menor de 5
4. No se debe aplicar cuando la variable dependiente o resultado es ORDINAL
5. Es un test no dirigido (test de planteamiento bilateral), que nos indica si existe o no relación entre dos variables cualitativas pero no en qué sentido se produce tal, ni cuantifica la intensidad de dicha asociación.

La prueba permite comprobar si dos variables cualitativas están o no asociadas. Si al final del estudio concluimos que las variables no están relacionadas podremos decir que ambas variables son independientes, con un determinado nivel de confianza, generalmente del 95%.

## Pregunta de investigación

¿Qué técnica quirúrgica es preferible en pacientes que deben ser intervenidos de colecistectomía vía laparoscópica?

Variable PREDICTORA  
Cualitativa dicotómica

- KTP/532 láser
- Electrocirugía monopolar



Variable RESULTADO  
Cualitativa dicotómica

Desarrollo de complicaciones graves postquirúrgicas

¿Cuál es el test estadístico más adecuado?

Lane GE, Lathrop JC. Comparison of results of KTP/532 laser versus monopolar electro-surgical dissection in laparoscopic cholecystectomy.  
J Laparoendosc Surg 1993; 3: 209-214.

¿Existe diferencia significativa respecto del porcentaje de complicaciones graves entre ambas técnicas de disección?

$H_0$  (hipótesis nula) = No hay diferencia entre ambas técnicas de disección.

$H_a$  (hipótesis alternativa) = Sí existe diferencia entre ambas técnicas de disección.

KTP/532 láser



Electrocirugía monopolar

Lane GE, Lathrop JC. Comparison of results of KTP/532 laser versus monopolar electro-surgical dissection in laparoscopic cholecystectomy.

J Laparoendosc Surg 1993; 3: 209-214.

## Test de $\chi^2$ de Pearson

Variable predictora	Variable resultado		
	Complicaciones postquirúrgicas graves		
Tipo de cirugía	SI (+)	NO (-)	Total
KTP/532 láser	16 (8%) P <sub>1</sub>	183	199
Electrocirugía monopolar	11 (2,6%) P <sub>2</sub>	411	422

$\chi^2_{\text{Pearson}} = 9,60 \rightarrow p = 0,0019$  Existe asociación

Si  $[P_1 - P_2]$  es mayor que el producto de 1,96 ( $Z_{95\%}$  error  $\alpha = 0,05$ ) multiplicado por el error estándar, concluimos que la diferencia es significativa. Por lo tanto, hemos de calcular el error estándar para luego compararlo con la diferencia observada en los grupos en estudio.

$$[P_1 - P_2] = [0,080 - 0,026] = 0,054$$

$$P = [P_1 - P_2] / 2 = [0,080 + 0,026] / 2 = 0,053$$

El error estándar se calcula de la siguiente forma:

$$\text{Error estándar (EE)} = \sqrt{p(1-p) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$$\text{Error estándar (EE)} = \sqrt{0,053(1-0,053) \left( \frac{1}{199} + \frac{1}{422} \right)} = 0,00035$$

$$\text{Error estándar (EE) multiplicado por } Z_{95\%} = 0,00035 * 1,96 = 0,00069$$

$$[P_1 - P_2]$$



$$Z_{95\%} \times \text{EE}$$

$$\text{EE} = \sqrt{p(1-p) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$$Z_{95\%} = 1,96$$

Si la diferencia de  $[P_1 - P_2]$  (0,054)  $>$  EE \*  $Z_{95\%}$  (0,00069) concluimos que existe una diferencia estadísticamente significativa entre los grupos en estudio ( $p < 0,05$ ); razón por la cual rechazamos  $H_0$ , por ende, aceptamos la  $H_a$ .

$$[P_1 - P_2] < Z_{95\%} \times \text{EE}$$

$$p > 0,1$$

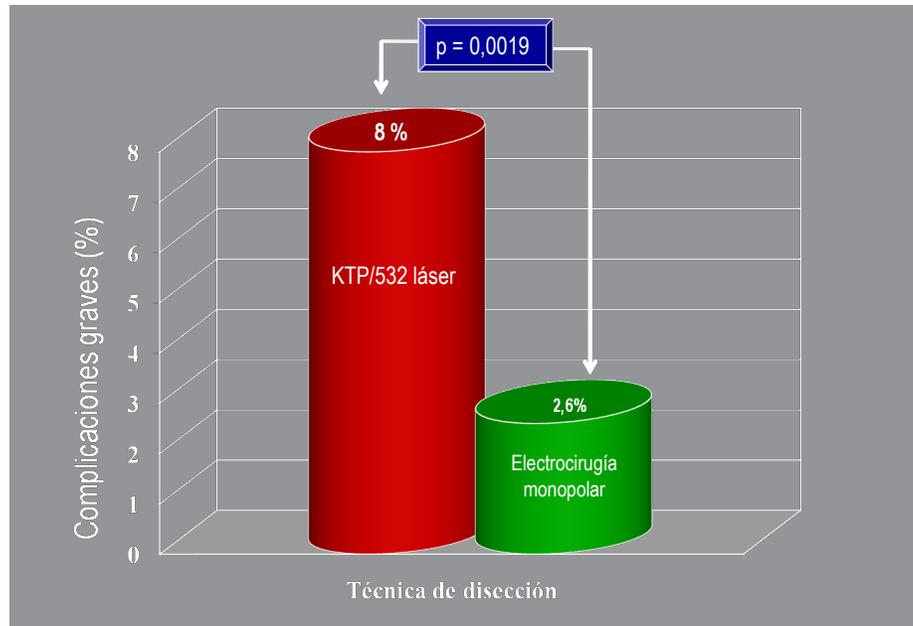
$H_0$  (hipótesis nula)

$$[P_1 - P_2] > Z_{95\%} \times \text{EE}$$

$$p < 0,05$$

$H_a$  (hipótesis alternativa)

Figura 1. Frecuencia de complicaciones graves en 641 pacientes colecistectomizados por vía laparoscópica. Distribución por técnicas de disección.



Test de  $\chi^2$  de Pearson

Lane GE, Lathrop JC. Comparison of results of KTP/532 laser versus monopolar electrocautery dissection in laparoscopic cholecystectomy. J Laparoendosc Surg 1993; 3: 209-214.

## Test de $\chi^2$ de Pearson

Resultados de un estudio imaginario que estudia la eficacia de dos tratamientos A y B en dos grupos de 30 pacientes		Variable resultado Porcentaje de respuestas a cada tratamiento		
Variable predictora Tipo de tratamiento	SI (+)	NO (-)	Total	
A	16 (8%) P <sub>1</sub>	183	199	
B	11 (2,6%) P <sub>2</sub>	411	422	

$$\chi^2_{\text{Pearson}} = 2,5 \rightarrow p = 0,12$$

No existe asociación

## Comparación de proporciones:

### Importancia del intervalo de confianza

Resultados de un estudio imaginario que estudia la eficacia de dos tratamientos A y B en dos grupos de 30 pacientes

$$\text{Diferencia } [P_1 - P_2] = 0,70 - 0,50 = 0,20$$

$$\text{IC}_{95\%} = 0,20 \pm 0,24$$

$$-4\% \text{ a } +44\%$$

A 50%

20%

B 70%

Realidad 15: ...de lo estadísticamente significativo a lo clínicamente relevante.	¿Es clínicamente relevante?	
¿Es estadísticamente significativo?	Sí (éxito)	No (fracaso)
Sí ( $p < 0,05$ )	Estadísticamente significativo	Estadísticamente significativo
	Clínicamente relevante	Clínicamente irrelevante ( $p < 0,05$ )
No ( $p \geq 0,1$ )	Estadísticamente NO significativo  Clínicamente relevante	Estadísticamente NO significativo  Clínicamente irrelevante

## Test de $\chi^2$ de Pearson

Resultados de un estudio imaginario que estudia la eficacia de dos tratamientos A y B en dos grupos de 800 pacientes	Variable resultado Porcentaje de respuestas a cada tratamiento		
Variable predictora Tipo de tratamiento	SI (+)	NO (-)	Total
A	480 (60%) P <sub>1</sub>	320	800
B	520 (65%) P <sub>2</sub>	280	800

$$\chi^2_{\text{Pearson}} = 4,2 \rightarrow p = 0,04$$

Existe asociación

## Comparación de proporciones:

### Importancia del intervalo de confianza

Resultados de un estudio imaginario que estudia la eficacia de dos tratamientos A y B en dos grupos de 800 pacientes

$$\text{Diferencia } [ P_1 - P_2 ] = 0,65 - 0,60 = 0,05$$

$$\text{IC}_{95\%} = 0,05 \pm 0,047$$

$$+ 0,3\% \text{ a } + 9,7\%$$

A 60%

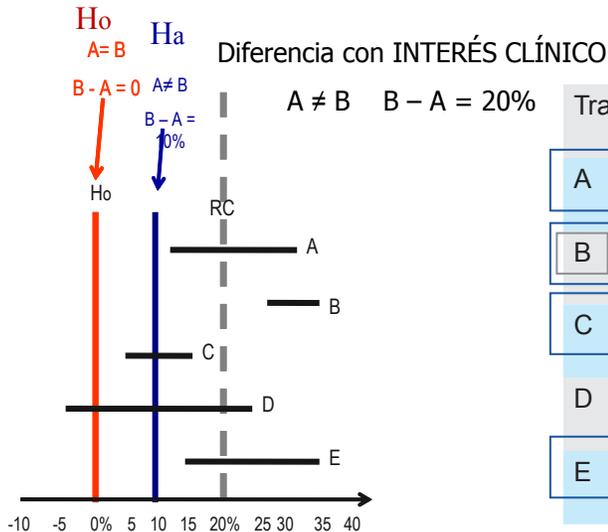
5%

B 65%

Realidad 15: ...de lo estadísticamente significativo a lo clínicamente relevante.	¿Es clínicamente relevante?	
¿Es estadísticamente significativo?	Sí (éxito)	No (fracaso)
Sí ( $p < 0,05$ )	Estadísticamente significativo Clínicamente relevante	Estadísticamente significativo Clínicamente irrelevante ( $p < 0,05$ )
No ( $p \geq 0,1$ )	Estadísticamente NO significativo Clínicamente relevante	Estadísticamente NO significativo Clínicamente irrelevante

## Importancia del intervalo de confianza

¿Cuál de los siguientes tratamientos es estadísticamente significativo?



Tratamiento	Diferencia observada	Significación estadística	IC95%
A	20%	$p < 0,05$	+10 a +30%
B	30%	$p < 0,05$	+25 a +35%
C	10%	$p < 0,05$	+5 a +15%
D	10%	NS	-5 a +25%
E	25%	$p < 0,05$	+15 a +35%

Resultados hipotéticos de un estudio que compara un tratamiento activo frente a un placebo y utiliza como medidas del efecto la diferencia entre los porcentajes de curación observados. Se considera que la mínima diferencia con relevancia clínica es del 20%. RC: Diferencia con relevancia clínica.

## Importancia del intervalo de confianza

Tratamiento	Diferencia observada	Significación estadística	IC95%	Interpretación
A	20%	$p < 0,05$	+10 a +30%	No concluyente Potencialmente importante (observado = relevante)
B	30%	$p < 0,05$	+25 a +35%	Importante
C	10%	$p < 0,05$	+5 a +15%	No concluyente
D	10%	NS	-5 a +25%	No concluyente
E	25%	$p < 0,05$	+15 a +35%	No concluyente Potencialmente importante (observado > relevante)

Resultados hipotéticos de un estudio que compara un tratamiento activo frente a un placebo y utiliza como medidas del efecto la diferencia entre los porcentajes de curación observados. Se considera que la mínima diferencia con relevancia clínica es del 20%. RC: Diferencia con relevancia clínica.

## Pregunta de investigación

¿Qué relación existe entre el grupo cultural/religioso de la ciudad de Ceuta y la prevalencia de inmunización frente al virus de la Rubeola en la población gestante?

Variable **PREDICTORA**  
Cualitativa dicotómica

- Cristianas
- Musulmanas



Variable **RESULTADO**  
Cualitativa dicotómica

Virus de la Rubeola

¿Cuál es el test estadístico más adecuado?

Díaz Portillo J, Torres E, Vadillo Andrade J, Laroca C.«PREVALENCIA DE ANTICUERPOS FRENTE AL VIRUS DE LA RUBÉOLA EN UNA POBLACIÓN GESTANTE DE CEUTA». Revista de Diagnóstico Biológico 1995, 44: 114-119.

## Test de $\chi^2$ de Pearson

Rubeola	Variable predictora Religión	
Variable resultado Inmunización	Cristiana	Musulmana
(+)	279 (87,2%)	372 (86,1%)
(-)	41	60

$$\chi^2_{\text{Pearson}} = 0,18; p=0,668$$

No existe asociación

## Resumen del procesamiento informático de los datos

Cálculos del programa estadístico tabla 2 x 2	Valor	Grados de libertad	Contraste bilateral	Contraste unilateral
$\chi^2$ Pearson	0,183	1	0,669	
Corrección de Yates	0,102	1	0,749	
Razón de verosimilitud	0,184		0,668	
Test exacto de Fisher		1	0,746	0,376
Nº Casos válidos	752			

$H_0$  (hipótesis nula) = La inmunización frente al virus de la rubeola NO está asociada al grupo cultural de la gestante.

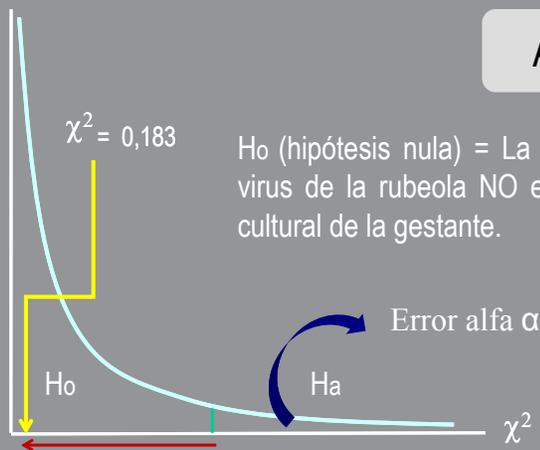
$H_a$  (hipótesis alternativa) = La inmunización frente al virus de la rubeola está asociada al grupo cultural de la gestante.



## Test de $\chi^2$ de Pearson (1 grado de libertad)

$$\chi^2_{\text{Pearson calculado}} = 0,183 < \chi^2_{\text{Pearson tabulado}} = 3,84$$

AZAR



$H_0$  (hipótesis nula) = La inmunización frente al virus de la rubeola NO está asociada al grupo cultural de la gestante.

AZAR

$$\chi^2_{\text{Pearson tabulado}} = 3,84$$

## Pregunta de investigación

¿Qué relación existe entre el hábito tabáquico y el cáncer de mama?

Variable PREDICTORA  
Cualitativa dicotómica

Hábito tabáquico



Variable RESULTADO  
Cualitativa dicotómica

Cáncer de mama

¿Cuál es la prueba estadística más adecuada?

Tabla de contingencia cáncer de mama versus tabaquismo	Variable RESULTADO Cáncer de mama		
	Si	No	Total
Variable PREDICTORA			
Fumadora	15	20	35
No fumadora	10	40	50
Total	25	60	85

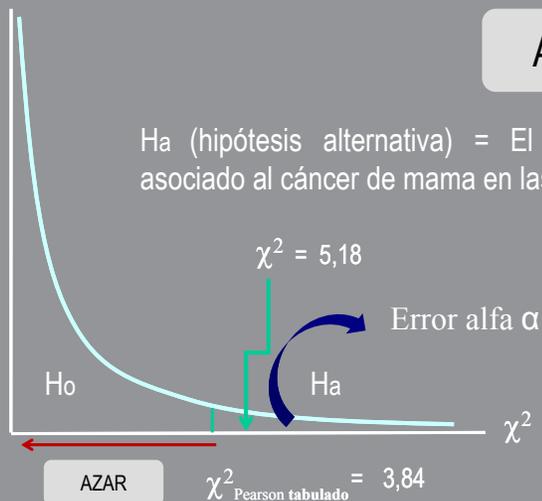
Cálculos del programa estadístico tabla 2 x 2	Valor	Grados de libertad	Contraste bilateral	Contraste unilateral
$\chi^2$ Pearson	5,18	1	0,023	
Corrección de Yates	4,14	1	0,042	
Razón de verosimilitud	5,14		0,023	
Test exacto de Fisher		1	0,030	0,021
Nº Casos válidos	85			

## Test de $\chi^2$ de Pearson (1 grado de libertad)

$$\chi^2_{\text{Pearson calculado}} = 5,18 > \chi^2_{\text{Pearson tabulado}} = 3,84$$

AZAR

Ha (hipótesis alternativa) = El hábito tabáquico está asociado al cáncer de mama en las mujeres de mi estudio.



AZAR

$$\chi^2_{\text{Pearson tabulado}} = 3,84$$

## Pregunta de investigación

¿El hábito tabáquico de los alumnos del I curso de metodología de investigación depende del sexo?

Variable PREDICTORA  
Cualitativa dicotómica

Sexo de los alumnos



Variable RESULTADO  
Cualitativa dicotómica

Hábito tabáquico

¿Cuál es la prueba estadística más adecuada?

## Test de $\chi^2$ de Pearson

Alumnos del curso	Variable RESULTADO		
Variable PREDICTORA	Fumador	No fumador	Total
Masculino	5	28	33
Femenino	18	10	28
Total	23	38	61

$H_0$  (hipótesis nula) = El hábito tabáquico es independiente de sexo del alumno del I curso de Metodología de Investigación.

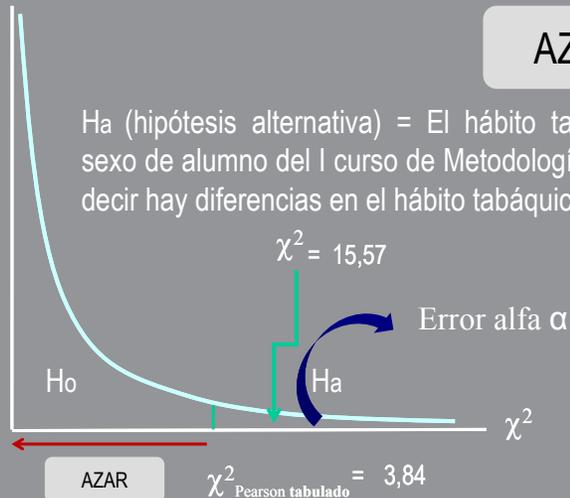
$H_a$  (hipótesis alternativa) = El hábito tabáquico depende del sexo del alumno del I curso de Metodología de Investigación.

## Test de $\chi^2$ de Pearson (1 grado de libertad)

$$\chi^2_{\text{Pearson calculado}} = 15,57 > \chi^2_{\text{Pearson tabulado}} = 3,84$$

AZAR

Ha (hipótesis alternativa) = El hábito tabáquico depende del sexo de alumno del I curso de Metodología de Investigación. Es decir hay diferencias en el hábito tabáquico en función del sexo.



## Test de $\chi^2$ de Pearson

### Resumen del procesamiento informático de los datos

Cálculos del programa estadístico	Valor	Grados de libertad	Contraste bilateral	Contraste unilateral
$\chi^2$ Pearson (*)	15,57	1	0,000	
Corrección de Yates	13,54	1	0,000	
Razón de verosimilitud	16,28		0,000	
Test exacto de Fisher		1	0,000	0,000
Nº Casos válidos	61			

(\*) 0 casillas (0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 10,56

$H_a$  (hipótesis alternativa) = El hábito tabáquico depende del sexo de alumno del I curso de Metodología de Investigación. Es decir, hay diferencias en el hábito tabáquico en función del sexo.

## Pregunta de investigación

¿El hábito tabáquico es homogéneo en los alumnos del II curso de metodología de investigación en función del sexo?

Variable PREDICTORA  
Cualitativa dicotómica

Sexo de los alumnos



Variable RESULTADO  
Cualitativa dicotómica

Hábito tabáquico

¿Cuál es la prueba estadística más adecuada?

Alumnos del curso	Variable RESULTADO		
	(+)	(-)	Total
Variable PREDICTORA			
Si	8	7	15
No	24	4	28
Total	32	11	43

$H_0$  (hipótesis nula) = El hábito tabáquico es homogéneo en hombres y mujeres del II curso de Metodología de Investigación.

$H_a$  (hipótesis alternativa) = El hábito tabáquico NO es homogéneo en hombres y mujeres del II curso de Metodología de Investigación.

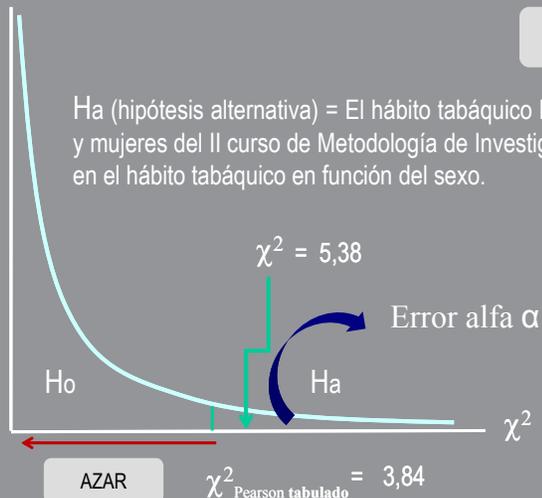
Test de  $\chi^2$  de Pearson

## Test de $\chi^2$ de Pearson (1 grado de libertad)

$$\chi^2_{\text{Pearson calculado}} = 5,38 > \chi^2_{\text{Pearson tabulado}} = 3,84$$

AZAR

$H_a$  (hipótesis alternativa) = El hábito tabáquico NO es homogéneo en hombres y mujeres del II curso de Metodología de Investigación. Es decir hay diferencias en el hábito tabáquico en función del sexo.



## Mito I: “La sonrisa de la esfinge”

¿Se cumplen todos los requisitos para el uso del test de  $\chi^2$  Pearson?

He observado que en uno de los casilleros el valor obtenido es inferior a 5 observaciones.

¿Cómo justificas el empleo del test de  $\chi^2$  Pearson?

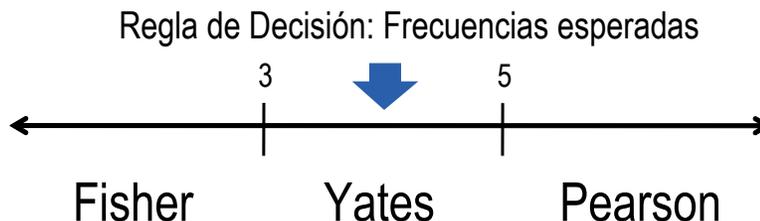
La corrección de Yates permite analizar si dos variables dicotómicas están asociadas cuando la muestra a estudiar es pequeña y no se cumplen las condiciones necesarias para que la aplicación del test  $\chi^2$  Pearson sea adecuada (uno de las casillas es inferior a 5). El test de  $\chi^2$  Pearson exige que todos los valores esperados de las celdas en una tabla de contingencia sean mayores de 5.

## Test de $\chi^2$ de Yates:

### Corrección de continuidad

En la frecuencia esperada en una de las cuatro celdas es menor de 5, por lo que NO resulta adecuado aplicar el test  $\chi^2$  Pearson, aunque sí el test de Yates. Las variables sexo y tabaquismo son independientes, porque la probabilidad asociada a los datos del estudio ( $p = 0,051$ ) es inferior a 0,05 (error alfa). Por tanto no podemos descartar la hipótesis nula ( $H_0$ ).

$H_0$  (hipótesis nula) = El hábito tabáquico es homogéneo en hombres y mujeres del II curso de Metodología de Investigación



## Resumen del procesamiento informático de los datos

Cálculos del programa estadístico	Valor	Grados de libertad	Contraste bilateral	Contraste unilateral
$\chi^2$ Pearson (*)	5,38	1	0,020	
Corrección de Yates	3,81	1	0,051	
Razón de verosimilitud	5,21		0,022	
Test exacto de Fisher		1	0,031	0,027
Nº Casos válidos	43			

(\*) 1 casillas (25%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 3,84

$H_0$  (hipótesis nula) = El hábito tabáquico NO depende del sexo de alumno del I curso de Metodología de Investigación. Es decir NO hay diferencias en el hábito tabáquico en función del sexo.

## Test exacto de Fisher

Variable PREDICTORA	Variable RESULTADO		Total
	(+)	(-)	
Si	a	b	a + b
No	c	d	c + d
Total	a + c	b + d	n

$$p = \frac{(a + b)! * (c + d)! * (a + c)! * (b + d)!}{n! * a! * b! * c! * d!}$$

## Pregunta de investigación

¿El hábito tabáquico es homogéneo en los alumnos del III curso de metodología de investigación en función del sexo?

Variable PREDICTORA  
Cualitativa dicotómica

Sexo de los alumnos



Variable RESULTADO  
Cualitativa dicotómica

Hábito tabáquico

¿Cuál es la prueba estadística más adecuada?

Alumnos del curso	Variable RESULTADO		
Variable PREDICTORA	Fumador	No fumador	Total
Masculino	2	5	7
Femenino	8	2	10
Total	10	7	17

$H_0$  (hipótesis nula) = El hábito tabáquico es homogéneo en hombres y mujeres del III curso de Metodología de Investigación.

$H_a$  (hipótesis alternativa) = El hábito tabáquico NO es homogéneo en hombres y mujeres del III curso de Metodología de Investigación.

Corrección de continuidad de Yates

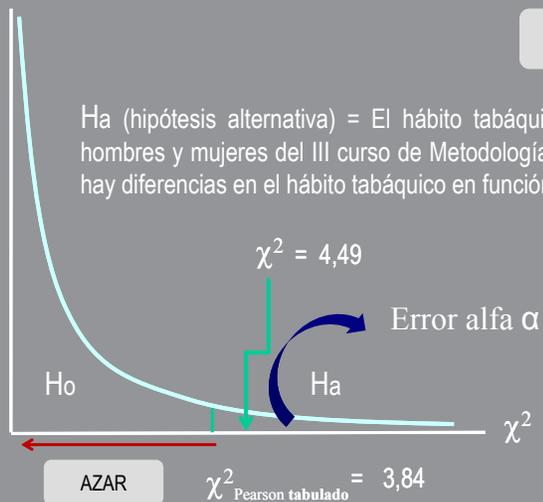
Test de  $\chi^2$  de Pearson

## Test de $\chi^2$ de Pearson (1 grado de libertad)

$$\chi^2_{\text{Pearson calculado}} = 4,49 > \chi^2_{\text{Pearson tabulado}} = 3,84$$

AZAR

Ha (hipótesis alternativa) = El hábito tabáquico NO es homogéneo en hombres y mujeres del III curso de Metodología de Investigación. Es decir hay diferencias en el hábito tabáquico en función del sexo.



## Mito I: “La sonrisa de la esfinge”

¿Se cumplen todos los requisitos para el uso del test de  $\chi^2$  Pearson?

He observado que en varios casilleros el valor es igual o inferior a 5 observaciones.

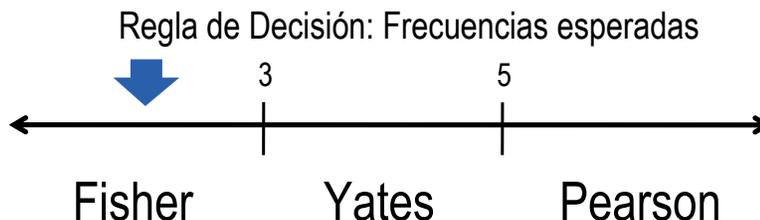
¿Cómo justificas el empleo del test de  $\chi^2$  Pearson?

El test exacto de Fisher permite analizar si dos variables dicotómicas están asociadas cuando la muestra a estudiar es demasiado pequeña y no se cumplen las condiciones necesarias para que la aplicación del test  $\chi^2$  Pearson sea adecuada. Estas condiciones exigen que los valores esperados de las celdas en una tabla de contingencia sean mayores de 5.

## Test exacto de Fisher

En esta tabla  $a=2$ ,  $b=5$ ,  $c=8$  y  $d=2$ . La frecuencia esperada en dos de las cuatro celdas es menor de 5, por lo que NO resulta adecuado aplicar el test  $\chi^2$  Pearson, ni la corrección de Yates, aunque sí el test exacto de Fisher. Si las variables sexo y tabaquismo fuesen independientes, la probabilidad asociada a los datos que han sido observados vendría dada por  $p = 0,058$  (bilateral).

$H_0$  (hipótesis nula) = El hábito tabáquico es homogéneo en hombres y mujeres del III curso de Metodología de Investigación.



Cálculos del programa estadístico	Valor	Grados de libertad	Contraste bilateral	Contraste unilateral
$\chi^2$ Pearson	4,49	1	0,034	
Corrección de Yates	2,62	1	0,105	
Razón de verosimilitud	4,65		0,031	
Test exacto de Fisher		1	0,058	0,052
Nº Casos válidos	17			

(\*) 3 casillas (75%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 2,88.

En ambos casos la  $H_0$  es la misma: la prevalencia de tabaquismo es igual en hombres y mujeres ( $H_0$ ). En el contraste unilateral la hipótesis alternativa ( $H_a$ ) dice de que la prevalencia es mayor en las mujeres que en los hombres ( $p=0,052$ ). Cuando el planteamiento se hace con una perspectiva bilateral, la hipótesis alternativa ( $H_a$ ) consiste en asumir que existen diferencias en la prevalencia de tabaquismo entre sexos, pero sin especificar de antemano en qué sentido se producen dichas diferencias, ( $p=0,058$ ), que tampoco es significativo. Las dos formas de cálculo nunca proporcionan los mismos resultados. El contraste unilateral (más exigente) siempre resultará en un valor de  $p$  menor al contraste bilateral. En ambos casos, y a la vista de los resultados del programa estadístico, NO existe evidencia estadística de asociación entre el sexo y el hábito tabáquico en los alumnos del curso de metodología de investigación de mi estudio.

## Pregunta de investigación

¿Qué utilidad clínica tiene la prueba diagnóstica X en el diagnóstico clínico de una determinada enfermedad E?

Objetivo: Evaluación de la prueba X en el diagnóstico de la enfermedad E.

Técnica analítica a valorar

PRUEBA X



Técnica de referencia  
(diagnóstico definitivo)

ENFERMEDAD E

¿Cuál es el test estadístico más acertado?

## Razón de Verosimilitud de $\chi^2$

- Si la proporción o frecuencia observada  $f_o$  es en realidad igual a la proporción esperada  $f_e$  según la hipótesis nula ( $H_o$ ), las 2 probabilidades calculadas serán iguales y su razón será igual a 1.
- Cuanto mayor sea la diferencia entre  $f_o$  y  $f_e$ , mayor será la razón entre ambas. La razón de estas 2 probabilidades o verosimilitudes puede utilizarse como estadístico para medir el grado de ajuste entre las frecuencias observadas y las esperadas.
- Esta prueba se calcula de la siguiente manera:

$$RV_{\chi^2} = 2 * \sum f_o * \ln \frac{f_e}{f_o}$$

## Razón de Verosimilitud de $\chi^2$

Estudio Casos - Controles	Diagnóstico definitivo		
	Enfermos	Sanos	Total
Prueba X			
Positiva (+)	9 (2,1%)	19 (4,4%)	28 (6,5%)
Negativa (-)	67 (15,7%)	335 (93,5%)	402 (93,5%)
Total	76 (17,7%)	354 (82,3)	430

$H_0$  (hipótesis nula) = Los resultados de la prueba diagnóstica X NO están relacionados con el diagnóstico definitivo.

$H_a$  (hipótesis alternativa) = Los resultados de la prueba diagnóstica X están relacionados con el diagnóstico definitivo.

## Razón de Verosimilitud de $\chi^2$

Cálculos del programa estadístico	Valor	Grados de libertad	Contraste bilateral	Contraste unilateral
$\chi^2$ Pearson	4,309	1	0,038	
Corrección de Yates	3,311	1	0,069	
Razón de verosimilitud	3,706		0,054	
Test exacto de Fisher		1	0,067	0,041
Nº Casos válidos	430			

A pesar de que el test de  $\chi^2$  de Pearson es significativo ( $p < 0,05$ ) no podemos rechazar la hipótesis nula ya que la razón de verosimilitud es mayor de 0,05 (diseño de casos y controles).

$H_0$  (hipótesis nula) = Los resultados de la prueba diagnóstica X NO están relacionados con el diagnóstico definitivo

Un equipo de traumatólogos desarrollaron una nueva técnica quirúrgica para el reemplazo de la cabeza del fémur que suponían superior a la técnica tradicional.

¿Existen diferencias significativa respecto al porcentaje de complicaciones postoperatorias inmediatas?

$H_0$  (hipótesis nula) = No hay diferencia entre ambas técnicas de quirúrgicas.

$H_a$  (hipótesis alternativa) = Sí existe diferencia entre ambas técnicas de quirúrgicas.

Técnica CLÁSICA



Técnica NUEVA

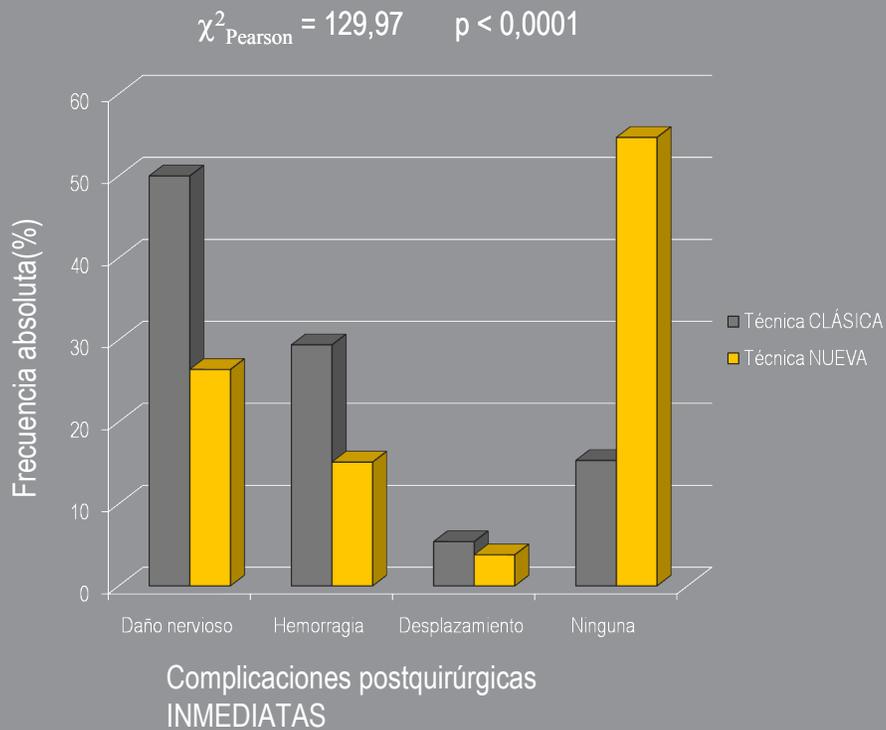
Variable resultado	Variable predictora	Tipo de cirugía
Complicaciones INMEDIATAS	Técnica CLÁSICA	Técnica NUEVA
Daño nervioso	485 (50%)	28 (26,4%)
Hemorragia	285 (29,4%)	16 (15,1%)
Desplazamiento	52 (5,4%)	4 (3,8%)
Ninguna	148 (15,3%)	58 (54,7%)
Total	970	106

¿Cuál es el test estadístico más acertado?

Test de  $\chi^2$  de Pearson

$$\chi^2_{\text{Pearson}} = 129,97 \rightarrow p < 0,0001$$

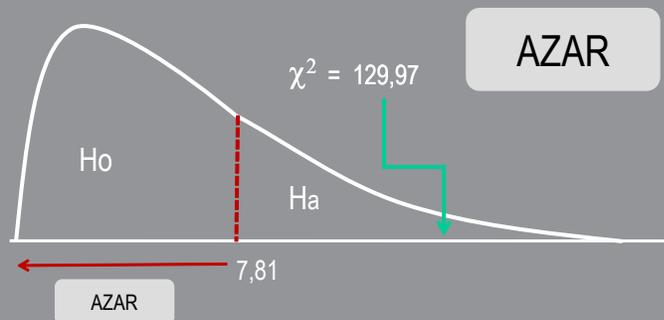
Existe asociación



## Test de $\chi^2$ de Pearson

$$\text{Grados de libertad} = [K - 1][F - 1] = (2-1)(4-1) = 3$$

$$\chi^2_{\text{Pearson calculado}} = 129,97 > \chi^2_{\text{Pearson tabulado}} = 7,81$$



$H_a$  (hipótesis alternativa) = La frecuencia de las complicaciones postquirúrgicas inmediatas en la nueva técnica NO eran iguales que en la técnica clásica.

## Pregunta de investigación

¿Qué relación existe entre el país de procedencia y la prevalencia de infección tuberculosa en el colectivo de inmigrantes del campamento de Calamocarro de Ceuta?

Variable **PREDICTORA**  
Cualitativa policotómica

Países africanos de origen



Variable **RESULTADO**  
Cualitativa dicotómica

Infección tuberculosa

¿Cuál es el test estadístico más adecuado?

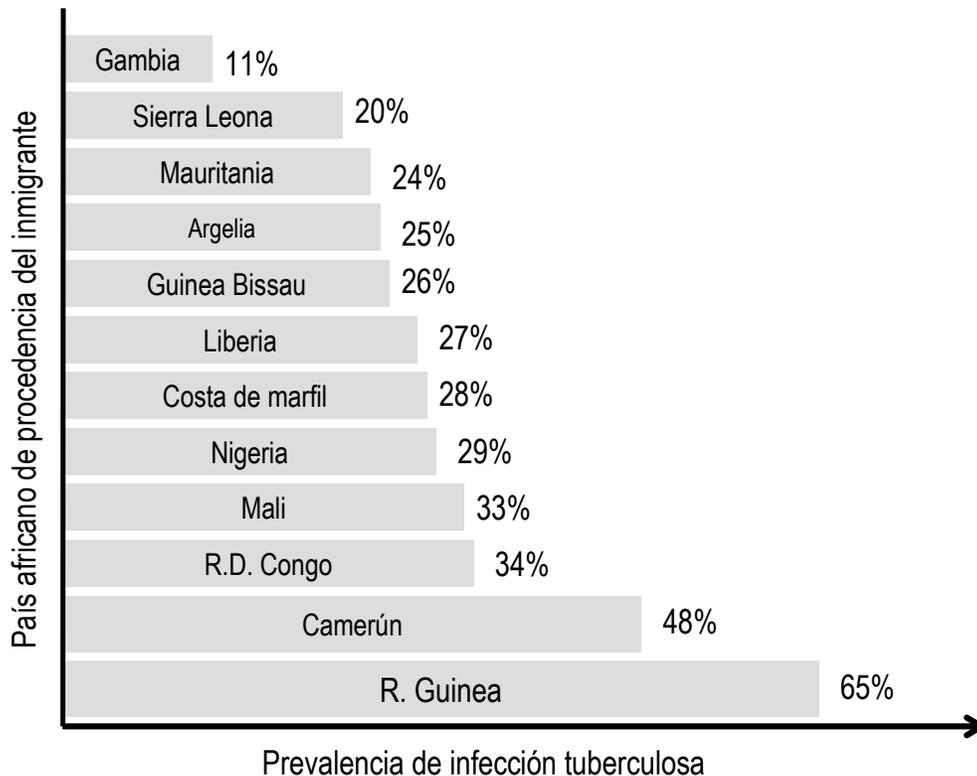
Fernández SanFrancisco T, Díaz Portillo J, Sánchez Romero JM et al. PREVALENCIA DE INFECCIÓN TUBERCULOSA EN LA POBLACIÓN DE INMIGRANTES EN CEUTA. Rev Esp Salud Pública 2001; 75: 501-558.

## Test de $\chi^2$ de Pearson

País de procedencia	N	Prevalencia
Argelia	133	25,6
Camerún	128	48,4
Costa de marfil	18	27,8
Gambia	107	11,2
Guinea Bissau	186	26,3
Liberia	105	27,4
Mali	291	32,7
Mauritania	55	24,1
Nigeria	597	29,1
R.D. Congo	146	34,0
R. Guinea	45	65,1
Sierra Leona	177	20,8

$$\chi^2_{\text{Pearson}} = 165; p < 0,00001$$

Existe asociación



Fernández SanFrancisco T, Díaz Portillo J, Sánchez Romero JM et al. PREVALENCIA DE INFECCIÓN TUBERCULOSA EN LA POBLACIÓN DE INMIGRANTES EN CEUTA. Rev Esp Salud Pública 2001; 75: 501-558.

## Pregunta de investigación

¿Existen diferencias en la prevalencia de infección tuberculosa en el colectivo de inmigrantes según la ciudad de destino (Ceuta, Barcelona, Toledo)?

Variable PREDICTORA  
Cualitativa policotómica

Destino  
- Ceuta  
- Barcelona  
- Toledo



Variable RESULTADO  
Cualitativa dicotómica

Infección tuberculosa

¿Cuál es el test estadístico más adecuado?

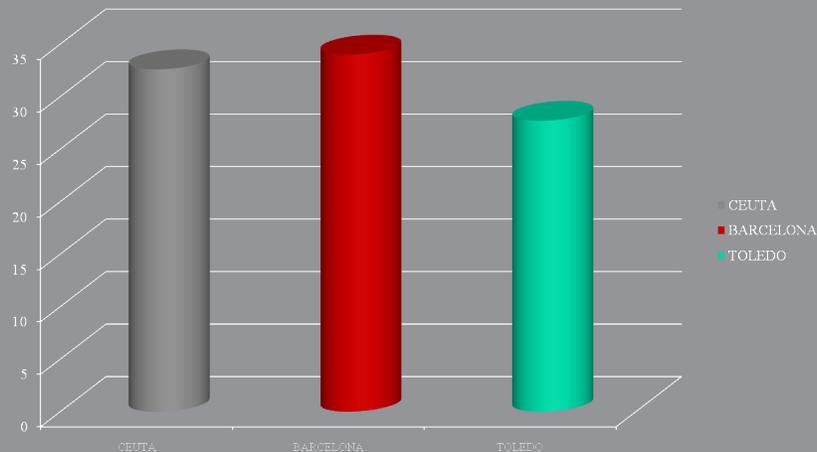
## Test de $\chi^2$ de Pearson

Variable PREDICTORA	Variable RESULTADO Prueba de Mantoux		Total
	(+)	(-)	
Ceuta	722 (32,6%)	1501	2223
Barcelona	506 (34%)	983	1489
Toledo	75 (27,7%)	269	344

$$\chi^2_{\text{Pearson}} = 19,29; p = 0,000065$$

Ha: Existen diferencias en la prevalencia de infección tuberculosa en el colectivo de inmigrantes según la ciudad de destino (Ceuta, Barcelona, Toledo).

## PREVALENCIA DE INFECCIÓN TUBERCULOSA EN INMIGRANTES



Fernández Sanfrancisco MT, Díaz Portillo J, Sánchez Romero JM, et al. Prevalencia de infección tuberculosa en la población inmigrante de Ceuta. *Rev Esp Salud Pública* 2001; 75: 551-558.

Duran E, Cabezos J, Ros M, et al. Tuberculosis en los inmigrantes recién llegados a Barcelona. *Med Clin (Barc)* 1996; 106: 525-528.

Alonso Moreno FJ, García Bajo MC, Lougedo Calderón MJ, et al. Prevalencia de infección tuberculosa en las personas inmigrantes del área sur de Toledo. *Rev Esp Salud Pública* 2004; 78: 593-600.

## Realidad 3

...de lo estadísticamente significativo a lo clínicamente relevante

Realidad 3: ...de lo estadísticamente significativo a lo clínicamente relevante.	¿Es clínicamente relevante?	
¿Es estadísticamente significativo?	Sí (éxito)	No (fracaso)
Sí ( $p < 0,05$ )	Clínicamente relevante Estadísticamente significativo	Clínicamente irrelevante ( $p < 0,05$ ) Estadísticamente significativo ¿Error Alfa?
No ( $p \geq 0,1$ )	Clínicamente relevante Estadísticamente NO significativo ¿Error Beta?	Clínicamente irrelevante Estadísticamente NO significativo

## Pregunta de investigación

¿Qué relación y qué grado de asociación existe entre el grupo cultural/religioso de la ciudad de Ceuta y la prevalencia de inmunización frente a *Toxoplasma goondi* en la población gestante?

Variable PREDICTORA  
Cualitativa policotómica

Grupo cultural:  
- Cristianas  
- Musulmanas



Variable RESULTADO  
Cualitativa dicotómica

Infección *Toxoplasma goondi*

¿Cuál es el test estadístico más adecuado?

Díaz Portillo J, Torres E, Vadillo Andrade J. SEROEPIDEMIOLOGÍA DE LA TOXOPLASMOSIS EN UNA POBLACIÓN GESTANTE DE CEUTA. Revista de Diagnóstico Biológico 1998; 47: 106-116.

## Test de $\chi^2$ de Pearson

Variable resultado Inmunización	Variable predictorora Religión	
	Cristiana	Musulmana
(+)	112	220
(-)	210	148

$$\chi^2_{\text{Pearson}} = 43,99; p < 0,0000001$$

Existe asociación

No cuantifica el grado de intensidad de la asociación.

## Test de McNemar para datos apareados

(Prueba de la significación de los cambios)

McNemar	DESPUÉS	
ANTES	ÉXITO (+)	FRACASO (-)
ÉXITO (+)	A	B
FRACASO (-)	C	D

$$\chi^2_{\text{McMemar}} = \frac{[(B - C) - 1]^2}{B + C}$$

## ¿Existen evidencias de un mejor efecto analgésico en los dos fármacos protocolizados para el cólico nefrítico?

### - Diseño APAREADO

- Un solo grupo de pacientes ( $n=20$ ) con cólico nefrítico recidivante (dos episodios de cólico nefrítico).
- Dos mediciones repetidas (antes y después).
- Gran potencia estadística (mínima variabilidad aleatoria).

- Hipótesis nula  $H_0$ : la eficacia de ambos analgésicos es la misma Metamizol = Ketorolaco

Test de McNemar

N = 20 pacientes

KETOROLACO



METAMIZOL

## Test de McNemar para datos apareados

McNemar	METAMIZOL	
KETOROLACO	ÉXITO = 1	FRACASO = 0
ÉXITO = 1	A	B
FRACASO = 0	C	D

$$\chi^2_{\text{McMemar}} = \frac{[(B - C) - 1]^2}{B + C}$$

$$Z_{\text{McNemar}} = \frac{(B - C) - 1}{\sqrt{B + C}}$$

## Test de McNemar para datos apareados

McNemar	METAMIZOL	
KETOROLACO	ÉXITO = 1	FRACASO = 0
ÉXITO = 1	8	9
FRACASO = 0	2	1

$$\chi^2_{\text{McMemar}} = \frac{[(9 - 2) - 1]^2}{9 + 2} = 3,27 \quad p > 0,05$$

## Procesamiento informático de los datos

Test de McNemar para datos apareados		
	Valor	Significación bilateral (*)
Prueba de McNemar	3,27	0,065
N casos válidos	20	

(\*) Utilizada la distribución binomial

Hipótesis nula  $H_0$ : La eficacia de ambos analgésicos es la misma, es decir  
Metamizol = Ketorolaco

¿Es posible un error  $\beta$ ?

¿Qué ocurre si aumentamos el tamaño muestral?

N = 30 pacientes

KETOROLACO



METAMIZOL

## Test de McNemar para datos apareados

McNemar	METAMIZOL	
KETOROLACO	ÉXITO = 1	FRACASO = 0
ÉXITO = 1	11	14
FRACASO = 0	3	2

$$\chi^2_{\text{McMemar}} = \frac{[(14 - 3) - 1]^2}{14 + 3} = 5,88$$
$$p = 0,013$$

## Procesamiento informático de los datos

Test de McNemar para datos apareados		
	Valor	Significación bilateral (*)
Prueba de McNemar	5,88	0,013
N casos válidos	30	

(\*) Utilizada la distribución binomial

Hipótesis alternativa  $H_a$ : La eficacia de ambos analgésicos es NO la misma, es decir Metamizol  $\neq$  Ketorolaco

## Pregunta de investigación

¿Existen evidencias de una mejoría en los conocimientos teóricos de Metodología de investigación en los alumnos después del curso?

Examen de 14 preguntas a los 8 alumnos antes y después de realizar el curso

Aprobado: [ > 7 respuestas correctas de 14 ] = 1

Suspense: [ < 8 respuestas correctas de 14 ] = 0

	DESPUÉS del curso	
ANTES del curso	APROBADO (> 7)	SUSPENSO (< 8)
APROBADO (> 7)	A	B
SUSPENSO (< 8)	C	D

## Test de McNemar para datos apareados

(Prueba de la significación de los cambios)

McNemar	DESPUÉS del curso	
ANTES del curso	APROBADO (< 7)	SUSPENSO (< 8)
APROBADO (< 7)	A	B
SUSPENSO (< 8)	C	D

$$\chi^2_{\text{McMemar}} = \frac{[(B - C) - 1]^2}{B + C}$$

$$Z_{\text{McNemar}} = \frac{(B - C) - 1}{\sqrt{B + C}}$$

¿Existen evidencias de una mejoría en los conocimientos teóricos de metodología de investigación en los alumnos después del curso?

Nº del Alumno	Evaluación ANTES del curso	Evaluación DESPUÉS del curso
1º	APROBADO (10)	APROBADO (14)
2º	SUSPENSO (7)	APROBADO (14)
3º	SUSPENSO (7)	APROBADO (14)
4º	SUSPENSO (5)	APROBADO (14)
5º	SUSPENSO (4)	APROBADO (13)
6º	SUSPENSO (3)	APROBADO (13)
7º	SUSPENSO (2)	APROBADO (13)
8º	SUSPENSO (2)	APROBADO (11)

¿Cuál es el test estadístico más acertado?

## Test de McNemar para datos apareados

(Prueba de la significación de los cambios)

McNemar	DESPUÉS del curso	
	APROBADO (< 7)	SUSPENSO (< 8)
ANTES del curso		
APROBADO (< 7)	1	0
SUSPENSO (< 8)	7	0

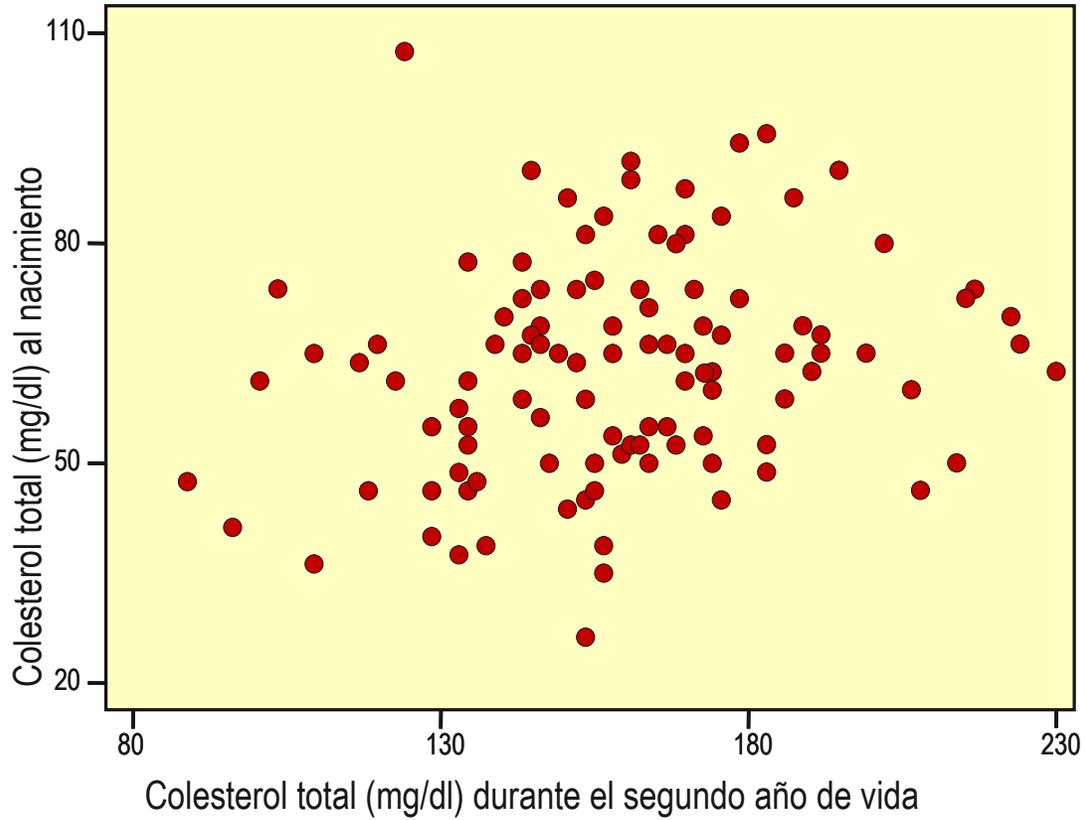
$$\chi^2_{\text{McMemar}} = \frac{[(0 - 7) - 1]^2}{0 + 7} = 9,14$$

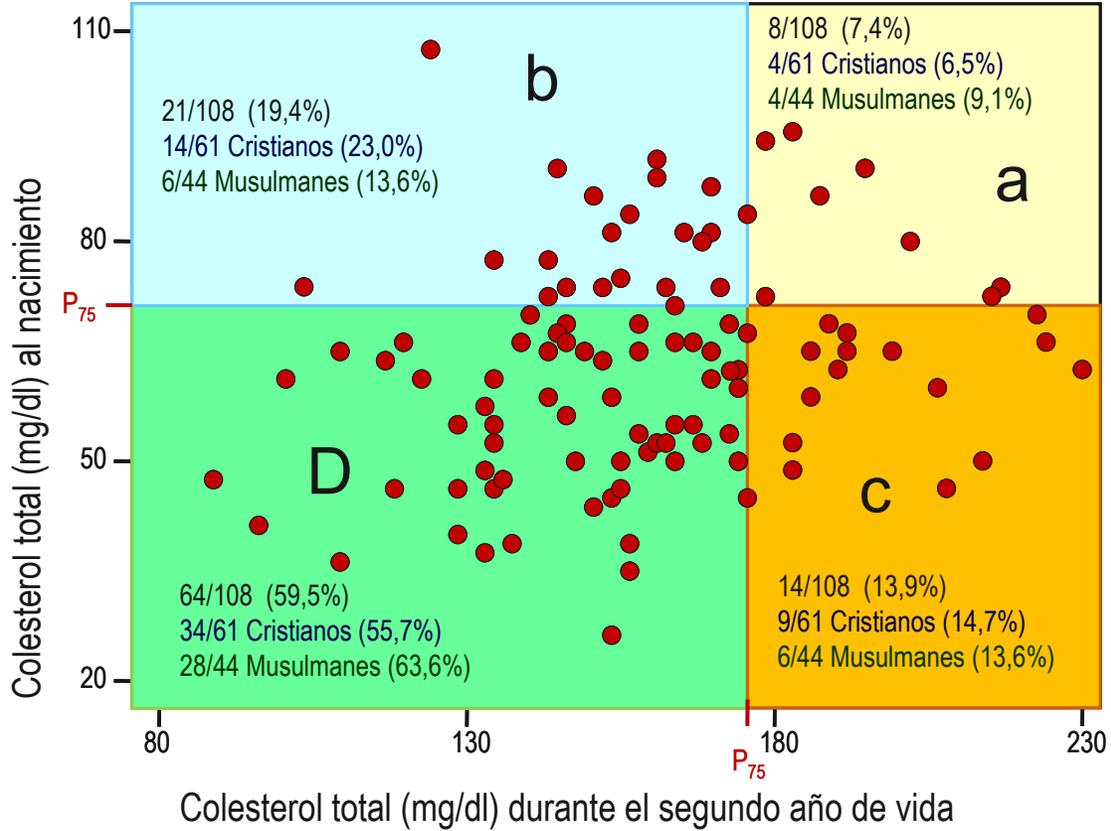
$$Z_{\text{McNemar}} = \frac{(B - C) - 1}{\sqrt{B + C}}$$

Z	Valor	$\alpha$
$Z_{95\%}$	1,96	0,05
$Z_{98\%}$	2,32	0,02
$Z_{99\%}$	2,56	0,01
$Z_{99,5\%}$	3,29	0,005

$$Z_{\text{McNemar}} = \frac{(0 - 7) - 1}{\sqrt{0 + 7}} = 3,03 > Z_{99\%} = 2,56$$

Conclusión: Como el estadístico Z calculado 3,03 es superior a Z tabulado para un error alfa de 0,01 (2,56), podemos decir que las dos variables evaluadas NO son independientes, sino que están asociadas ( $p < 0,01$ ), es decir, la acción formadora mejora los conocimientos de los alumnos.





## Test de McNemar para datos apareados

(Prueba de la significación de los cambios)

McNemar	SEGUNDO AÑO DE VIDA	
NEONATOS	CT > P <sub>75</sub>	CT ≤ P <sub>75</sub>
CT > P <sub>75</sub>	A	B
CT ≤ P <sub>75</sub>	C	D

$$\chi^2_{\text{McMemar}} = \frac{[(B - C) - 1]^2}{B + C}$$

$$Z_{\text{McNemar}} = \frac{(B - C) - 1}{\sqrt{B + C}}$$

## Test de McNemar para datos apareados

(Prueba de la significación de los cambios)

McNemar	SEGUNDO AÑO DE VIDA	
NEONATOS	CT > P <sub>75</sub>	CT ≤ P <sub>75</sub>
CT > P <sub>75</sub>	8	21
CT ≤ P <sub>75</sub>	14	64

Test de McNemar para datos apareados			
	Valor	Significación bilateral	Significación unilateral
Prueba de McNemar	1,03	0,311	0,155
N casos válidos	108		

## Test de McNemar para datos apareados

(Prueba de la significación de los cambios)

McNemar	SEGUNDO AÑO DE VIDA	
	CT > P <sub>75</sub>	CT ≤ P <sub>75</sub>
NEONATOS		
CT > P <sub>75</sub>	8	21
CT ≤ P <sub>75</sub>	14	64

$$\chi^2_{\text{McMemar}} = \frac{[(21 - 14) - 1]^2}{21 + 14} = 1,03$$

Jacobo Díaz Portillo. "Influencia del grupo étnico en el patrón lipoproteico de una determinada población pediátrica de Ceuta", Tesis doctoral. Facultad de Medicina de la Universidad de Cádiz. 1997

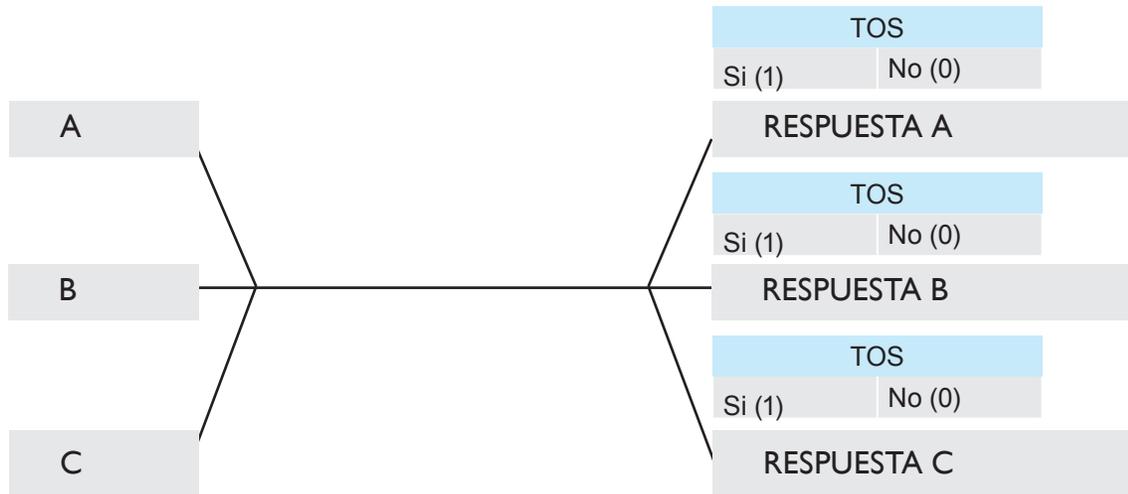
$$Z_{\text{McNemar}} = \frac{(B - C) - 1}{\sqrt{B + C}}$$

$$Z_{\text{McNemar}} = \frac{(21 - 14) - 1}{\sqrt{21 + 14}} = 1,014$$

Z	Valor	$\alpha$
$Z_{95\%}$	1,96	0,05
$Z_{98\%}$	2,32	0,02
$Z_{99\%}$	2,56	0,01
$Z_{99,5\%}$	3,29	0,005

Conclusión: Como el estadístico Z calculado 1,014 es INFERIOR a Z tabulado para un error alfa de 0,05 (1,96), podemos decir que las dos variables evaluadas son independientes, es decir, NO están asociadas, y por tanto no hay cambios importantes en la distribución del colesterol total al pasar del nacimiento al segundo año de vida. Este comportamiento también se repite en los dos grupos culturales (cristianos y musulmanes).

¿Existen evidencias de un mejor efecto antitusígeno en los 3 fármacos ensayados?



¿Cuál es el test estadístico más adecuado?

## Q de Cochran

### Para tres o más muestras dependientes

- La prueba Q de Cochran es una extensión de la prueba de McNemar, que se utiliza en los modelos experimentales con 3 o más muestras dependientes o relacionadas entre sí, es decir, esta población sirve como su propio control, en el que existe un período previo y otro ulterior; además, el tipo de escala debe ser nominal.
- El valor calculado en la prueba Q de Cochran se distribuye igual que la ji cuadrada, por lo cual el símbolo utilizado será  $X^2_Q$ .

$$X^2_Q = \frac{(K - 1) [K \sum Gn^2 - (\sum Gn)^2]}{K \sum Lc - \sum Lc^2}$$

Donde:

$X^2_Q$  = estadístico ji cuadrada de la prueba Q de Cochran.

K = número de tratamientos.

Gn = número total de respuestas de cambio de cada tratamiento o columna.

Lc = número total de respuestas de cambio por individuo de la muestra o hileras.

## 25 pacientes reciben 3 fármacos para mejorar su tos

Paciente	Fármaco A	Fármaco B	Fármaco C
1	1	1	1
2	1	1	0
3	0	1	0
4	1	1	1
5	0	0	1
6	1	1	0
7	0	1	0
8	1	1	0
9	0	0	0
10	0	1	1
11	1	1	0
12	1	1	1
13	1	0	0
14	1	1	1

Paciente	Fármaco A	Fármaco B	Fármaco C
15	0	1	1
16	1	0	0
17	0	0	1
18	1	1	0
19	0	0	0
20	1	0	1
21	1	1	1
22	0	1	0
23	1	1	0
24	1	0	1
25	0	1	1

0 = Mejora la tos

1 = NO mejora la tos

	N	Media	DE	Mínimo	Máximo
Fármaco A	25	0,60	0,500	0	1
Fármaco B	25	0,65	0,478	0	1
Fármaco C	25	0,48	0,510	0	1

	Valores	
	0	1
Fármaco A	10	15
Fármaco B	8	17
Fármaco C	13	12

Test estadísticos	
N	250
Q de Cochran	2,111
gl	2
Significación	0,348

$$\text{Grados de libertad} = [K - 1] [F - 1]$$

Hipótesis nula  $H_0$ : No hay diferencia entre los 3 fármacos para evitar la tos. 

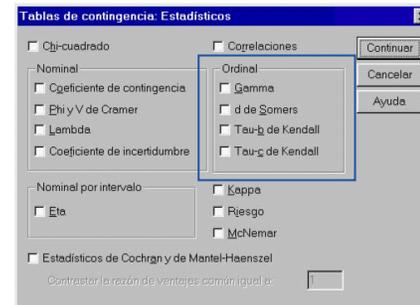
Hipótesis alternativa  $H_a$ : Hay diferencia entre los 3 fármacos para evitar la tos.

Conclusiones: No hay diferencias estadísticamente significativas entre los 3 fármacos para evitar la tos en los pacientes.

## Datos categóricos

### Cualitativa versus cualitativa ORDINALES

- Test de  $\chi^2$  de tendencia lineal
- Tau-B de Kendall o Stuart
- Tau-C de Kendall
- Gamma Y de Goodman y Kruskal
- D de Somers
- Correlación de Spearman



Estos estadísticos NO se pueden utilizar cuando alguna de las variables sea NOMINAL

## Capítulo 6

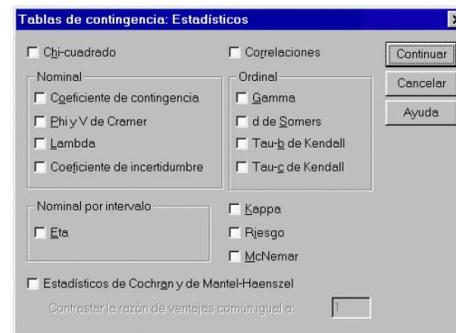
Publicación Bioestadística: Análisis bivariante.  
Test de contraste de hipótesis.  
Cualitativa versus cualitativa.  
Medidas de asociación.

Dr. Jacobo Díaz Portillo  
Coordinador de Formación Continuada  
Hospital Universitario de Ceuta

## Datos categóricos:

### Cualitativa versus cualitativa. Nominales

- Test de  $\chi^2$  de Pearson
  - Relación entre dos variables cualitativas nominales o policotómicas
  - Comparación de proporciones
- Corrección de Yates
- Test exacto de Fisher
  - Valores esperados menores de 5
- Razón de verosimilitud
  - Evaluación de pruebas diagnósticas
- Test de  $\chi^2$  de McNemar
  - Datos apareados o emparejados
- Coeficiente PHI ( $\Phi$ )
- V de Cramer
- Coeficiente de contingencia
- Coeficiente de incertidumbre
- Tau de Goodman y Kruskal
- Lambda de Goodman y Kruskal (coeficiente de predictibilidad de Guttman)

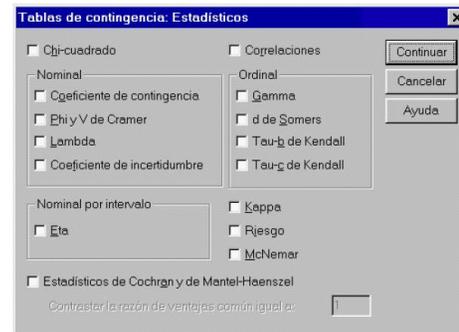


## Datos categóricos:

### Cualitativa versus cualitativa. Ordinales

- Test de  $\chi^2$  de tendencia lineal
- Tau-B de Kendall o Stuart
- Tau-C de Kendall
- Coeficiente Gamma Y de Goodman y Kruskal
- D de Somers
- Correlación de Spearman

Estos estadísticos NO se pueden utilizar cuando alguna de las variables sea NOMINAL

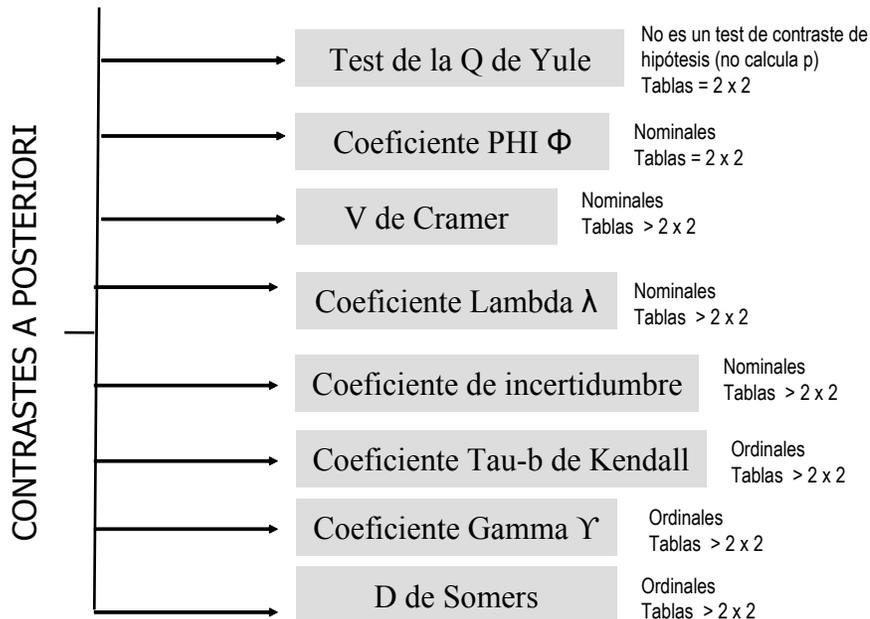


## Medidas de asociación

Son indicadores estadísticos que miden la FUERZA o INTENSIDAD con la que una determinada enfermedad o evento de salud (VARIABLE RESULTADO) está asociada o relacionada con un determinado factor (que se presume como su causa, VARIABLE PREDICTORA).

Estadístico	Transversales	Cohortes	Casos-contróles
OR (Odds ratio)	Limitaciones	NO	SI
RR (Riesgo Relativo)	NO	SI	NO
Q de Yule	SI	SI	SI
Coefficiente PHI $\Phi$	SI	SI	SI
V de Cramer	SI	SI	SI
Coefficiente de contingencia	SI	SI	SI
Lambda $\lambda$	SI	SI	SI
Tau de Goodman y Kruskal	SI	SI	SI
Coefficiente de incertidumbre	SI	SI	SI
Gamma $\delta$	SI	SI	SI
D de Somers	SI	SI	SI
Tau-b de Kendall	SI	SI	SI

## Test de $\chi^2$ de independencia de Pearson



## Test de la Q de Yule

Variable predictorora Religión		
Variable resultado	Cristiana	Musulmana
Inmunización		
(+)	A	B
(-)	C	D

$$Q = \frac{[(A * D) - (B * C)]}{[(A * D) + (B * C)]}$$

# Test de la Q de Yule

## Coeficiente Gamma

### - Interpretación:

- $Q = 0$      $\longrightarrow$     independencia
- $Q > 0$      $\longrightarrow$     asociación positiva
- $Q < 0$      $\longrightarrow$     asociación negativa

### - Intervalo de Q: (-1, +1)

### - Información:

- Medida de la intensidad de la asociación
- Sentido de la misma

## Características de la Q de Yule

- Medida de asociación.
  - Valores mayores de 0 (asociación positiva).
  - Valores menores de 0 (asociación negativa).
  - Valor = 0 o próximos a 0 (es probable que no exista asociación).
- Válida para cualquier tipo de Estudios (diseño).
- Alcanza valores extremos (+1, -1) bajo condiciones de asociación perfectas.
- Es independiente de los cambios de ESCALA de filas y columnas.
- Si se intercambian entre si las filas o las columnas se mantiene la magnitud de la INTENSIDAD de la asociación, sólo cambia la dirección de la misma, es decir el signo de la Q.
- Es más valida con un test previo de contraste de hipótesis que indique relación entre ambas variables.

## Test de la Q de Yule

### Coeficiente Gamma

Rubeola Variable resultado Inmunización	Religión	
	Cristiana	Musulmana
(+)	279 (87,2%)	372 (86,1%)
(-)	41	60

$$Q = \frac{[(279 * 60) - (372 * 41)]}{[(279 * 60) + (372 * 41)]} = 0,047$$

Gamma (Q de Yule)	EEk	T <sub>aprox</sub>	p
0,047	0,113	0,430	0,667

## Mito I: “La sonrisa de la esfinge”

La Q de Yule es un test que cuantifica la magnitud de la intensidad de la asociación entre dos variables cualitativas, pero NO es un test de contraste de hipótesis, y por tanto, que para usarlo se debe cumplir previamente la asociación estadística entre ambas variables. ¿Se cumple?

He observado en tus resultados anteriores que el test de contraste de hipótesis de  $\chi^2$  Pearson tiene un valor de tan sólo 0,18, con  $p = 0,67$ . ¿Cómo justificas el empleo de la Q de Yule si NO existe asociación estadística entre ambas variables?

## Test de la Q de Yule

Toxoplasma goondi	Religión	
Variable resultado	Cristiana	Musulmana
Inmunización		
(+)	112 (34,8%)	220 (59,8%)
(-)	210	148

$$Q = \frac{[(112 * 148) - (220 * 210)]}{[(112 * 148) + (220 * 210)]} = 0,47$$

Gamma (Q de Yule)	EEk	T <sub>aprox</sub>	p
- 0,47	0,041	- 6,780	0,000

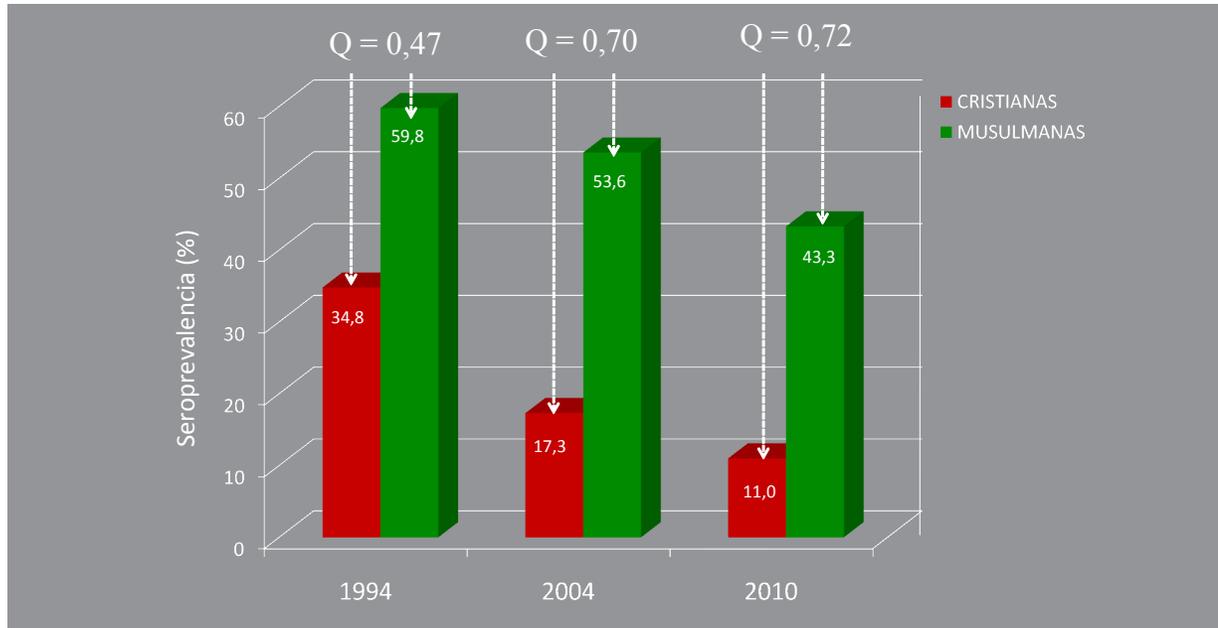


Figura .- Evolución de la seroprevalencia de la toxoplasmosis en gestante de Ceuta (periodos 1994, 2004, 2010) (Q= Q de Yule; gamma;  $p < 0,001$ )

## Ejercicio: ¿Qué grado de asociación existe entre el infarto de miocardio y el sexo del paciente?

Q de Yule	Acontecimiento de infarto	
Sexo	(+)	(-)
Hombre	32	10
Mujer	11	27

$$Q = \frac{[(32 * 27) - (11 * 10)]}{[(32 * 27) + (11 * 10)]} = 0,77$$

Gamma (Q de Yule)	EEk	T <sub>aprox</sub>	p
0,774	0,102	4,782	0,000

## Ejercicio: ¿Qué grado de asociación existe entre el infarto de miocardio y el sexo del paciente?

Q de Yule

El valor de 0,77 sugiere un grado de asociación entre los factores sexo y acontecimiento del infarto bastante elevado, ya que 0,77 está próximo a 1, y además indica que dicha asociación es positiva, es decir los varones desarrollan con más frecuencia el infarto que las mujeres.

## Pregunta de investigación

¿En qué grupo cultural/religioso es mayor la prevalencia de inmunización frente a *Toxoplasma goondi* en la población gestante de Ceuta?  
¿Musulmanas o cristianas?

Variable **PREDICTORA**  
Cualitativa policotómica

Grupo cultural:  
- Cristianas  
- Musulmanas



Variable **RESULTADO**  
Cualitativa dicotómica

Infección *Toxoplasma goondi*

¿Cuál es el test estadístico más adecuado?

Díaz Portillo J, Torres E, Vadillo Andrade J. SEROEPIDEMIOLOGÍA DE LA TOXOPLASMOSIS EN UNA POBLACIÓN GESTANTE DE CEUTA. Revista de Diagnóstico Biológico 1998; 47: 106-116.

Variable resultado	Variable predictorora		
	Positiva (+)	Negativa (-)	Total
Positiva (+)	a	b	a + b
Negativa (-)	c	d	c + d
Total	a + c	b + d	n

$$\Phi = \sqrt{\frac{\chi^2}{n}} \quad \text{Coeficiente PHI } \Phi$$

$$\Phi = \frac{(a * d) - (b * c)}{\sqrt{(a + c) (b + d) (a + b) (c + d)}}$$

## Coeficiente PHI $\Phi$

Coeficiente PHI ( $\Phi$ )	Variable predictorora		
	Variable resultado	Positiva (+)	Negativa (-)
Positiva (+)	112 (34,8%)	220 (59,8%)	332
Negativa (-)	210	148	358
Total	322	368	690

$$\Phi = \sqrt{\frac{\chi^2}{n}} \quad \Phi = \sqrt{\frac{43,99}{690}} = 0,25$$

Coeficiente  $\Phi$  es un test de contraste de hipótesis  $p < 0,01$

## Coeficiente PHI $\Phi$

Coeficiente PHI ( $\Phi$ )	Variable predictorora		
	Positiva (+)	Negativa (-)	Total
Positiva (+)	112 (34,8%)	220 (59,8%)	332
Negativa (-)	210	148	358
Total	322	368	690

$$\Phi = \frac{(112 * 148) - (220 * 210)}{\sqrt{(112 + 210) (220 + 148) (112 + 220) (210 + 148)}} = -0,25$$

Ha: La inmunización a Toxoplasma se encuentra asociada a la cultura musulmana.

## Coeficiente PHI $\Phi$

- Condiciones: Tablas tetracóricas (2 x 2).
- Interpretación:
  - $\Phi = 0$   $\longrightarrow$  independencia.
  - $\Phi > 0$   $\longrightarrow$  asociación positiva.
  - $\Phi < 0$   $\longrightarrow$  asociación negativa.
- Intervalo de  $\Phi$  : entre -1 y 1.
- Información: Es un test de contraste de hipótesis que ofrece una medida de la intensidad de la asociación. Cuando se acercan a cero, indican independencia o asociación muy débil entre las variables. Cuando se acercan a uno, indican una relación más fuerte o intensa entre las variables asociadas.

## Pregunta de investigación

¿El resultado del tratamiento depende del fármaco empleado?  
¿La condición de asintomático se encuentra asociado al uso de Mosaprida?

Variable PREDICTORA  
Cualitativa policotómica

- Metoclopramida (n = 99)
- Cisaprida (n = 102)
- Mosaprida (n = 99)



Variable RESULTADO  
Cualitativa dicotómica

- Asintomático
- Síntomas
- Mejoría clínica

¿Cuál es el test estadístico más adecuado?

## Nominal x Nominales (Tablas > 2 x 2)

V de Cramer

Coefficiente de contingencia

Lambda  $\lambda$

Tau de Goodman y Kruskal

Coefficiente de incertidumbre

## ¿El resultado del tratamiento depende del fármaco empleado?

Variable PREDICTORA	Variable RESULTADO Resultado del tratamiento a la semana		
	Asintomático	Sintomático	Mejoría
Tipo de medicamento			
Metoclopropamida	24 (8%)	54 (18%)	21 (7%)
Cisaprida	27 (9%)	18 (6%)	57 (19%)
Mosaprida	51 (17%)	27 (9%)	21 (7%)

$$\chi^2_{\text{Pearson}} = 60,12 \longrightarrow p = 0,000 \longrightarrow \text{Existe asociación}$$

Ha: El resultado del tratamiento depende del fármaco empleado

¿La condición de asintomático se encuentra asociado al uso de Mosaprida?

V de Cramer

## V de Cramer

$$V = \sqrt{\frac{\chi^2}{n(k-1)}} \quad V = \sqrt{\frac{60,12}{300(3-1)}} = 0,317$$

Nominal por nominal	Valor del estadístico	Significación (p)
Coefficiente PHI	0,448	0,000
V de Cramer	0,317	0,000
Nº de casos válidos	300 pacientes	

$H_a$ : La condición de asintomático se encuentra asociado al uso de Mosaprida.

## V de Cramer

- Definición: V de Cramer es una manera de calcular la correlación en los cuadros que tienen más de 2x2 filas y columnas. Se utiliza como posterior a la prueba para determinar la intensidad de la asociación después de que el test de *chi-cuadrado* ha sido significativa. Chi-cuadrado, dice que existe una relación significativa entre las variables, pero no dice nada del grado de significancia ni de su importancia. V de Cramer es un post-test para dar esta información adicional.
- Condiciones: Tablas de más de 2 x 2. Cuando la tabla es de 2 x 2 (tetracórica), utilice el coeficiente PHI.
- Interpretación:
  - $V = 0$   $\longrightarrow$  independencia.
  - $V > 0$   $\longrightarrow$  asociación positiva.
- Intervalo de V : entre 0 y 1.
- Información: Es una medida de la intensidad de la asociación. Cuando se acercan a cero, indican independencia o asociación muy débil entre las variables. Cuando se acercan a uno, indican una relación más fuerte o intensa entre las variables estudiadas.

## Limitaciones del test $\chi^2$ de Pearson como medida de asociación

Una vez que tenemos hallado el valor del  $\chi^2$  podemos preguntarnos si existe o no asociación. No se puede emplear el valor de  $\chi^2$  como una medida de asociación ya que este aumenta con el tamaño muestral. El valor absoluto de  $\chi^2$  no permite tener una idea clara de la fuerza de la asociación ya que sus valores no están acotados en un intervalo definido.

Un coeficiente que sí tiene la propiedad de que sus valores varíen dentro de un intervalo es el COEFICIENTE DE CONTINGENCIA.

### Coeficiente de contingencia

El valor de este coeficiente nunca será inferior a 0 ni superior a 1. El 0 corresponderá a ausencia de asociación, con  $\chi^2 = 0$ . No en todos los casos la dependencia extrema necesariamente debe dar 1, dependerá del número de categorías  $k$ .

## Coeficiente de contingencia (C)

El coeficiente de contingencia se utiliza para saber la asociación de variables cualitativas NOMINALES, que tienen dos o más categorías. Este coeficiente requiere del cálculo previo del estadístico  $\chi^2$  de Pearson.

$$C_{\text{máx}} = \sqrt{\frac{k - 1}{k}}$$

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

Donde:

C = coeficiente de contingencia.

$\chi^2$  = valor de chi-cuadrado de Pearson calculado previamente.

n = tamaño o número de la muestra.

C = 0       $\longrightarrow$       Significa que NO hay asociación.

C > 0,30       $\longrightarrow$       Indica una buena asociación entre las variables.

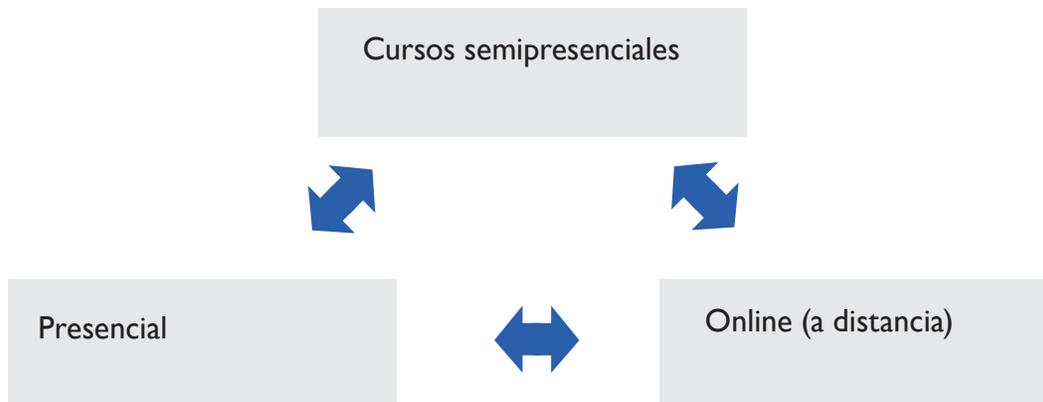
$C_{\text{máx}}$  No alcanza el valor 1 aunque las dos variables están totalmente relacionadas.

## Coeficiente de contingencia (C)

El coeficiente de contingencia se utiliza para saber la asociación de variables cualitativas NOMINALES, que tienen dos o más categorías. Este coeficiente requiere del cálculo previo del estadístico  $\chi^2$  de Pearson.

$$C_{\text{ajustado}} = \frac{C}{C_{\text{máx}}}$$
$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$
$$C_{\text{máx}} = \sqrt{\frac{k-1}{k}}$$

¿La ADHERENCIA al curso de bioestadística depende del modelo docente utilizado (presencial, semipresencial o online)?



## Nominal x Nominales (Tablas > 2 x 2)

V de Cramer

**Coeficiente de contingencia**

Lambda  $\lambda$

Tau de Goodman y Kruskal

Coeficiente de incertidumbre

Variable resultado	Variable predictora TIPO DE CURSO		
	PRESENCIAL	ONLINE	semiPRESENCIAL
Adherencia al curso			
Seguimiento	190	170	180
Abandonos	10 (5%)	30 (15%)	20 (10%)
Total	200	200	200

$\chi^2_{\text{Pearson}} = 11,11 \Rightarrow p = 0,004 \Rightarrow$  **Existe asociación**

$H_a$ : La adherencia al curso es dependiente del tipo de modalidad de enseñanza empleado.

¿El modelo PRESENCIAL se encuentra asociado a una mayor adherencia del alumno al curso?

## ¿El modelo presencial se encuentra asociado a una mayor adherencia del alumno al curso?

Variable resultado	Variable predictora TIPO DE CURSO		
	PRESENCIAL	ONLINE	semiPRESENCIAL
Adherencia al curso			
Seguimiento	190	170	180
Abandonos	10 (5%)	30 (15%)	20 (10%)
Total	200	200	200

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

$$C = \sqrt{\frac{11,11}{11,11 + 600}} = 0,135$$

$$C < 0,30$$

$H_a$ : La adherencia al curso es mayor en la modalidad PRESENCIAL.

Nominal por nominal	Valor del estadístico	Significación (p)
Coefficiente de contingencia C	0,135	0,004
V de Cramer	0,136	0,004
Nº de casos válidos	600 pacientes	

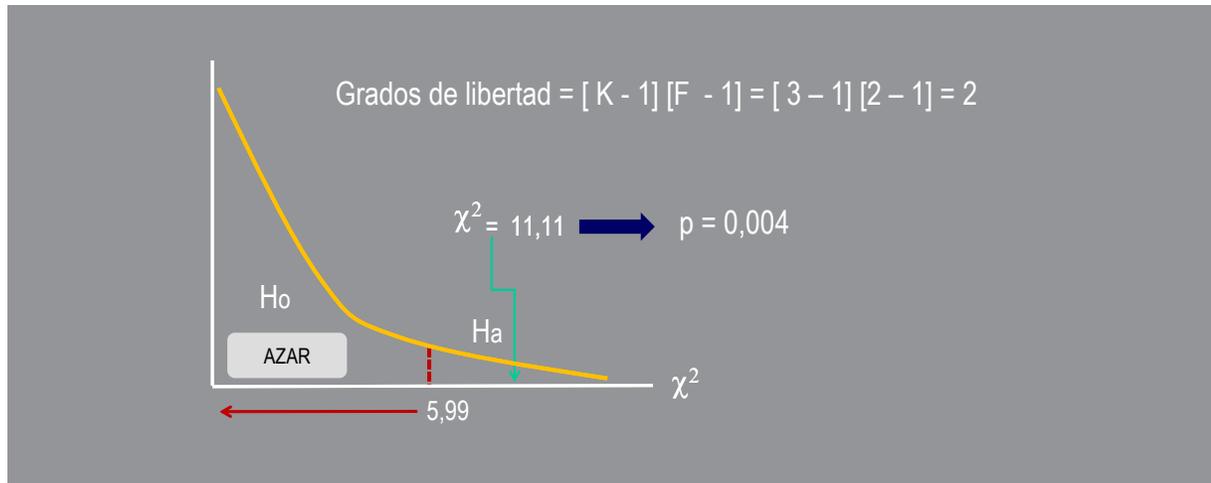
$$C_{\text{máx}} = \sqrt{\frac{k-1}{k}}$$

$$C_{\text{máx}} = \sqrt{\frac{3-1}{3}} = 0,816$$

$$C_{\text{ajustado}} = \frac{0,135}{0,816} = 0,165 \text{ (16,5\%)}$$

$H_a$ : La adherencia al curso es mayor en la modalidad PRESENCIAL.

Decisión: El coeficiente de contingencia como medida de asociación para una escala nominal, es una función de  $\chi^2$ . La decisión de aceptar o rechazar la hipótesis se establece mediante el valor de  $\chi^2 = 11,11$  con dos grados de libertad, y se obtiene una probabilidad en su distribución de  $p < 0,05$ . Entonces se acepta  $H_a$  y se rechaza  $H_o$ ; además, el coeficiente 0,135 es distinto de 0, aunque muy lejos de su valor máximo (0,816), ya que sólo supone el 16,5% de dicho valor, y también está lejos del punto de corte (0,30) por lo que la asociación encontrada es débil.



## Interpretación del coeficiente de contingencia

La asociación existente entre la ADHERENCIA al curso del alumno y el tipo de modalidad docente utilizada es verdadera, en la cual conforme sea el tipo de modalidad empleada, cambia la frecuencia de abandonos de manera significativa, siendo mayor el porcentaje de abandonos en el modelo ONLINE, con un valor de coeficiente de contingencia de 0,135.

PRESENCIAL

## Nominal x Nominales (Tablas > 2 x 2)

V de Cramer

Coefficiente de contingencia

Lambda  $\lambda$

**Tau de Goodman y Kruskal**

Coefficiente de incertidumbre

## Tau de Goodman y Kruskal

Al igual que el coeficiente de incertidumbre y lambda, el coeficiente Tau (de Goodman y Kruskal) es una medida de asociación basada en la reducción proporcional del error. Es una medida que expresa el grado de incertidumbre que conseguimos reducir cuando utilizamos una variable para efectuar pronósticos sobre otra.

Nominal por nominal	Valor Tau	EE	Significación
Tipo de modalidad de curso variable resultado	0,009	0,005	0,004
Rendimiento del curso variable resultado	0,019	0,010	0,004

$H_a$ : La adherencia al curso es mayor en la modalidad PRESENCIAL.

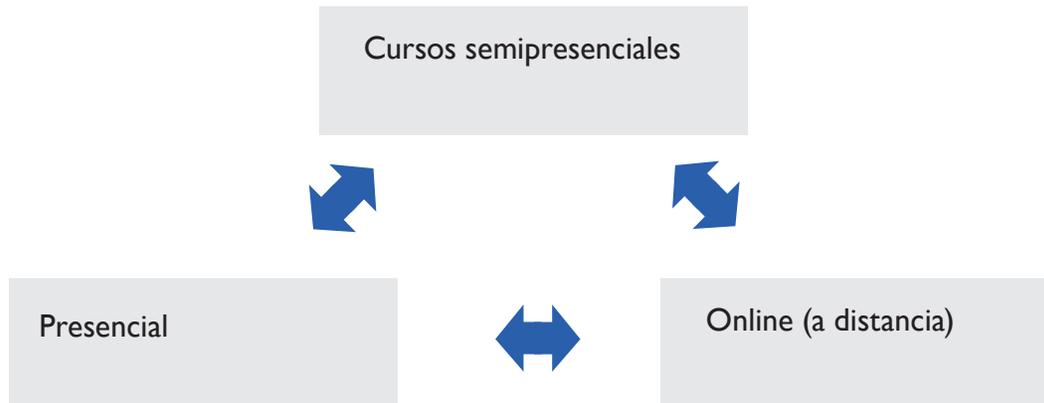
## Coeficiente de incertidumbre

Al igual que el coeficiente Tau y lambda, el coeficiente de incertidumbre es una medida de asociación basada en la reducción proporcional del error. Es una medida que expresa el grado de incertidumbre que conseguimos reducir cuando utilizamos una variable para efectuar pronósticos sobre otra.

Nominal por nominal	Valor $\lambda$	EE	T <sub>aprox</sub>	Significación
Simétrica	0,014	0,008	1,762	0,003
Tipo de modalidad de curso variable resultado	0,009	0,005	1,762	0,003
Adhesión al curso variable resultado	0,030	0,017	1,762	0,003

H<sub>a</sub>: La adherencia al curso es mayor en la modalidad PRESENCIAL.

¿El RENDIMIENTO académico (adquisición de conocimientos) al curso de bioestadística depende del modelo docente utilizado (presencial, semipresencial o online)?



## Nominal x Nominales (Tablas > 2 x 2)

V de Cramer

Coefficiente de contingencia

Lambda  $\lambda$

Tau de Goodman y Kruskal

Coefficiente de incertidumbre

¿El RENDIMIENTO académico (adquisición de conocimientos) al curso de bioestadística depende del modelo docente utilizado (presencial, semipresencial o online)?

Variable resultado	Variable predictora TIPO DE CURSO			Total
	ONLINE	SEMIpresencial	PRESENCIAL	
Evaluación de los conocimientos				
Aprobados	15 (37,5%)	20 (38,4%)	60 (80%)	95
Suspensos	25	32	15	72
Total	40	52	75	167

$$\chi^2_{\text{Pearson}} = 29,68 \Rightarrow p = 0,0000004 \Rightarrow \text{Existe asociación}$$

$H_a$ : El rendimiento del curso es dependiente del tipo de modalidad de enseñanza empleado.

¿El modelo PRESENCIAL se encuentra asociado a un mayor rendimiento del alumno?

## ¿El modelo PRESENCIAL se encuentra asociado a un mayor rendimiento del alumno?

Variable resultado	Variable predictora TIPO DE CURSO			Total
	ONLINE	SEMIpresencial	PRESENCIAL	
Evaluación de los conocimientos				
Aprobados	15 (37,5%)	20 (38,4%)	60 (80%)	95
Suspensos	25	32	15	72
Total	40	52	75	167

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

$$C = \sqrt{\frac{29,68}{29,68 + 167}} = 0,388$$

$$C > 0,30$$

H<sub>a</sub>: El rendimiento del curso es mayor en la modalidad PRESENCIAL.

Nominal por nominal	Valor del estadístico	Significación (p)
Coefficiente de contingencia C	0,388	0,000
V de Cramer	0,421	0,000
Nº de casos válidos	167 alumnos	

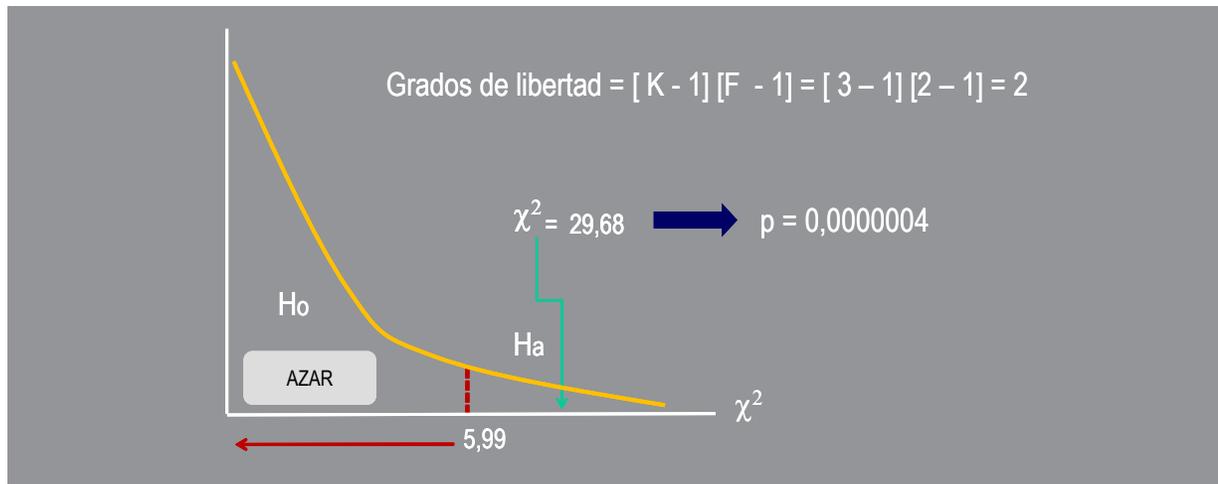
$$C_{\text{máx}} = \sqrt{\frac{k-1}{k}}$$

$$C_{\text{máx}} = \sqrt{\frac{3-1}{3}} = 0,816$$

$$C_{\text{ajustado}} = \frac{0,388}{0,816} = 0,475 \text{ (47,5\%)}$$

$H_a$ : El rendimiento del curso es mayor en la modalidad PRESENCIAL.

Decisión: El coeficiente de contingencia como medida de asociación para una escala nominal, es una función de  $\chi^2$ . La decisión de aceptar o rechazar la hipótesis se establece mediante el valor de  $\chi^2 = 29,68$  con dos grados de libertad, y se obtiene una probabilidad en su distribución de  $p < 0,00001$ , dado que la zona de rechazo es mayor que 0,05. Entonces se acepta  $H_a$  y se rechaza  $H_0$ ; además, el coeficiente 0,388 es distinto de 0, aunque no muy lejos de su valor máximo (0,816), ya que supone el 47,5% del valor máximo.



## Interpretación del coeficiente de contingencia

La asociación existente entre el RENDIMIENTO académico del alumno y el tipo de curso realizado es verdadera, aumentando la frecuencia de aprobados de manera significativa ( $p < 0,00001$ ) en el modelo PRESENCIAL, con un elevado valor de coeficiente de contingencia de 0,388 ( $C > 0,30$ ).

PRESENCIAL

## Coeficiente Lambda $\lambda$

Al igual que el coeficiente de incertidumbre, lambda es una medida que expresa el grado de incertidumbre que conseguimos reducir cuando utilizamos una variable para efectuar pronósticos sobre otra. Lambda tiene tres versiones dos asimétricas (cuando conocemos la variable predictora y la resultado) y una simétrica cuando no sabemos cual es la variable predictora y cual es la variable resultado).

Nominal por nominal	Valor $\lambda$	EE	T <sub>aprox</sub>	Significación
Simétrica	0,238	0,077	2,801	0,005
Tipo de modalidad de curso variable resultado	0,185	0,067	2,527	0,012
Adhesión al curso variable resultado	0,306	0,111	2.331	0,020

H<sub>a</sub>: El rendimiento del curso es dependiente del tipo de modalidad de enseñanza empleado.

## Tau de Goodman y Kruskal

Al igual que el coeficiente de incertidumbre y lambda, el coeficiente Tau (de Goodman y Kruskal) es una medida de asociación basada en la reducción proporcional del error. Es una medida que expresa el grado de incertidumbre que conseguimos reducir cuando utilizamos una variable para efectuar pronósticos sobre otra.

Nominal por nominal	Valor Tau	EE	Significación
Tipo de modalidad de curso variable resultado	0,103	0,034	0,000
Rendimiento del curso variable resultado	0,178	0,057	0,000

$H_a$ : El rendimiento del curso es dependiente del tipo de modalidad de enseñanza empleado.

## Coeficiente de incertidumbre

Al igual que el coeficiente Tau y lambda, el coeficiente de incertidumbre es una medida de asociación basada en la reducción proporcional del error. Es una medida que expresa el grado de incertidumbre que conseguimos reducir cuando utilizamos una variable para efectuar pronósticos sobre otra.

Nominal por nominal	Valor $\lambda$	EE	T <sub>aprox</sub>	Significación
Simétrica	0,106	0,036	2,958	0,000
Tipo de modalidad de curso variable resultado	0,087	0,030	2,958	0,000
Adhesión al curso variable resultado	0,136	0,046	2,958	0,000

H<sub>a</sub>: El rendimiento del curso es dependiente del tipo de modalidad de enseñanza empleado.

## Pregunta de investigación

¿El resultado del tratamiento depende del fármaco empleado?  
¿La condición de asintomático se encuentra asociado al uso de Mosaprida?

Variable PREDICTORA  
Cualitativa policotómica

- Metoclopramida (n = 99)
- Cisaprida (n = 102)
- Mosaprida (n = 99)



Variable RESULTADO  
Cualitativa policotómica

- Asintomático
- Síntomas
- Mejoría clínica

¿Cuál es el test estadístico más adecuado?

## Nominal x Nominales (Tablas > 2 x 2)

V de Cramer

Coefficiente de contingencia

**Lambda  $\lambda$**

Tau de Goodman y Kruskal

Coefficiente de incertidumbre

## ¿El resultado del tratamiento depende del fármaco empleado?

Tipo de medicamento	Variable RESULTADO Resultado del tratamiento a la semana		
	Asintomático	Sintomático	Mejoría
Metoclopropamida	24 (8%)	54 (18%)	21 (7%)
Cisaprida	27 (9%)	18 (6%)	57 (19%)
Mosaprida	51 (17%)	27 (9%)	21 (7%)

$$\chi^2_{\text{Pearson}} = 60,12 \longrightarrow p = 0,000 \longrightarrow \text{Existe asociación}$$

$H_a$ : El resultado del tratamiento depende del fármaco empleado.

¿La condición de asintomático se encuentra asociada al uso de Mosaprida?

Coeficiente Lambda  $\lambda$

## Coeficiente Lambda $\lambda$

Al igual que el coeficiente de incertidumbre, lambda es una medida que expresa el grado de incertidumbre que conseguimos reducir cuando utilizamos una variable para efectuar pronósticos sobre otra. Lambda tiene tres versiones dos asimétricas (cuando conocemos la variable predictor y la resultado) y una simétrica cuando no sabemos cual es la variable predictor y cual es la variable resultado).

Nominal por nominal	Valor $\lambda$	EE	T <sub>aprox</sub>	Significación
Simétrica	0,303	0,048	5,814	0,000
Tipo de Fármaco variable resultado	0,303	0,052	5,108	0,000
Resultado del tratamiento a la semana variable resultado	0,303	0,054	4,899	0,000

H<sub>a</sub>: La condición de asintomático se encuentra asociada al uso de Mosaprida.

## Nominal x Nominales (Tablas > 2 x 2)

V de Cramer

Coefficiente de contingencia

Lambda  $\lambda$

Tau de Goodman y Kruskal

Coefficiente de incertidumbre

## Pregunta de investigación

¿El resultado del tratamiento depende del fármaco empleado?  
¿La condición de asintomático se encuentra asociada al uso de Mosaprida?

Variable PREDICTORA  
Cualitativa policotómica

- Metoclopramida (n = 99)
- Cisaprida (n = 102)
- Mosaprida (n = 99)



Variable RESULTADO  
Cualitativa policotómica

- Asintomático
- Síntomas
- Mejoría clínica

¿Cuál es el test estadístico más adecuado?

## ¿El resultado del tratamiento depende del fármaco empleado?

Variable PREDICTORA Tipo de medicamento	Variable RESULTADO Resultado del tratamiento a la semana		
	Asintomático	Sintomático	Mejoría
Metoclopropamida	24 (8%)	54 (18%)	21 (7%)
Cisaprida	27 (9%)	18 (6%)	57 (19%)
Mosaprida	51 (17%)	27 (9%)	21 (7%)

$$\chi^2_{\text{Pearson}} = 60,12 \longrightarrow p = 0,000 \longrightarrow \text{Existe asociación}$$

$H_a$ : El resultado del tratamiento depende del fármaco empleado.

¿La condición de asintomático se encuentra asociada al uso de Mosaprida?

Coeficiente de incertidumbre

## Coeficiente de incertidumbre

Al igual que el coeficiente Tau y lambda, el coeficiente de incertidumbre es una medida de asociación basada en la reducción proporcional del error. Es una medida que expresa el grado de incertidumbre que conseguimos reducir cuando utilizamos una variable para efectuar pronósticos sobre otra.

Nominal por nominal	Valor	EE	T <sub>aprox</sub>	Significación
Simétrica	0,086	0,023	3,801	0,000
Tipo de Fármaco variable resultado	0,086	0,023	3,801	0,000
Resultado del tratamiento a la semana variable resultado	0,086	0,023	3,801	0,000

H<sub>a</sub>: La condición de asintomático se encuentra asociada al uso de Mosaprida.

## Odds Ratio (OR)

Exposición (Variable predictora)	Enfermedad (Variable resultado)	
	Casos	Controles
(+)	A	B
(-)	C	D

A: Personas enfermas que están expuestas

B: Personas sanas que están expuestas

C: Personas enfermas que no están expuestas

D: Personas sanas que no están expuestas

$$\text{OR} = \frac{\text{Tasa de exposición observada en los CASOS}}{\text{Tasa de exposición observada en los CONTROLES}}$$

$$\text{OR} = \frac{[A / C]}{[B / D]}$$

$$\text{OR} = \frac{[(A * D)]}{[(B * C)]}$$

## Características de la Odds Ratio (OR)

- Medida de asociación.
  - Valores mayores de 1 (asociación positiva) (factor de riesgo).
  - Valores menores de 1 (asociación negativa) (factor de protección).
  - Valor = 1 (no existe asociación).
- Estudios casos-controles.
- OR: es la razón entre la probabilidad de que un suceso ocurra y la probabilidad de que no ocurra.
- OR : cociente entre la tasa de exposición observada en el grupo de casos (A/C) y la tasa de exposición observada en el grupo control (B/D).
- OR: refleja la probabilidad de que los casos (cáncer de pulmón) estén expuestos al factor de estudio (tabaco) comparada con la probabilidad de que lo estén los controles.

## Pregunta de investigación

¿Qué relación y qué medida de asociación existe entre la cloración del agua y el cáncer de vejiga?

Variable PREDICTORA  
CUALITATIVA dicotómica

Cloración del agua



Variable RESULTADO  
CUALITATIVA dicotómica

Cáncer de vejiga

¿Cuál es el test estadístico más adecuado?

## Odds Ratio (OR)

Exposición	Cáncer de vejiga	
	Casos	Controles
Cloración (+)	A	B
Cloración (-)	C	D

A: Personas enfermas de cáncer de vejiga expuestas al cloro

B: Personas sanas expuestas al cloro

C: Personas enfermas de cáncer de vejiga no expuestas al cloro

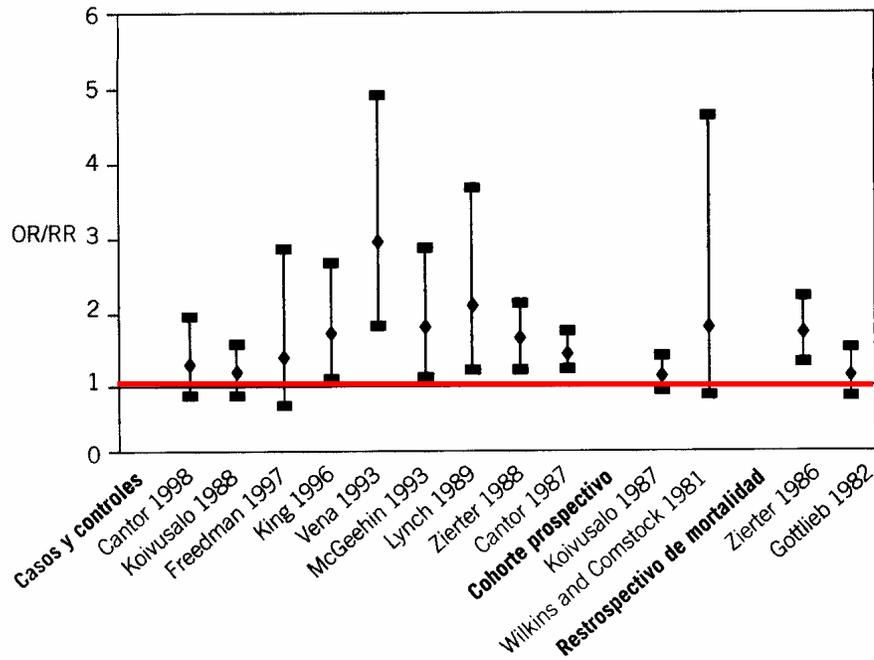
D: Personas sanas no expuestas al cloro

$$\text{OR} = \frac{\text{Tasa de CLORACIÓN observada en los enfermos de CÁNCER de vejiga (casos)}}{\text{Tasa de CLORACIÓN observada en SANOS (controles)}}$$

$$\text{OR} = \frac{[A / C]}{[B / D]}$$

$$\text{OR} = \frac{[(A * D)]}{[(B * C)]}$$

# Influencia de la metodología de la evaluación de la exposición



## COMUNICACIÓN BREVE

### **Actividad laboral agrícola y estrongiloidiasis. Estudio caso-control**

D. Rodríguez Calabuig\*, R. Igual Adell\*\*, C. Oltra Alcaraz\*\*\*, P. Sánchez Sánchez\*\*\*\*,  
M. Bustamante Balen\*\*\*\*\*, F. Parra Godoy\* y E. Nagore Enguidanos\*\*\*\*\*

\*Centro de Salud Jaume Roig, Oliva. Valencia. \*\*Servicio de Microbiología. \*\*\*Centro de Salud Pública.  
Unidad de Salud Laboral. Alcoy. Alicante. \*\*\*\*Medicina de Familia. Olocau. Valencia. \*\*\*\*\*Especialista Digestivo.  
\*\*\*\*\*Dermatólogo. Hospital F. Borja. Gandía. Valencia.

Rodríguez Calabuig D, Igual Adell R, Oltra Alcaraz C, Sánchez Sánchez P, Bustamante Balen M, Parra Godoy F y Nagore Enguidanos E. Actividad laboral agrícola y estrongiloidiasis. Estudio caso-control. Rev Clín Esp 2001; 201:81-84

## Ejemplo de uso de la OR:

### Estudio caso-control

- CASOS: Individuos con *Strongiloides stercoralis* en las heces ante la presencia de eosinofilia en el hemograma.
- CONTROLES: Individuos de igual edad y sexo sin eosinofilia ni parásitos en heces.

Rodríguez Calabuig D, Igual Adell R, Oltra Alcaraz C, Sánchez Sánchez P, Bustamante Balen M, Parra Godoy F y Nagore Enguidanos E. Actividad laboral agrícola y estrogiloidiasis. Estudio caso-control. Rev Clín Esp 2001; 201:81-84.

# Odds Ratio (OR)

## Ejemplo estudio caso-control

TABLA 1  
**Características generales. Patología crónica,  
 síntomas y uso de corticoides**

Datos generales	Casos 47	Controles 47	Significación OR (IC 95%)
Edad media	66 ± 10	66 ± 10	
IMC	28 ± 4	27 ± 3	
Nacidos en Oliva/comarca	45	42	
Viajes al extranjero	30	24	
Sintomatología			
Tos	23 (49%)	12 (25%)	2,80 (1,08-7,37) ←
Prurito	22 (47%)	5 (11%)	7,39 (2,29-27,60) ←
Dispepsia	13 (28%)	8 (17%)	NS
Enfermedades asociadas			
Pulmonares	14	8	
Digestivas	17	5	
Tratamiento con corticoides			
3 años anteriores	7 (15%)	2 (4%)	

IMC: índice de masa corporal; NS: no significativo.

## Odds Ratio (OR)

### Ejemplo estudio caso-control

**TABLA 2**  
**Situación laboral y diferentes tipos de cultivos**

	<b>Casos n (%)</b>	<b>Controles n (%)</b>	<b>Significación OR (IC 95%)</b>
Pensionistas	36 (77%)	38 (81%)	
Agricultores	32 (68%)	31 (66%)	
(años de dedicación)	41 ± 16	35 ± 19	
<b>Tipos de cultivo</b>			
Cultivo de arroz	25 (53%)	13 (28%)	2,97 (1,16-7,71) ◀
Limpieza de acequias	17 (36%)	9 (19%)	2,39 (0,85-6,82) ◀
Cultivo de cítricos	30 (64%)	28 (60%)	NS
Cultivo de hortalizas	9 (19%)	15 (32%)	NS

OR: odds ratio; IC: intervalo de confianza; NS: no significativo.

# Actividad laboral agrícola y estrombiloidiasis

## Estudio caso-control

### Discusión

“Un gran número de personas trabajaban descalzas en la siembra, cuidando y recolectando el arroz, con los pies hundidos en el fango durante largos períodos de tiempo, lo que facilitaría la entrada del nematodo. Los resultados obtenidos, especificando el tipo de actividad, muestran una asociación de la estrombiloidiasis con el cultivo de arroz OR de 2,97 (IC 95%: 1,16-7,71), lo que representa un mayor riesgo de infestación. En la limpieza de acequias se obtuvo también una tendencia casi significativa OR de 2,39 (IC 95%: 0,85-6,82), en cambio en el cultivo de cítricos y de hortalizas, que habitualmente no se realizaban descalzos ni en tierras fangosas, no se han encontrado estas diferencias”.

Rodríguez Calabuig D, Igual Adell R, Oltra Alcaraz C, Sánchez Sánchez P, Bustamante Balen M, Parra Godoy F y Nagore Enguidanos E. Actividad laboral agrícola y estrombiloidiasis. Estudio caso-control. Rev Clín Esp 2001; 201:81-84.

## Mito I: “La sonrisa de la esfinge”

En las tablas 1 y 2 presentadas se han incluido sólo el valor de la OR y su IC95%, pero no se ha incluido el valor del estadístico (test de contraste de hipótesis) ni la significación estadística  $p$  para cada variable predictora presuntamente significativa.

¿Cuál es el valor de la  $p$ ?

¿Por qué dices que la práctica de la limpieza de acequias está relacionada con la estrongiloidiasis cuando el valor del IC95% de la OR incluye a la unidad?

¿Cuál es el test de contraste de hipótesis que has empleado en cada asociación?

Test de  $\chi^2$  Pearson

He calculado el valor del estadístico  $\chi^2$  Pearson NO hay la significación estadística.

¿Será un error beta  $\beta$ ?

¿Cuál es la potencia estadística del estudio (valor  $\beta$ )?

## ¿Qué relación y qué medida de asociación existe entre la práctica del cultivo de arroz y la estrongiloidiasis?

Test de $\chi^2$ de Pearson	Variable RESULTADO Estrongiloidiasis		
	Casos	Controles	Total
Variable PREDICTORA Cultivo de arroz			
Si	25 (53%)	13 (28%)	38
No	22	34	56
Total	47	47	94

$$\chi^2_{\text{Pearson}} = 6,36; p = 0,012 \quad \text{OR} = 2,97 (1,16-7,71)$$

$$\text{Q de Yule} = 0,50$$

## Resumen del procesamiento informático de los datos

Cálculos del programa estadístico tabla 2 x 2	Valor	Grados de libertad	Contraste bilateral	Contraste unilateral
$\chi^2$ Pearson	6,361	1	0,012	
Corrección de Yates	5,345	1	0,021	
Razón de verosimilitud	6,447	1	0,011	
Test exacto de Fisher			0,020	0,010
Nº Casos válidos	94			

$H_0$  (hipótesis nula) = La práctica del cultivo de arroz NO está asociada a la estrongiloidiasis en el grupo estudiado.

$H_a$  (hipótesis alternativa) = La práctica del cultivo de arroz está asociada a la estrongiloidiasis en el grupo estudiado.

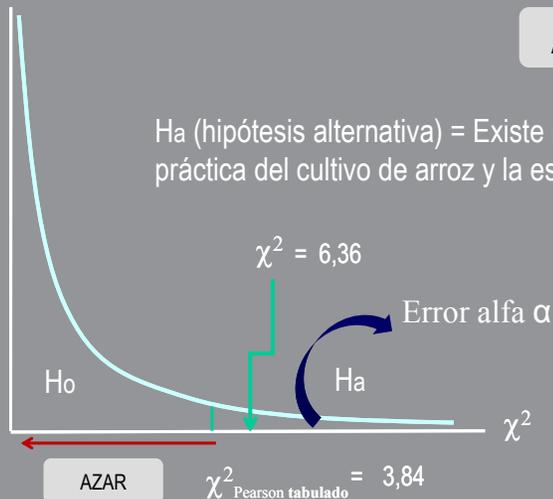


## Test de $\chi^2$ de Pearson (1 grado de libertad)

$$\chi^2_{\text{Pearson calculado}} = 6,36 > \chi^2_{\text{Pearson tabulado}} = 3,84$$

AZAR

$H_a$  (hipótesis alternativa) = Existe relación entre la práctica del cultivo de arroz y la estrongiloidiasis.



AZAR

$$\chi^2_{\text{Pearson tabulado}} = 3,84$$

## Pregunta de investigación

¿Qué relación y qué medida de asociación existe entre la práctica del cultivo del arroz y la estrongiloidiasis?

Variable PREDICTORA  
CUALITATIVA dicotómica

Cultivo de arroz



Variable RESULTADO  
CUALITATIVA dicotómica

Estrongiloidiasis

$$\chi^2_{\text{Pearson}} = 6,36; \quad \Rightarrow \quad p = 0,012$$

$$\text{OR} = 2,97$$

$$\text{IC}_{95\%} = 1,16 - 7,71$$

$$\text{Q de Yule} = 0,50$$

## ¿Qué relación y qué medida de asociación existe entre la práctica de la limpieza de acequias y la estrogiloidiasis?

Test de $\chi^2$ de Pearson	Variable RESULTADO Estrongiloidiasis		
	Casos	Controles	Total
Variable PREDICTORA Limpieza de acequia			
Si	17 (36%)	9 (19%)	26
No	30	38	68
Total	47	47	94

$$\chi^2 = 3,40 ; \Rightarrow p = 0,065$$

No existe asociación

## Resumen del procesamiento informático de los datos

Cálculos del programa estadístico tabla 2 x 2	Valor	Grados de libertad	Contraste bilateral	Contraste unilateral
$\chi^2$ Pearson	3,403	1	0,065	
Corrección de Yates	2,605	1	0,107	
Razón de verosimilitud	3,445		0,063	
Test exacto de Fisher		1	0,106	0,053
Nº Casos válidos	94			

$H_0$  (hipótesis nula) = La limpieza de acequias NO está asociada a la strongiloidiasis en el grupo estudiado. 

$H_a$  (hipótesis alternativa) = La limpieza de acequias está asociada a la strongiloidiasis en el grupo estudiado.

## Pregunta de investigación

¿Qué relación y qué medida de asociación existe entre la práctica de la limpieza de acequias y la estrongiloidiasis?

Variable PREDICTORA  
CUALITATIVA dicotómica

Limpieza de acequias



Variable RESULTADO  
CUALITATIVA dicotómica

Estrongiloidiasis

$$\chi^2 = 3,40 ; \Rightarrow p = 0,065$$

$$OR = 2,39$$

$$IC_{95\%} = 0,85 - 6,82$$

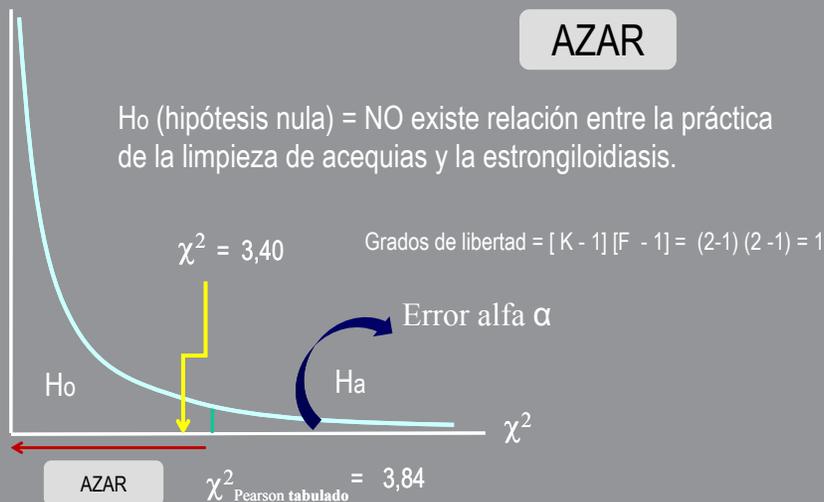
$$Q \text{ de Yule} = 0,41$$

## Test de $\chi^2$ de Pearson (1 grado de libertad)

$$\chi^2_{\text{Pearson calculado}} = 3,40 < \chi^2_{\text{Pearson tabulado}} = 3,84$$

AZAR

$H_0$  (hipótesis nula) = NO existe relación entre la práctica de la limpieza de acequias y la strongiloidiasis.



## Riesgo Relativo (RR)

Exposición (F)	Enfermedad (E)	
	Casos	Controles
Tabaco (+)	A	B
Tabaco (-)	C	D

A: Personas enfermas de cáncer de pulmón fumadoras

B: Personas sanas fumadoras

C: Personas enfermas de cáncer de pulmón no fumadoras

D: Personas sanas no fumadoras

$$\begin{aligned}
 \text{RR} &= \frac{[A / (A + B)]}{[C / (C + D)]} \\
 &\downarrow \\
 \text{RR} &= \frac{[A * D]}{[B * C]}
 \end{aligned}$$

$A \ll B \quad C \ll D$   
 $A + B \approx B \quad C + D \approx D$

Es la probabilidad de que una enfermedad (E) se desarrolle en el grupo expuesto (F) en relación a la del grupo no expuesto.

## Pregunta de investigación

¿Qué relación y qué medida de asociación existe entre la exposición al humo del tabaco durante la gestación y el bajo peso al nacimiento del RN?

Variable PREDICTORA  
CUALITATIVA dicotómica

Exposición al humo del  
tabaco



Variable RESULTADO  
CUALITATIVA dicotómica

Peso del RN

$$RR = 2,71 (1,38-5,34)$$

$$Q \text{ de Yule} = 0,47$$

$$\chi^2_{\text{Pearson}} = 9.03; p = 0.00265$$

Martín TR, Bracken MB. Association of low birth weight with passive smoke exposure in pregnancy. Am J. Epidemiol 1986; 124: 633-642.

## Riesgo Relativo (RR)

Exposición (F)	Enfermedad (E)	
	Casos	Controles
Tabaco (+)	A (20)	B (833)
Tabaco (-)	C (14)	D (1606)

A: RN de bajo peso de madres expuestas al tabaco

B: RN normopeso de madres expuestas al tabaco

C: RN de bajo peso de madres NO expuestas al tabaco

D: RN normopeso de madres NO expuestas al tabaco

En la Tabla, exponemos los resultados de un estudio de seguimiento donde 853 mujeres estuvieron pasivamente expuestas al humo del tabaco durante la gestación y 1620 no lo estuvieron, y su asociación con el bajo peso al nacer.

$$\chi^2 = 9.03; p = 0.00265$$

$$95\% \text{ IC } (1.38; 5.34)$$

$$Q \text{ de Yule} = 0,47$$

$$RR = \frac{I_e}{I_o} = \frac{20/853}{14/1620} = 2,713$$

El RR es 2.71, significa que las expuestas al factor F (exposición pasiva al tabaco) tienen 2,71 veces más probabilidades de tener niños de bajo peso que las no expuestas.

Martín TR, Bracken MB. Association of low birth weight with passive smoke exposure in pregnancy. Am J Epidemiol 1986; 124: 633-642.

## Pregunta de investigación

¿Qué relación y qué medida de asociación existe entre el sedentarismo (factor de riesgo) y el desenlace (muerte) en pacientes que han sufrido un IAM?

Variable PREDICTORA  
CUALITATIVA dicotómica

Ejercicio moderado/  
Sedentarismo

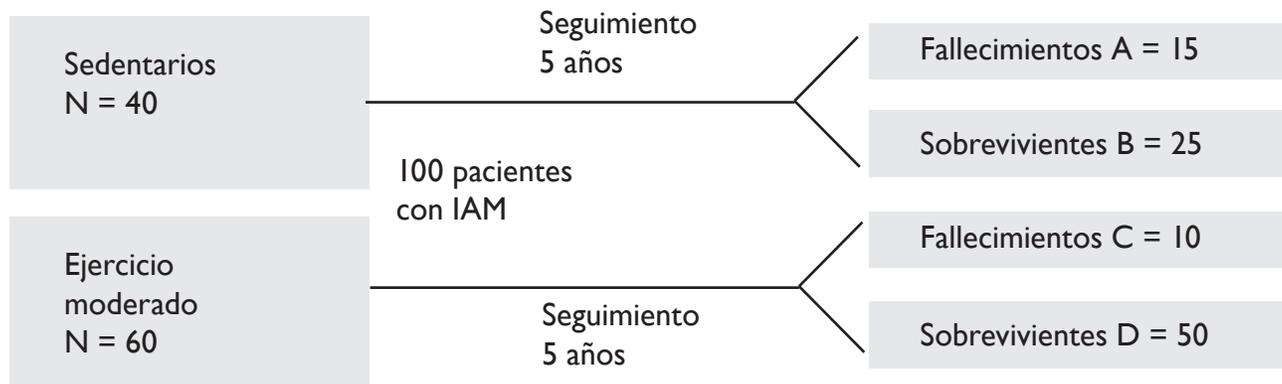


Variable RESULTADO  
CUALITATIVA dicotómica

Muerte / Supervivencia

100 pacientes  
con IAM

Evolución de 100 pacientes con IAM a los 5 años en función del ejercicio físico realizado.  
¿Existe asociación entre el factor de riesgo (SEDENTARISMO) y el desenlace (MUERTE)?  
¿Cuál es la magnitud de dicha asociación?



Evolución de 100 pacientes con IAM a los 5 años en función del ejercicio físico		Desenlace (E) Variable resultado	
Exposición (F) Variable predictora		Muerte	Sobreviviente
Sedentarismo (+)		A = 15	B = 25
Sedentarismo (-)		C = 10	D = 50

¿Existe asociación entre el factor de riesgo (SEDENTARISMO) y el desenlace (MUERTE)?

100 pacientes con IAM

A: Fallecidos con sedentarismo

B: Sobrevivientes sedentarios

C: Fallecidos deportistas

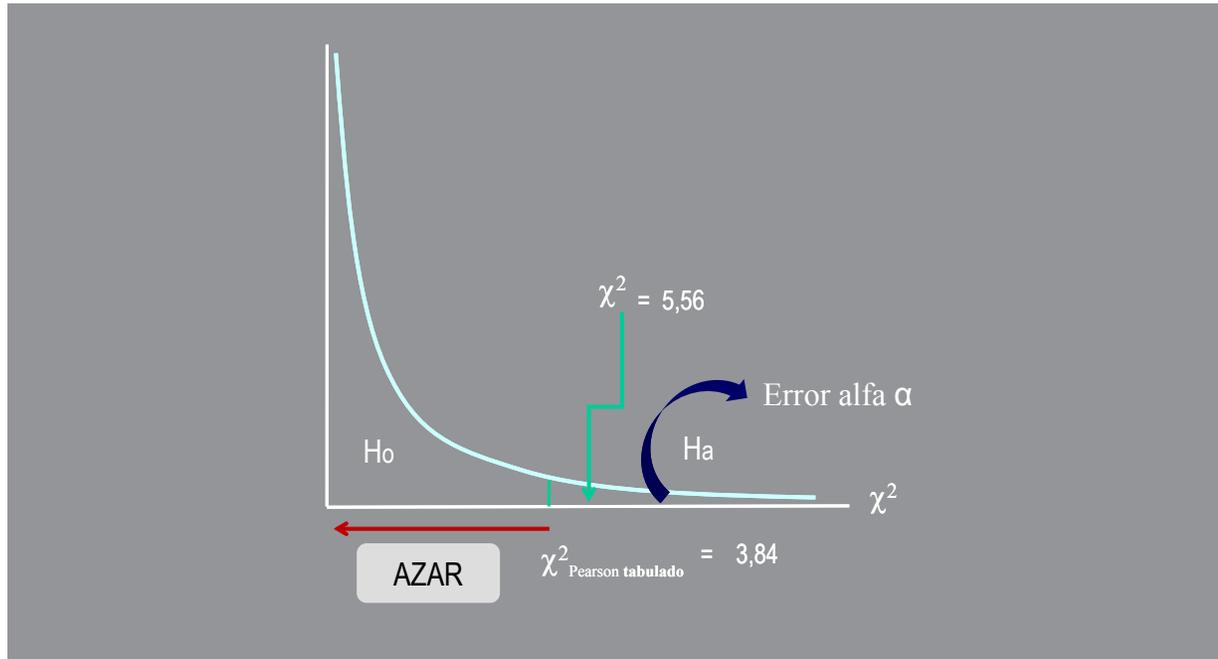
D: Sobrevivientes deportistas

¿Cuál es la prueba estadística más adecuada?

## Test de $\chi^2$ de Pearson

Cálculos del programa estadístico	Valor	Grados de libertad	Contraste bilateral	Contraste unilateral
$\chi^2$ Pearson (*)	5,56	1	0,018	
Corrección de Yates	4,50	1	0,034	
Razón de verosimilitud	5,47		0,019	
Test exacto de Fisher		1	0,033	0,017
Nº Casos válidos	100			

$H_a$ : Existe asociación entre el factor de riesgo (SEDENTARISMO) y el desenlace (MUERTE).



$$\chi^2_{\text{Pearson calculado}} = 5,56 > \chi^2_{\text{Pearson tabulado}} = 3,84 \quad \text{Existe asociación} \quad \rightarrow \quad p < 0,05$$

Evolución de 100 pacientes con IAM a los 5 años en función del ejercicio físico	Desenlace (E) Variable resultado	
	Muerte	Sobreviviente
Exposición (F) Variable predictora		
Sedentarismo (+)	A = 15	B = 25
Sedentarismo (-)	C = 10	D = 50

100 pacientes con IAM

A: Fallecidos con sedentarismo

B: Sobrevivientes sedentarios

C: Fallecidos deportistas

D: Sobrevivientes deportistas

¿Cuál es la intensidad o magnitud de dicha asociación?

$$RR = \frac{[ A / (A + B) ]}{[ C / (C + D) ]}$$

Es la probabilidad de morir (E) en el grupo de sedentarios (expuestos al factor de riesgo) en relación con la probabilidad de muerte en deportistas (grupo no expuesto).

$$\text{Riesgo Relativo (RR)} = \frac{\text{Tasa de Incidencia de muertes en sedentarios (en expuestos)}}{\text{Tasa de Incidencia de muertes en deportistas (NO expuestos)}}$$

$$\text{RR} = \frac{I_e}{I_o}$$

$$\text{RR} = \frac{[A / (A + B)]}{[C / (C + D)]} \quad \text{RR} = \frac{[15 / (15 + 25)]}{[10 / (10 + 50)]} = 2,25$$

$$\text{RR} = 2,25$$

$$\text{Q de Yule} = 0,50$$

$$\chi^2_{\text{Pearson}} = 5,56; \quad p = 0,018$$

## Pregunta de investigación

¿Qué relación y qué medida de asociación existe entre el ejercicio físico moderado (factor de riesgo) y el desenlace (muerte) en pacientes que han sufrido un IAM?

Variable PREDICTORA  
CUALITATIVA dicotómica

Ejercicio moderado/  
Sedentarismo

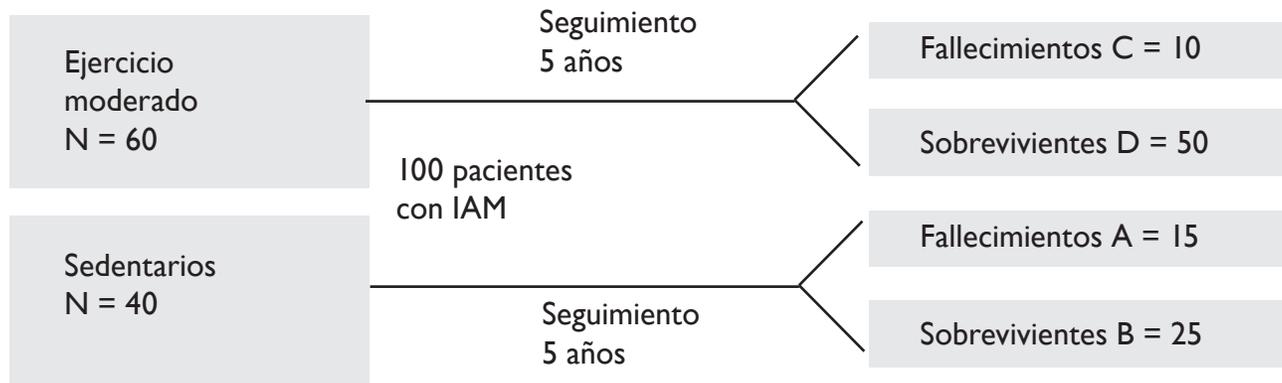


Variable RESULTADO  
CUALITATIVA dicotómica

Muerte / Supervivencia

100 pacientes  
con IAM

Evolución de 100 pacientes con IAM a los 5 años en función del ejercicio físico realizado.  
¿Existe asociación entre el factor de riesgo (EJERCICIO FÍSICO) y el desenlace (MUERTE)? ¿Cuál es la magnitud de dicha asociación?



Evolución de 100 pacientes con IAM a los 5 años en función del ejercicio físico	Desenlace (E) Variable resultado	
	Muerte	Sobreviviente
Exposición (F) Variable predictora		
Ejercicio físico (+)	A = 10	B = 50
Ejercicio físico (-)	C = 15	D = 25

100 pacientes con IAM

A: Fallecidos con ejercicio físico moderado

B: Sobrevivientes con ejercicio físico moderado

C: Fallecidos sedentarios (NO deportistas)

D: Sobrevivientes sedentario (NO deportistas)

$$RR = \frac{[ A / (A + B) ]}{[ C / (C + D) ]}$$

Es la probabilidad de morir (E) en el grupo de deportistas (expuestos al factor de riesgo) en relación con la probabilidad de muerte en sedentarios -NO deportistas- (grupo no expuesto).

$$\text{Riesgo Relativo (RR)} = \frac{\text{Tasa de Incidencia de muertes en deportistas moderados (en expuestos)}}{\text{Tasa de Incidencia de muertes en sedentarios (NO expuestos)}}$$

$$\text{RR} = \frac{I_e}{I_o}$$

$$\text{RR} = \frac{[A / (A + B)]}{[C / (C + D)]} \quad \text{RR} = \frac{[10 / (10 + 50)]}{[15 / (15 + 25)]} = 0,44$$



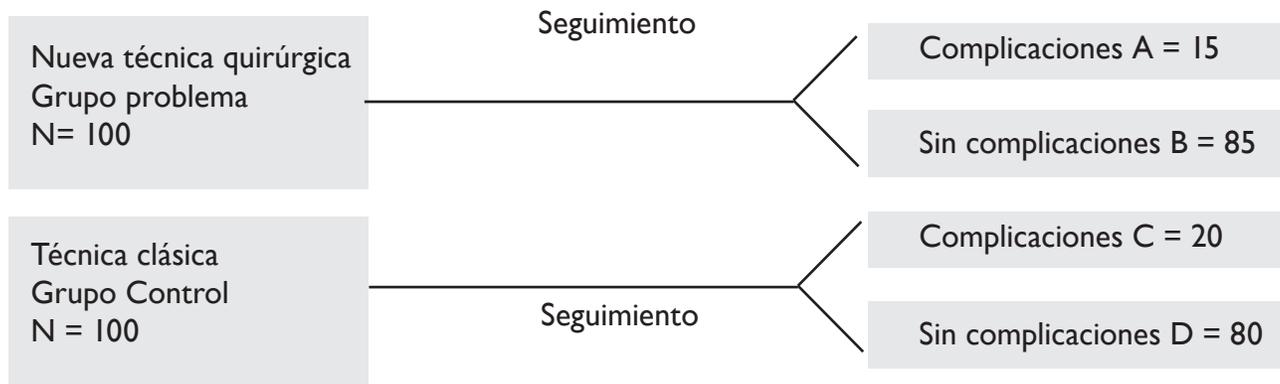
$$\text{RR} = 0,44$$



$$\text{Q de Yule} = - 0,50$$

$$\chi^2_{\text{Pearson}} = 5,56; p = 0,018$$

Evolución de 100 pacientes operados con una NUEVA técnica quirúrgica frente a 100 pacientes operados con la técnica CLÁSICA.



$$\text{Riesgo Relativo (RR)} = \frac{\text{Tasa de Incidencia de complicaciones con la NUEVA técnica}}{\text{Tasa de Incidencia de complicaciones con la técnica CLASICA}}$$

$$RR = \frac{I_e}{I_o}$$

$$RR = \frac{[A / (A + B)]}{[C / (C + D)]} \quad RR = \frac{[15 / (15 + 85)]}{[20 / (20 + 80)]} = 0,75$$

$$RR = 0,75$$

$$RRR = [1 - RR] * 100$$

$$RRR = 25\%$$

$$RRA = [I_N - I_C] * 100$$

$$RRA = 5\%$$

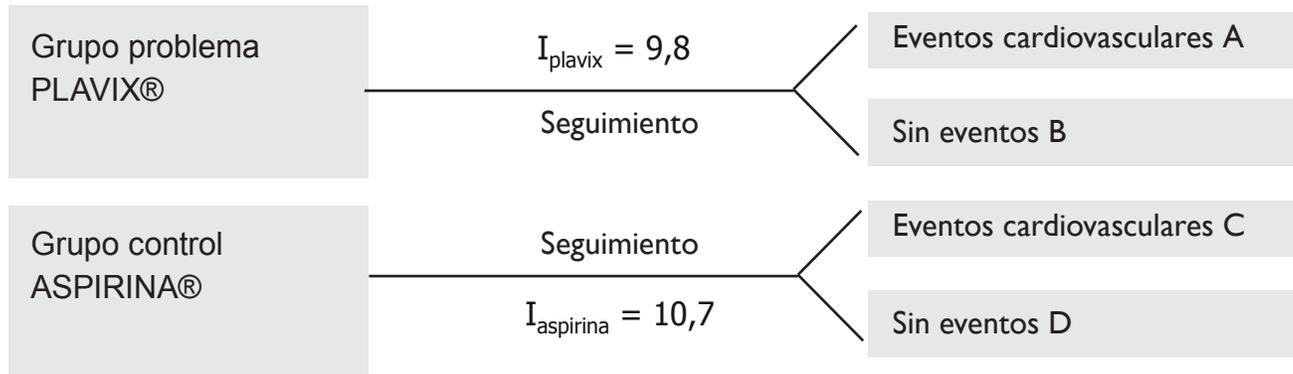
	Fórmula	Evento en estudio
Incidencia de complicaciones quirúrgicas con la NUEVA técnica	$I_N = [ A / (A + B) ]$	15%
Incidencia de complicaciones quirúrgicas con la técnica CLÁSICA	$I_C = [ C / (C + D) ]$	20%
RIESGO RELATIVO (RR)	$RR = \frac{[ A / (A + B) ]}{[ C / (C + D) ]}$	0,75
Reducción del RIESGO RELATIVO (RRR)	$RRR = [ 1 - RR ] * 100$	25%
Reducción ABSOLUTA del RIESGO (RAR)	$RAR = [ I_N - I_C ] * 100$	5%
Número de pacientes que se necesita operar (NNT)	$NNT = [ 1 / RAR ] * 100$	20

Escenario clínico: Un día muy ocupado en su centro y usted está sentado con las piernas en alto, hojeando un número reciente de “Médicos en su tiempo libre”. Se encuentra con un anuncio de PLAVIX® que indica que este medicamento produce una reducción del riesgo de eventos cardiovasculares en un 8% comparado con aspirina. Usted piensa si debería cambiar todos sus pacientes a Plavix®.

Eventos resultantes del análisis del estudio CAPRIE			
	PLAVIX®	ASPIRINA®	DIFERENCIA
Pacientes	9599	9586	
ICTUS	4,6%	4,8%	- 0,2%
IAM	2,9%	3,5%	- 0,6%
Otras muertes vasculares	2,4%	2,4%	0%
Total de eventos vasculares	9,8%	10,7%	- 0,9%

Como se muestra en el cuadro PLAVIX® se asoció con una menor incidencia de eventos resultantes de cualquier clase. La reducción del Riesgo Relativo fue de 8,7% con un IC95% = 0,3-16,5% p=0,045. Sin embargo la reducción del riesgo absoluto (9,8% contra 10,7%) fue de tan sólo 0,9%.

Aparición de eventos cardiovasculares después de un periodo de seguimiento en pacientes tratados con PLAVIX (grupo problema) frente a los pacientes tratados con ASPIRINA (grupo control).



$$\text{Riesgo Relativo (RR)} = \frac{\text{Tasa de Incidencia de eventos cardiovasculares con el PLAVIX}}{\text{Tasa de Incidencia de eventos cardiovasculares con ASPIRINA}}$$

$$RR = \frac{I_P}{I_A}$$

$$RR = \frac{I_P}{I_A} = \frac{9,8}{10,7} = 0,916 \quad RR = 0,916$$

$$RRR = [ 1 - RR ] * 100$$

$$RRR = 8,7\%$$

$$IC_{95\%} = 0,3-16,5\% \quad p = 0,045$$

$$RAR = [ I_P - I_A ] * 100$$

$$RAR = 0,9\%$$

$$NNT = [ 1 / RAR ] * 100$$

$$NNT = 111 \text{ pacientes}$$

	Fórmula	Evento en estudio
Incidencia de eventos cardiovasculares con el PLAVIX	$I_{\text{Plavix}} = [ A / (A + B) ]$	Complicaciones quirúrgicas 9,8%
Incidencia de eventos cardiovasculares con la ASPIRINA	$I_{\text{Aspirina}} = [ C / (C + D) ]$	10,7%
RIESGO RELATIVO (RR)	$RR = \frac{[ A / (A + B) ]}{[ C / (C + D) ]}$	0,916
Reducción del RIESGO RELATIVO (RRR)	$RRR = [ 1 - RR ] * 100$	8,7%
Reducción ABSOLUTA del RIESGO (RAR)	$RAR = [ I_N - I_C ] * 100$	0,9%
Número de pacientes que se necesita operar (NNT)	$NNT = [ 1 / RAR ] * 100$	111
Aumento del coste mensual del Plavix respecto a la aspirina	Ensayo CAPRIE (Lancet)	+ 112 \$ (94 €)

## Resolución del escenario

El 8,7% de reducción del riesgo que publica el artículo es la Reducción del Riesgo Relativo (RRR). La Reducción Absoluta del Riesgo (RAR) fue un ridículo 0,9%, con un NNT muy elevado (111), con un amplio IC95% (de 57 a 2500). El resultado fue estadísticamente significativo casi marginalmente,  $p = 0,045$ , donde el IC95% de la RRR fue del 0.3% al 16.5%, que ni siquiera aparecía en el texto (sólo a pie de tabla). Aunque es estadísticamente significativo, sin embargo la significación CLÍNICA es muy cuestionable. La RRR del 8,7% de Plavix frente a aspirina se dibuja en un gráfico con una área que es casi igual que la del 25% del RRR de la aspirina frente al placebo, violando claramente de la ley de Tufte (factor mentiroso de Tufte = Tamaño del efecto mostrado en el gráfico / Tamaño del efecto en los datos  $\neq 1$ ).

La publicación es un estudio real, el ensayo CAPRIE publicado en Lancet, y sin embargo, después de todo usted lógicamente decide, ahorrarles a sus pacientes y a la Seguridad Social los 112 dólares (94€) al mes que cuesta Plavix y seguir tratando a sus pacientes con aspirina.

**Objetivo:** Evaluar la eficacia de la disminución de colesterol con simvastatina en la reducción de la mortalidad y morbilidad en pacientes con enfermedad coronaria (EC).

**Diseño:** Ensayo clínico controlado, aleatorizado en dos ramas, a doble ciego con placebo y un seguimiento medio de 5,4 años.

**Ámbito:** Pacientes procedentes de 95 centros médicos de Escandinavia.

**Sujetos de estudio:** 4.444 pacientes de entre 35 y 70 años, 81% hombres, 51% mayores de 60 años, con antecedentes de infarto de miocardio previo a angina de pecho, colesterol sérico de 5,5 a 8,8 mmol/l, concentraciones medias de HDL de 1,18 mmol/l, LDL 4,87 mmol/l, triglicéridos 1,50 mmol/l y dieta hipolipemiente concurrente.

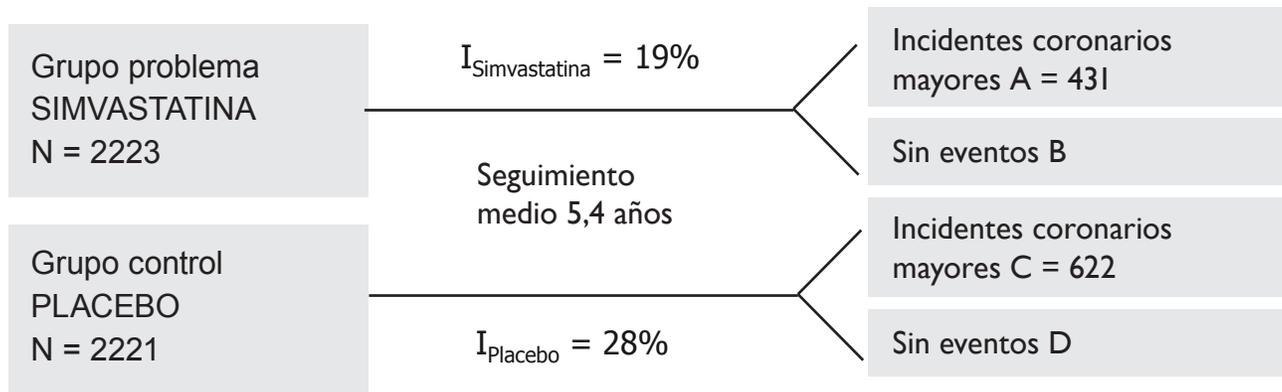
**Intervención:** 2.223 sujetos se les administraron 20 mgr de simvastatina y a 2.221 el correspondiente placebo.

**Parámetros principales de eficacia:** Mortalidad total, mortalidad de causa coronaria, infarto de miocardio no letal probable o definitivo verificado.

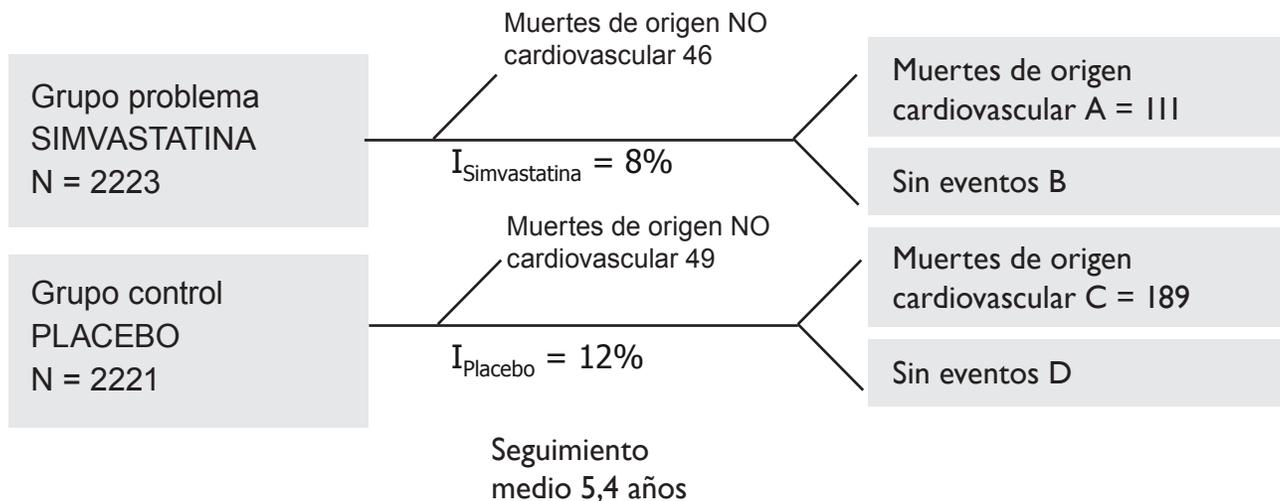
**Resultados principales:** Durante los 5,4 años en promedio de seguimiento la simvastatina produjo unos cambios medios de colesterol total, LDL colesterol y HDL colesterol de -25%, -35% y +8%, respectivamente. 256 (12%) sujetos en el grupo de placebo murieron en comparación con 182 (8%) en el grupo de simvastatina. Hubo 189 muertes de origen coronario en el grupo de placebo y 111 en el de simvastatina y 49 y 46 muertes de origen no cardiovascular en dichos grupos, respectivamente. 622 sujetos (28%) en el grupo con placebo y 431 (19%) en el de simvastatina tuvieron uno o más incidentes coronarios mayores.

**Conclusiones:** La disminución de la concentración de colesterol con simvastatina reduce la mortalidad total y la incidencia de incidentes cardiovasculares mayores.

Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandivarian Simvastatin Survival Study (4S). Lancet 1994; 334: 1383-1389.



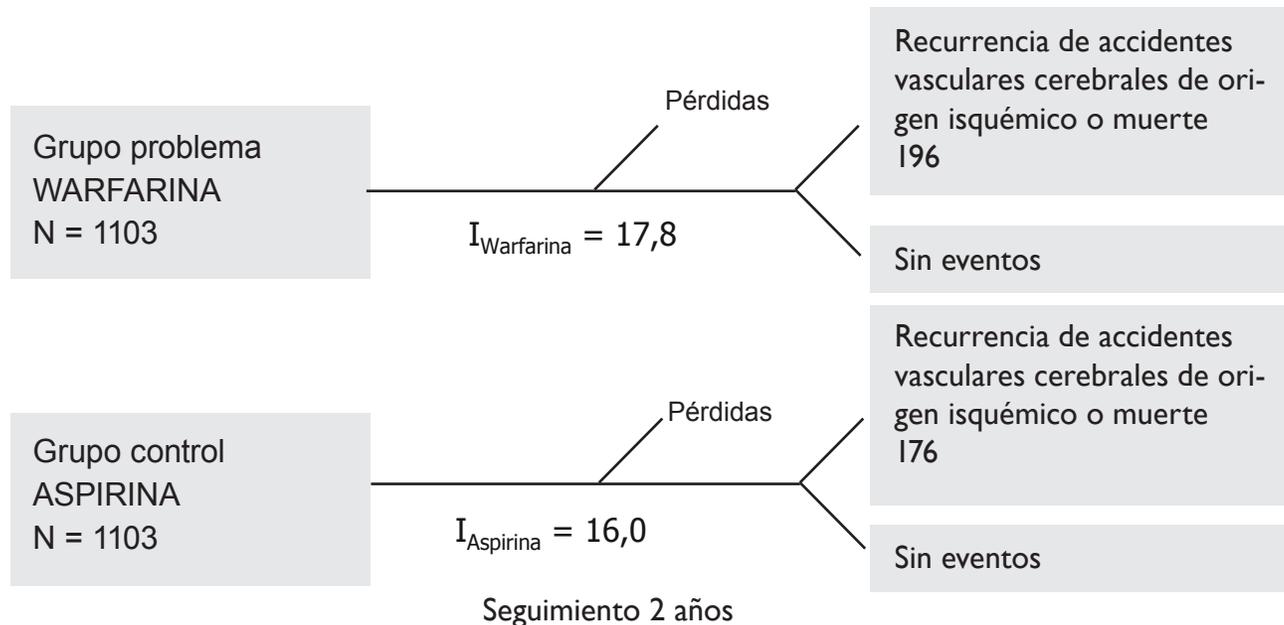
Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). Lancet 1994; 334: 1383-1389.



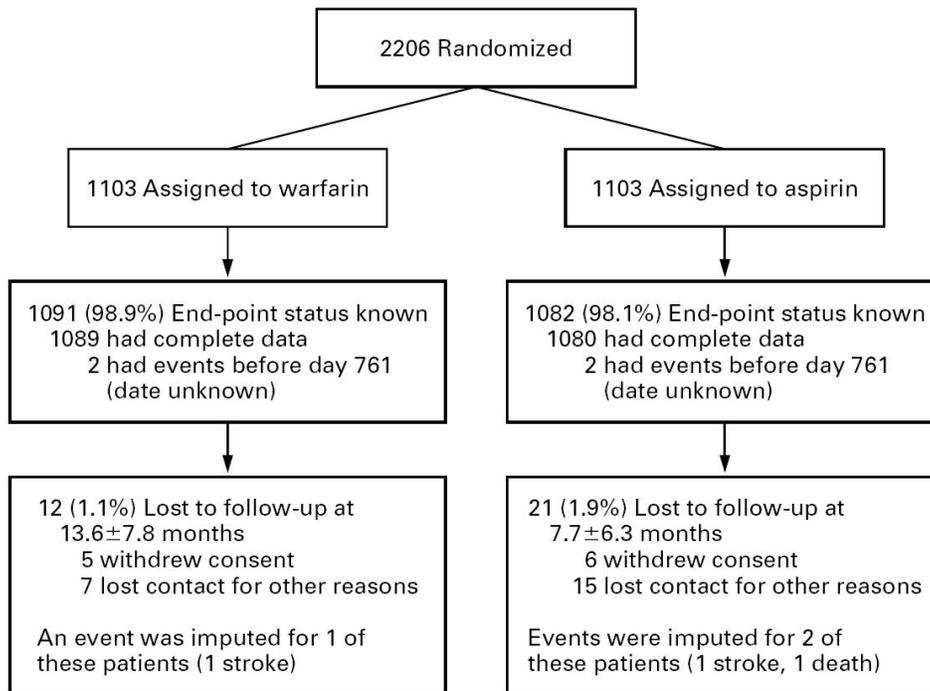
Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). Lancet 1994; 334: 1383-1389.

	Simvastatina	Placebo	RR	RRR	RAR	NNT
	Incidencia en tratados con Simvastatina $I_S$	Incidencia en tratados con placebo $I_P$	$RR = \frac{I_S}{I_A}$	$[ 1 - RR ] * 100$	$I_S - I_P$	$[ 1 / RAR ] * 100$
Incidentes cardiovasculares mayores	19%	28%	0,68	32%	9%	11
Mortalidad	8%	12%	0,66	33%	4%	25

Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandivarian Simvastatin Survival Study (4S). Lancet 1994; 334: 1383-1389.



Mohr JP, Thompson JLP, Lazar RM, Levin B, Sacco RL, Furie KL, et al. A comparison of warfarin and aspirin for the prevention of recurrent ischemic stroke. N Engl J Med 2001;345:1444-51.



**Figure 1.** Follow-up of Patients and Imputation of Events.

Events for which exact dates were unknown were considered to have occurred at the midpoints of the calendar periods during which they occurred. Plus-minus values are means  $\pm$  SE.

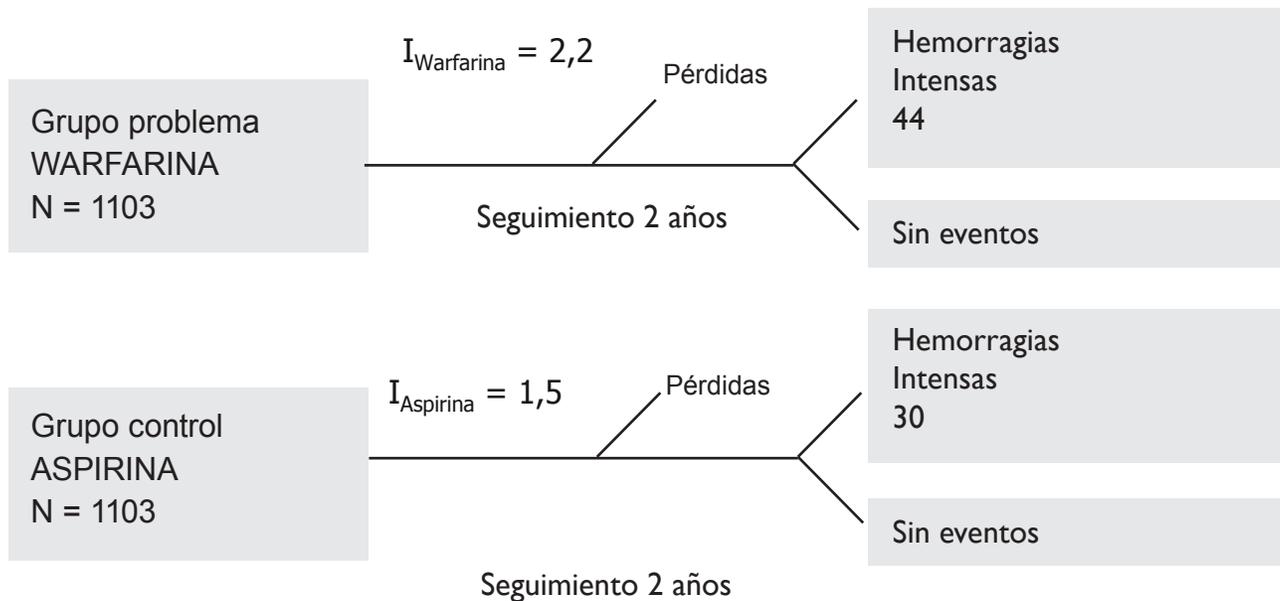
$$\text{Riesgo Relativo (RR)} = \frac{\text{Tasa de Incidencia de Recurrencia de accidentes vasculares cerebrales de origen isquémico o muerte con WARFARINA}}{\text{Tasa de Incidencia de Recurrencia de accidentes vasculares cerebrales de origen isquémico o muerte con ASPIRINA}} \quad \text{RR} = \frac{I_W}{I_A}$$

$$\text{RR} = \frac{17,8}{16,0} = 1,13$$

$$\text{RR} = 1,13 ; \text{IC95\%} = 0,92 - 1,38 \quad p = 0,25$$

$$\text{RRR} = [ 1 - \text{RR} ] * 100 \quad \text{RRR} = - 13\%$$

Mohr JP, Thompson JLP, Lazar RM, Levin B, Sacco RL, Furie KL, et al. A comparison of warfarin and aspirin for the prevention of recurrent ischemic stroke. N Engl J Med 2001;345:1444-51.



Mohr JP, Thompson JLP, Lazar RM, Levin B, Sacco RL, Furie KL, et al. A comparison of warfarin and aspirin for the prevention of recurrent ischemic stroke. *N Engl J Med* 2001;345:1444-51.

$$\text{Riesgo Relativo (RR)} = \frac{\text{Tasa de Incidencia de HEMORRAGIAS intensas con WARFARINA}}{\text{Tasa de Incidencia de HEMORRAGIAS intensas con ASPIRINA}} \quad \text{RR} = \frac{2,2}{1,5}$$

$$\text{RR} = \frac{I_P}{I_A} = 1,48$$

$$\text{RR} = 1,48; \text{ IC95\%} = 0,93-2,44 \quad p = 0,10$$

$$\text{RRR} = [ 1 - \text{RR} ] * 100$$

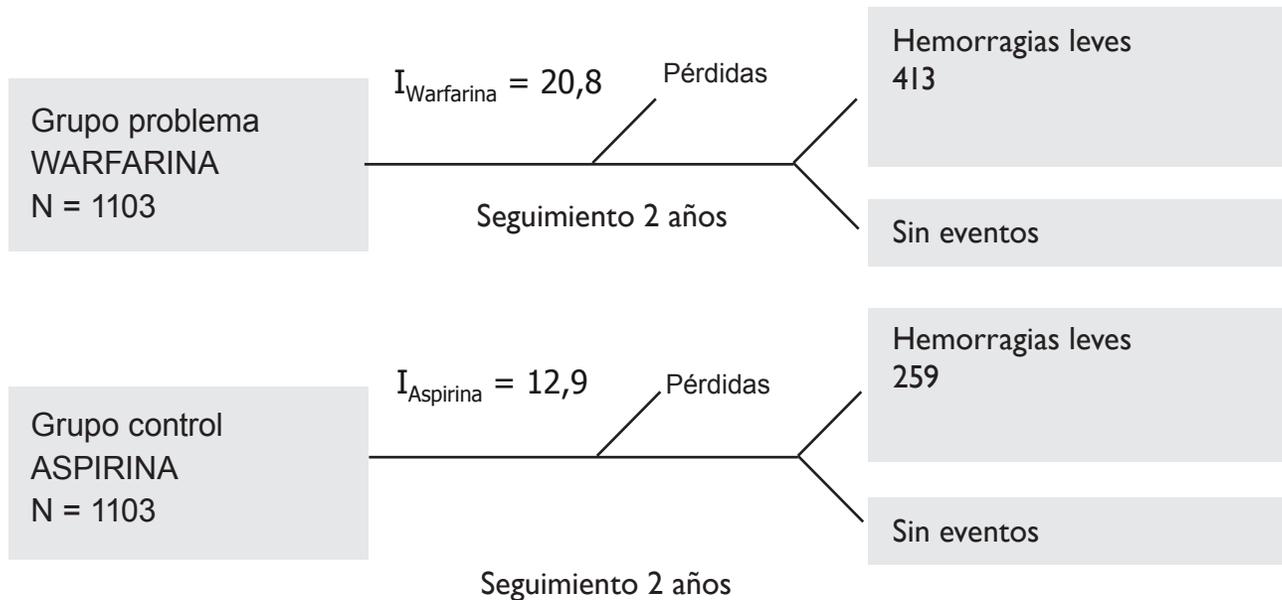
$$\text{RRR} = - 48\%$$

Mohr JP, Thompson JLP, Lazar RM, Levin B, Sacco RL, Furie KL, et al. A comparison of warfarin and aspirin for the prevention of recurrent ischemic stroke. N Engl J Med 2001;345:1444-51.

	Warfarina	Aspirina	RR	RRR	RAR
	Incidencia en tratados con WARFARINA $I_W$	Incidencia en tratados con ASPIRINA $I_A$	$RR = \frac{I_S}{I_A}$	$[1 - RR] * 100$	$I_W - I_A$
Incidencia HEMORRAGIAS intensas	2,2	1,5	1,48	- 48%	0,7

Además, tal como comentaban los autores en la discusión del artículo, la warfarina tiene un coste mayor y requiere una monitorización más estrecha. Con estos datos, ¿se puede considerar que la warfarina y la aspirina son alternativas terapéuticas razonables para prevenir los accidentes cerebrovasculares isquémicos recurrentes?

Mohr JP, Thompson JLP, Lazar RM, Levin B, Sacco RL, Furie KL, et al.  
A comparison of warfarin and aspirin for the prevention of recurrent ischemic stroke.  
N Engl J Med 2001;345:1444-51.



Mohr JP, Thompson JLP, Lazar RM, Levin B, Sacco RL, Furie KL, et al. A comparison of warfarin and aspirin for the prevention of recurrent ischemic stroke. *N Engl J Med* 2001;345:1444-51.

$$\text{Riesgo Relativo (RR)} = \frac{\text{Tasa de Incidencia de HEMORRAGIAS leves con WARFARINA}}{\text{Tasa de Incidencia de HEMORRAGIAS leves con ASPIRINA}} \quad \text{RR} = \frac{20,8}{12,9}$$

$$\text{RR} = \frac{I_P}{I_A} = 1,61$$

$$\text{RR} = 1,61; \text{ IC95\%} = 1,38-1,89$$

$$p < 0,001$$

$$\text{RRR} = [ 1 - \text{RR} ] * 100$$

$$\text{RRR} = - 61\%$$

Mohr JP, Thompson JLP, Lazar RM, Levin B, Sacco RL, Furie KL, et al. A comparison of warfarin and aspirin for the prevention of recurrent ischemic stroke. N Engl J Med 2001;345:1444-51.

	Warfarina	Aspirina	RR	RRR	RAR
	Incidencia en tratados con WARFARINA $I_W$	Incidencia en tratados con ASPIRINA $I_A$	$RR = \frac{I_S}{I_A}$	$[1 - RR] * 100$	$I_W - I_A$
Incidencia HEMORRAGIAS leves	20,8	12,9	1,61	- 61%	7,9

Además, tal como comentaban los autores en la discusión del artículo, la warfarina tiene un coste mayor y requiere una monitorización más estrecha. Con estos datos, ¿se puede considerar que la warfarina y la aspirina son alternativas terapéuticas razonables para prevenir los accidentes cerebrovasculares isquémicos recurrentes?

Mohr JP, Thompson JLP, Lazar RM, Levin B, Sacco RL, Furie KL, et al.  
A comparison of warfarin and aspirin for the prevention of recurrent ischemic stroke.  
N Engl J Med 2001;345:1444-51.

The New England Journal of Medicine

---

## A COMPARISON OF WARFARIN AND ASPIRIN FOR THE PREVENTION OF RECURRENT ISCHEMIC STROKE

J.P. MOHR, M.D., J.L.P. THOMPSON, PH.D., R.M. LAZAR, PH.D., B. LEVIN, M.D., R.L. SACCO, M.D., K.L. FURIE, M.D., J.P. KISTLER, M.D., G.W. ALBERS, M.D., L.C. PETTIGREW, M.D., H.P. ADAMS, JR., M.D., C.M. JACKSON, M.D., AND P. PULLICINO, M.D., FOR THE WARFARIN-ASPIRIN RECURRENT STROKE STUDY GROUP\*

### ABSTRACT

*Background* Despite the use of antiplatelet agents, usually aspirin, in patients who have had an ischemic stroke, there is still a substantial rate of recurrence. Therefore, we investigated whether warfarin, which is effective and superior to aspirin in the prevention of cardiogenic embolism, would also prove superior in the prevention of recurrent ischemic stroke in patients with a prior noncardioembolic ischemic stroke.



The New England Journal of Medicine

---

**A COMPARISON OF WARFARIN AND ASPIRIN FOR THE PREVENTION  
OF RECURRENT ISCHEMIC STROKE**

J.P. MOHR, M.D., J.L.P. THOMPSON, PH.D., R.M. LAZAR, PH.D., B. LEVIN, M.D., R.L. SACCO, M.D., K.L. FURIE, M.D.,  
J.P. KISTLER, M.D., G.W. ALBERS, M.D., L.C. PETTIGREW, M.D., H.P. ADAMS, JR., M.D., C.M. JACKSON, M.D.,  
AND P. PULLICINO, M.D., FOR THE WARFARIN-ASPIRIN RECURRENT STROKE STUDY GROUP\*

*Methods* In a multicenter, double-blind, randomized trial, we compared the effect of warfarin (at a dose adjusted to produce an international normalized ratio of 1.4 to 2.8) and that of aspirin (325 mg per day) on the combined primary end point of recurrent ischemic stroke or death from any cause within two years.



The New England Journal of Medicine

---

## A COMPARISON OF WARFARIN AND ASPIRIN FOR THE PREVENTION OF RECURRENT ISCHEMIC STROKE

J.P. MOHR, M.D., J.L.P. THOMPSON, PH.D., R.M. LAZAR, PH.D., B. LEVIN, M.D., R.L. SACCO, M.D., K.L. FURIE, M.D., J.P. KISTLER, M.D., G.W. ALBERS, M.D., L.C. PETTIGREW, M.D., H.P. ADAMS, JR., M.D., C.M. JACKSON, M.D., AND P. PULLICINO, M.D., FOR THE WARFARIN-ASPIRIN RECURRENT STROKE STUDY GROUP\*

*Results* The two randomized study groups were similar with respect to base-line risk factors. In the intention-to-treat analysis, no significant differences were found between the treatment groups in any of the outcomes measured. The primary end point of death or recurrent ischemic stroke was reached by 196 of 1103 patients assigned to warfarin (17.8 percent) and 176 of 1103 assigned to aspirin (16.0 percent;  $P=0.25$ ; hazard ratio comparing warfarin with aspirin, 1.13; 95 percent confidence interval, 0.92 to 1.38). The rates of major hemorrhage were low (2.22 per 100 patient-years in the warfarin group and 1.49 per 100 patient-years in the aspirin group). Also, there were no significant treatment-related differences in the frequency of or time to the primary end point or major hemorrhage according to the cause of the initial stroke.



The New England Journal of Medicine

---

**A COMPARISON OF WARFARIN AND ASPIRIN FOR THE PREVENTION  
OF RECURRENT ISCHEMIC STROKE**

J.P. MOHR, M.D., J.L.P. THOMPSON, PH.D., R.M. LAZAR, PH.D., B. LEVIN, M.D., R.L. SACCO, M.D., K.L. FURIE, M.D.,  
J.P. KISTLER, M.D., G.W. ALBERS, M.D., L.C. PETTIGREW, M.D., H.P. ADAMS, JR., M.D., C.M. JACKSON, M.D.,  
AND P. PULLICINO, M.D., FOR THE WARFARIN-ASPIRIN RECURRENT STROKE STUDY GROUP\*

***Conclusions*** Over a two-year period, we found no difference between aspirin and warfarin in the prevention of recurrent ischemic stroke or death or in the rate of major hemorrhage. Consequently, we regard both warfarin and aspirin as reasonable therapeutic alternatives. (N Engl J Med 2001;345:1444-51.)

## ABSTRACT

**Background** Despite the use of antiplatelet agents, usually aspirin, in patients who have had an ischemic stroke, there is still a substantial rate of recurrence. Therefore, we investigated whether warfarin, which is effective and superior to aspirin in the prevention of cardiogenic embolism, would also prove superior in the prevention of recurrent ischemic stroke in patients with a prior noncardioembolic ischemic stroke.

**Methods** In a multicenter, double-blind, randomized trial, we compared the effect of warfarin (at a dose adjusted to produce an international normalized ratio of 1.4 to 2.8) and that of aspirin (325 mg per day) on the combined primary end point of recurrent ischemic stroke or death from any cause within two years.

**Results** The two randomized study groups were similar with respect to base-line risk factors. In the intention-to-treat analysis, no significant differences were found between the treatment groups in any of the outcomes measured. The primary end point of death or recurrent ischemic stroke was reached by 196 of 1103 patients assigned to warfarin (17.8 percent) and 176 of 1103 assigned to aspirin (16.0 percent;  $P=0.25$ ; hazard ratio comparing warfarin with aspirin, 1.13; 95 percent confidence interval, 0.92 to 1.38). The rates of major hemorrhage were low (2.22 per 100 patient-years in the warfarin group and 1.49 per 100 patient-years in the aspirin group). Also, there were no significant treatment-related differences in the frequency of or time to the primary end point or major hemorrhage according to the cause of the initial stroke.

**Conclusions** Over a two-year period, we found no difference between aspirin and warfarin in the prevention of recurrent ischemic stroke or death or in the rate of major hemorrhage. Consequently, we regard both warfarin and aspirin as reasonable therapeutic alternatives. (N Engl J Med 2001;345:1444-51.)

**TABLE 2. RESULTS OF PRIMARY AND SECONDARY ANALYSES.**

ANALYSIS	EVENTS		PROBABILITY OF EVENT AT 2 YR*		HAZARD RATIO (95% CI)†	P VALUE‡
	WARFARIN	ASPIRIN	WARFARIN	ASPIRIN		
	no. with events/total no.					
<b>Primary and secondary analyses</b>						
Recurrent ischemic stroke or death	196/1103	176/1103	17.8	16.0	1.13 (0.92–1.38)	0.25
Recurrent ischemic stroke or death or major hemorrhage	222/1103	196/1103	20.0	17.8	1.15 (0.95–1.39)	0.16
Recurrent ischemic stroke or death, with data from patients lost to follow-up censored	195/1103	174/1103	17.6	15.9	1.13 (0.92–1.39)	0.24
Recurrent ischemic stroke or death (model including interaction of treatment assignment and interruption of treatment)	196/1103	176/1103				
<b>Subgroup analyses for primary end point</b>						
Sex						
Male	122/656	101/653	18.5	15.4	1.23 (0.95–1.61)	0.12
Female	74/447	75/450	16.2	16.8	0.98 (0.71–1.36)	0.92
Race or ethnic group						
Black	70/338	59/325	20.2	18.4	1.14 (0.81–1.62)	0.45
White	98/627	90/626	15.5	14.3	1.10 (0.83–1.47)	0.50
Hispanic	21/105	21/118	20.1	17.9	1.14 (0.62–2.09)	0.66
Other	7/33	6/34	21.2	17.6	1.18 (0.40–3.50)	0.77
Cause of prior stroke						
Cryptogenic	42/281	48/295	15.0	16.5	0.92 (0.61–1.39)	0.68
Small vessel or lacunar	107/612	95/625	17.1	15.2	1.15 (0.88–1.52)	0.31
Large artery, severe stenosis, or occlusion	27/144	18/115	18.8	15.7	1.22 (0.67–2.22)	0.51
Other determined cause	11/30	7/33	36.7	21.2	1.99 (0.77–5.15)	0.15
Conflicting mechanism	9/36	8/35	25.0	23.0	1.14 (0.44–2.96)	0.79

\*Probabilities of events were derived from Kaplan–Meier curves.

†Hazard ratios were calculated by the discrete-time Cox model. CI denotes confidence interval.

‡P values were calculated with the log-rank test, except for those for the interruption-of-therapy model, which were calculated by the Wald test.

The New England Journal of Medicine

**A COMPARISON OF WARFARIN AND ASPIRIN FOR THE PREVENTION  
OF RECURRENT ISCHEMIC STROKE**

J.P. MOHR, M.D., J.L.P. THOMPSON, PH.D., R.M. LAZAR, PH.D., B. LEVIN, M.D., R.L. SACCO, M.D., K.L. FURIE, M.D.,  
J.P. KISTLER, M.D., G.W. ALBERS, M.D., L.C. PETTIGREW, M.D., H.P. ADAMS, JR., M.D., C.M. JACKSON, M.D.,  
AND P. PULLICINO, M.D., FOR THE WARFARIN-ASPIRIN RECURRENT STROKE STUDY GROUP\*

**TABLE 2.** RESULTS OF PRIMARY AND SECONDARY ANALYSES.

ANALYSIS	EVENTS		PROBABILITY OF EVENT AT 2 YR*		HAZARD RATIO (95% CI)†	P VALUE‡
	WARFARIN	ASPIRIN	WARFARIN	ASPIRIN		
	no. with events/total no.					
<b>Primary and secondary analyses</b>						
Recurrent ischemic stroke or death	196/1103	176/1103	17.8	16.0	1.13 (0.92–1.38)	0.25
Recurrent ischemic stroke or death or major hemorrhage	222/1103	196/1103	20.0	17.8	1.15 (0.95–1.39)	0.16
Recurrent ischemic stroke or death, with data from patients lost to follow-up censored	195/1103	174/1103	17.6	15.9	1.13 (0.92–1.39)	0.24
Recurrent ischemic stroke or death (model including interaction of treatment as- signment and interruption of treatment)	196/1103	176/1103				

**TABLE 3. ADVERSE EVENTS ACCORDING TO TREATMENT ASSIGNMENT.\***

EVENT	WARFARIN (N=1103)	ASPIRIN (N=1103)	ODDS RATIO (95% CI)	P VALUE†
	no. (%)			
Death	47 (4.3)	53 (4.8)	0.88 (0.58–1.32)	0.61
Related to hemorrhage	7 (0.6)	5 (0.4)	1.40 (0.42–5.13)	0.77
First hemorrhage‡				
Major	38 (3.4)	30 (2.7)	1.28 (0.78–2.10)	0.39
Minor	261 (23.7)	188 (17.0)	1.51 (1.22–1.87)	<0.001
			RATE RATIO (95% CI)	P VALUES
	no. of events (rate/100 patient-yr)			
All hemorrhages¶				
Major	44 (2.2)	30 (1.5)	1.48 (0.93–2.44)	0.10
Minor	413 (20.8)	259 (12.9)	1.61 (1.38–1.89)	<0.001

\*Maximal follow-up was 25 months. Hemorrhages occurring on the day of the primary event (death or recurrent ischemic stroke) are included. CI denotes confidence interval.

†P values were calculated by the exact test of two independent proportions.

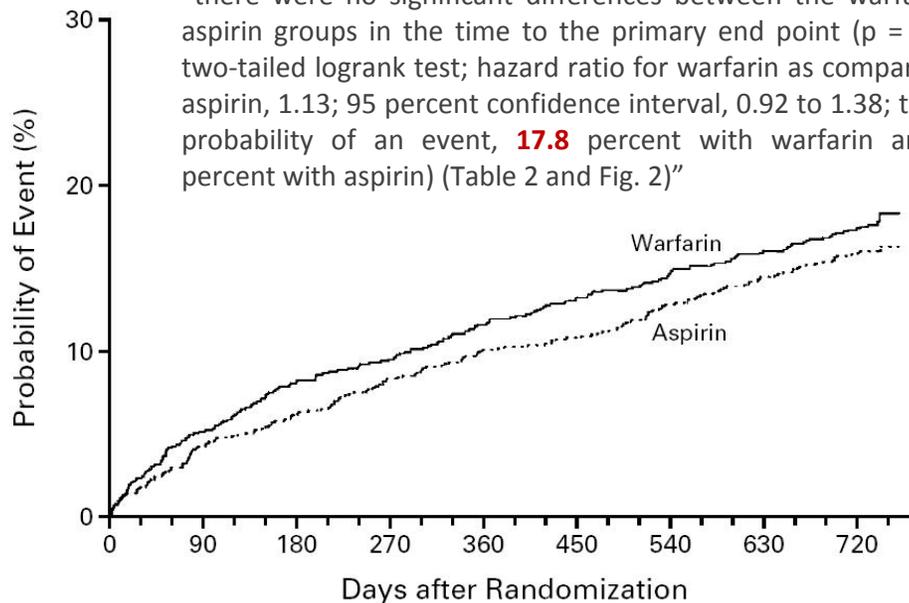
‡The first hemorrhage is the first or only hemorrhage for each patient.

§P values were calculated by the exact conditional binomial test for two independent Poisson processes.

¶All hemorrhages include all hemorrhages in any patient.



“there were no significant differences between the warfarin and aspirin groups in the time to the primary end point ( $p = 0,25$  by two-tailed logrank test; hazard ratio for warfarin as compared with aspirin, 1.13; 95 percent confidence interval, 0.92 to 1.38; two-year probability of an event, **17.8** percent with warfarin and **16.0** percent with aspirin) (Table 2 and Fig. 2)”



No. AT RISK

Warfarin	1103	1047	1013	998	972	956	939	924	885
Aspirin	1103	1057	1032	1004	984	974	951	932	900

**Figure 2.** Kaplan–Meier Analyses of the Time to Recurrent Ischemic Stroke or Death According to Treatment Assignment.

## RESULTS (fragmento del texto)

“The rates of major hemorrhage were low, with no significant differences between treatment groups; the annual rates were 2.22 per 100 patient-years for warfarin and 1.49 per 100 patient-years for aspirin (rate ratio, 1.48;  $p = 0,10$ ). Patients in the warfarin group had significantly more minor hemorrhages than did those in the aspirin group (Table 3). There was no significant difference between groups in the time to the first occurrence of major hemorrhage or the primary end point ( $p = 0,16$ ; hazard ratio with warfarin as compared with aspirin, 1.15; 95 percent confidence interval, 0.95 to 1.39) (Table 2).”

## DISCUSSION (fragmento del texto)

“We observed no significant difference between treatment with warfarin and treatment with aspirin in the prevention of recurrent ischemic stroke or death or in the occurrence of serious adverse events in this large cohort of patients with inferred noncardioembolic ischemic stroke”.

“Warfarin offered no additional benefit over aspirin in preventing recurrent ischemic stroke in the population we studied. Patients with other, established reasons for warfarin use may take comfort in the evidence of safety and lack of significant difference overall, as compared with aspirin. However, aspirin, either alone or in combination with some other antiplatelet agents, 38 appears to be a well-justified choice for the prevention of recurrent ischemic stroke.”

## Mito I: “La sonrisa de la esfinge”

¿Por qué concluyes que “la ausencia de evidencia es evidencia de ausencia”?

“La ausencia de evidencia NO es evidencia de ausencia”.

Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995;311:485.

Altman y Bland alertaron sobre un error común en la bibliografía biomédica: interpretar un resultado estadísticamente no significativo, cuando se comparan dos tratamientos, como sinónimo de su equivalencia terapéutica.

## Mito I: “La sonrisa de la esfinge”

En este estudio se comparaba la eficacia de la warfarina y de la aspirina para prevenir las recurrencias de los accidentes vasculares cerebrales de origen isquémico. Mohr JP, et al diseñaron el estudio para poder demostrar una reducción relativa del riesgo (RRR) del 30% en el grupo tratado con warfarina. Cuando finalizó el período de dos años de seguimiento, no se observaron diferencias estadísticamente significativas entre los dos grupos, y los autores llegaron a la conclusión de que la warfarina y la aspirina son opciones terapéuticas razonables en este tipo de situaciones clínicas. Sin embargo, la warfarina no sólo NO consiguió reducir el riesgo en un 30%, sino que lo INCREMENTÓ en un 13%, y el riesgo de HEMORRAGIAS importantes fue SUPERIOR en el grupo de la WARFARINA (2,22 por 100 pacientes/año) frente a la ASPIRINA (1,49 por 100 pacientes/año en el grupo aspirina) pero sin alcanzar significación. Con estos datos....

¿Crees que se puede considerar que la warfarina y la aspirina son alternativas terapéuticas razonables para prevenir los accidentes cerebrovasculares isquémicos recurrentes?  
NO.

Además, tal como comentaban los autores en la discusión del artículo, la warfarina tiene un coste mayor y requiere una monitorización más estrecha.

# The NEW ENGLAND JOURNAL of MEDICINE

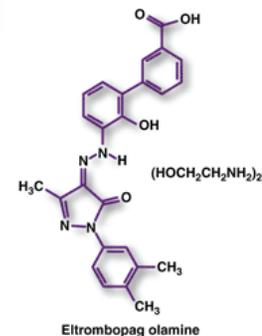
ESTABLISHED IN 1812

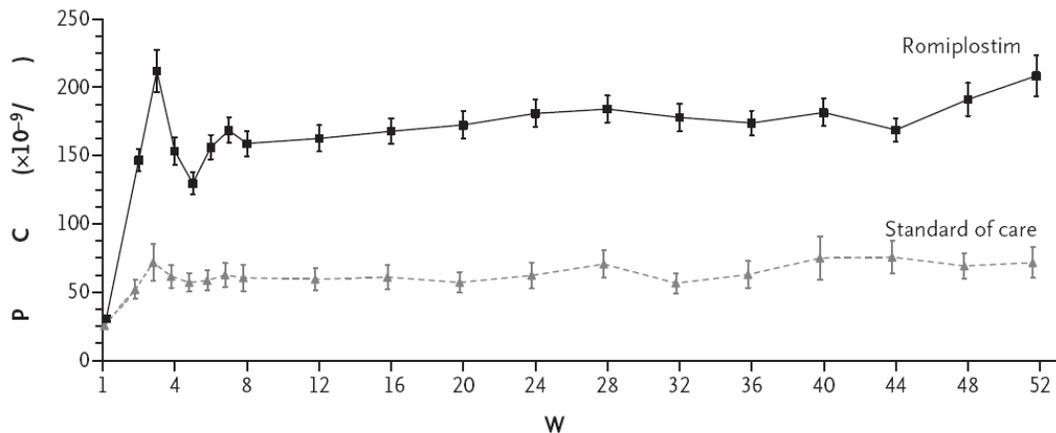
NOVEMBER 11, 2010

VOL. 363 NO. 20

## Romiplostim or Standard of Care in Patients with Immune Thrombocytopenia

David J. Kuter, M.D., D.Phil., Mathias Rummel, M.D., Ralph Boccia, M.D., B. Gail Macik, M.D.,  
Ingrid Pabinger, M.D., Dominik Selleslag, M.D., Francesco Rodeghiero, M.D., Beng H. Chong, M.D.,  
Xuena Wang, Ph.D., and Dietmar P. Berger, M.D., Ph.D.\*





N . D	72	68	62	59	57	54	53	54	51	51	49	51	46	38
Standard of care	72	68	62	59	57	54	53	54	51	51	49	51	46	38
Romiplostim	153	150	148	148	141	132	137	135	132	135	126	127	130	122

El Romiplostim produjo un aumento estadísticamente significativo del recuento de plaquetas durante el periodo de estudio, pero no se asoció a una disminución del las hemorragias de grado  $\frac{3}{4}$ .

## Efficacy of romiplostim in patients with chronic immune thrombocytopenic purpura: a double-blind randomised controlled trial

David J Kuter, James B Bussel, Roger M Lyons, Vinod Pullarkat, Terry B Gernsheimer, Francis M Senecal, Louis M Aledort, James N George, Craig M Kessler, Miguel A Sanz, Howard A Liebman, Frank T Slovic, J Th M deWaf, Emmanuelle Bourgeois, Troy H Guthrie Jr, Adrian Newland, Jeffrey S Wasser, Solomon I Hamburg, Carlos Grande, François Lefrère, Alan Eli Lichtin, Michael D Tarantino, Howard R Terrelba, Jean-François Viallard, Francis J Cuevas, Ronald S Go, David H Henry, Robert L Redner, Lawrence Rice, Martin R Schipperus, D Matthew Guo, Janet L Nichol

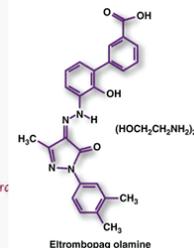
### Summary

**Background** Chronic immune thrombocytopenic purpura (ITP) is characterised by accelerated platelet destruction and decreased platelet production. Short-term administration of the thrombopoiesis-stimulating protein, romiplostim, has been shown to increase platelet counts in most patients with chronic ITP. We assessed the long-term administration of romiplostim in splenectomised and non-splenectomised patients with ITP.

**Methods** In two parallel trials, 63 splenectomised and 62 non-splenectomised patients with ITP and a mean of three platelet counts  $30 \times 10^9/L$  or less were randomly assigned 2:1 to subcutaneous injections of romiplostim (n=42 in splenectomised study and n=41 in non-splenectomised study) or placebo (n=21 in both studies) every week for 24 weeks. Doses of study drug were adjusted to maintain platelet counts of  $50 \times 10^9/L$  to  $200 \times 10^9/L$ . The primary objectives were to assess the efficacy of romiplostim as measured by a durable platelet response (platelet count  $\geq 50 \times 10^9/L$  during 6 or more of the last 8 weeks of treatment) and treatment safety. Analysis was per protocol. These studies are registered with ClinicalTrials.gov, numbers NCT00102323 and NCT00102336.

**Findings** A durable platelet response was achieved by 16 of 42 splenectomised patients given romiplostim versus none of 21 given placebo (difference in proportion of patients responding 38% [95% CI 23.4–52.8],  $p=0.0013$ ), and by 25 of 41 non-splenectomised patients given romiplostim versus one of 21 given placebo (56% [38.7–73.7],  $p<0.0001$ ). The overall platelet response rate (either durable or transient platelet response) was noted in 88% (36/41) of non-splenectomised and 79% (33/42) of splenectomised patients given romiplostim compared with 14% (three of 21) of non-splenectomised and no splenectomised patients given placebo ( $p<0.0001$ ). Patients given romiplostim achieved platelet counts of  $50 \times 10^9/L$  or more on a mean of 13.8 (SE 0.9) weeks (mean 12.3 [1.2] weeks in splenectomised group vs 15.2 [1.2] weeks in non-splenectomised group) compared with 0.8 (0.4) weeks for those given placebo (0.2 [0.1] weeks vs 1.3 [0.8] weeks). 87% (20/23) of patients given romiplostim (12/12 splenectomised and eight of 11 non-splenectomised patients) reduced or discontinued concurrent therapy compared with 38% (six of 16) of those given placebo (one of six splenectomised and five of ten non-splenectomised patients). Adverse events were much the same in patients given romiplostim and placebo. No antibodies against romiplostim or thrombopoietin were detected.

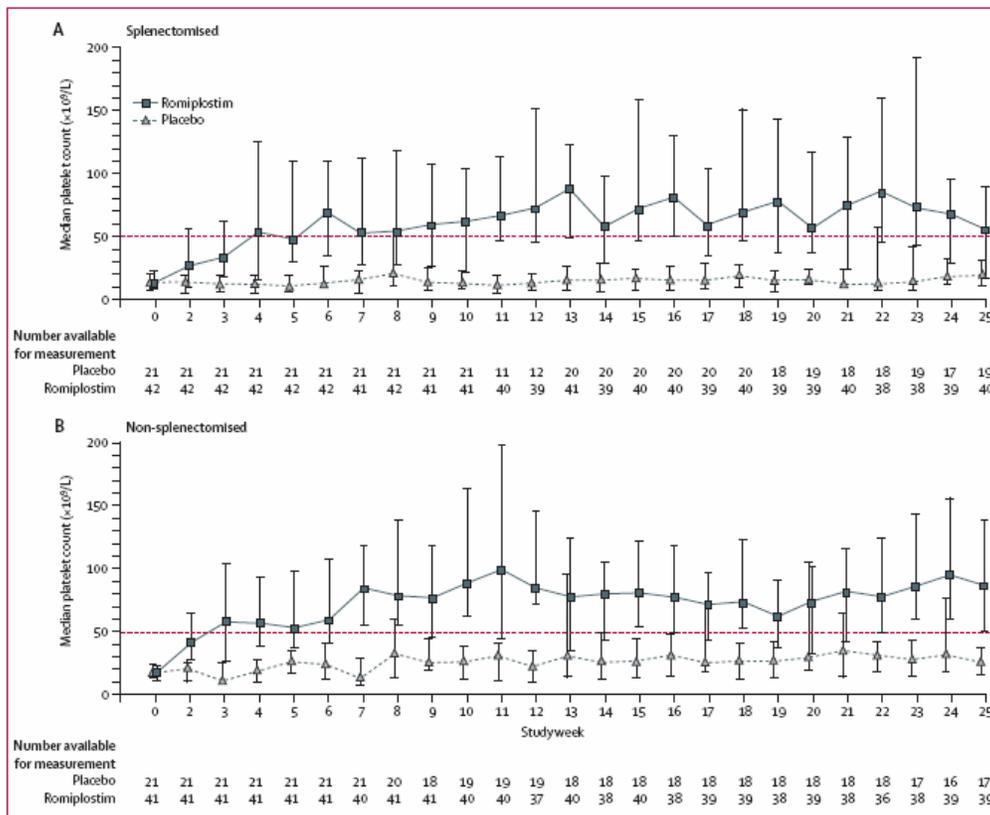
**Interpretation** Romiplostim was well tolerated, and increased and maintained platelet counts in splenectomised and non-splenectomised patients with ITP. Many patients were able to reduce or discontinue other ITP medications. Stimulation of platelet production by romiplostim may provide a new therapeutic option for patients with ITP.



**Lancet** 2008; **371**: 395–403

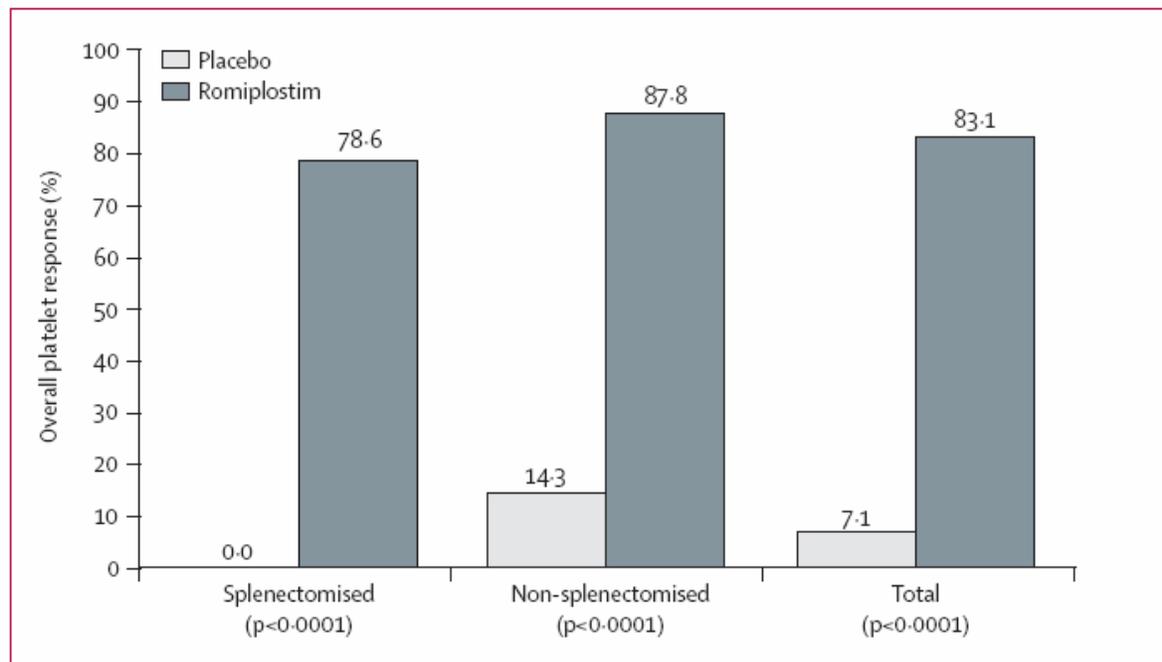
See Comment page 362

Massachusetts General Hospital, Boston, MA, USA (D J Kuter MD); New York Presbyterian Hospital, Division of Pediatrics, New York, NY, USA (J B Bussel MD); Hematology Oncology Associates of South Texas, San Antonio, TX, USA (R M Lyons MD); City of Hope, Division of Hematology, Duarte, CA, USA (V Pullarkat MD); Puget Sound Blood Center, Seattle, WA, USA (T B Gernsheimer MD); Northwest Medical Specialties, Tacoma, WA, USA (F M Senecal MD); Mount Sinai Hospital, New York, NY (L M Aledort MD); University of Oklahoma Health Sciences Center, Hematology Oncology Section, Oklahoma City, OK, USA (J N George MD); Georgetown University Medical Center, Washington, DC, USA (C M Kessler MD); Hospital La Fe, Valencia, Spain (M A Sanz MD); USC Keck School of Medicine, Division of Hematology, Los Angeles, CA, USA (H A Liebman MD); Heartland Hematology Oncology Associates, Inc., Kansas City, MO, USA (F T Slovic MD); Indiana University Medical Centre



**Figure 3:** Median platelet count at every weekly study visit for splenectomised (A) and non-splenectomised (B) patients

Data includes all patients, even those who received rescue drugs. Error bars indicate the range from the first to third quartiles. Dashed line indicates platelet count of  $50 \times 10^9/L$ .



**Figure 4: Patients with overall platelet response**

Overall platelet response defined as durable platelet response or four or more weekly platelet responses at any time during the study.

## Romiplostim. PTI

- Romiplostim. Ensayo pivotal. Fase III vs. placebo.
- Inclusión: < 30.000 plaquetas/mm<sup>3</sup>.
- Objetivo principal: > 50.000 plaquetas/mm<sup>3</sup>.



Efectos adversos:



Objetivo subrogado

	Romiplostim	Placebo	p
Hemorragia Grado 3/4	6 de 84 (7%)	5 de 41 (12%)	n.s.

David J Kuter, James B Bussel. Efficacy of romiplostim in patients with chronic immune thrombocytopenic purpura: a double-blind randomised controlled trial. Lancet 2008; 371: 395–403.

## Romiplostim. PTI

Hemorragia Grado 3/4	Romiplostim	Placebo	Total
SI	6 (7%)	5 (12%)	11
NO	78	36	114
Total	84	41	125

El Romiplostim produjo un aumento estadísticamente significativo del recuento de plaquetas durante el periodo de estudio, pero no se asoció a una disminución de las hemorragias de grado  $\frac{3}{4}$ .

$$\chi^2_{\text{Pearson}} = 0,88; p = 0,349$$

David J Kuter, James B Bussel. Efficacy of romiplostim in patients with chronic immune thrombocytopenic purpura: a double-blind randomised controlled trial. Lancet 2008; 371: 395–403.

## Odds Ratio (OR) versus Riesgo Relativo (RR)

Exposición (F)	Enfermedad (E)		Total
	Cáncer de páncreas (+)	No enfermos (-)	
Tabaco (+)	A (10)	B (9990)	10000
Tabaco (-)	C (10)	D (19990)	20000
Total	20	29980	30000

$$RR = \frac{[ A / (A + B) ]}{[ C / (C + D) ]}$$

$$RR = \frac{[ 10 / (10000) ]}{[ 10 / (20000) ]}$$

$$OR = \frac{[ A / C ]}{[ B / D ]}$$

$$OR = \frac{[ 10/10 ]}{[ 9990 / 19990 ]}$$

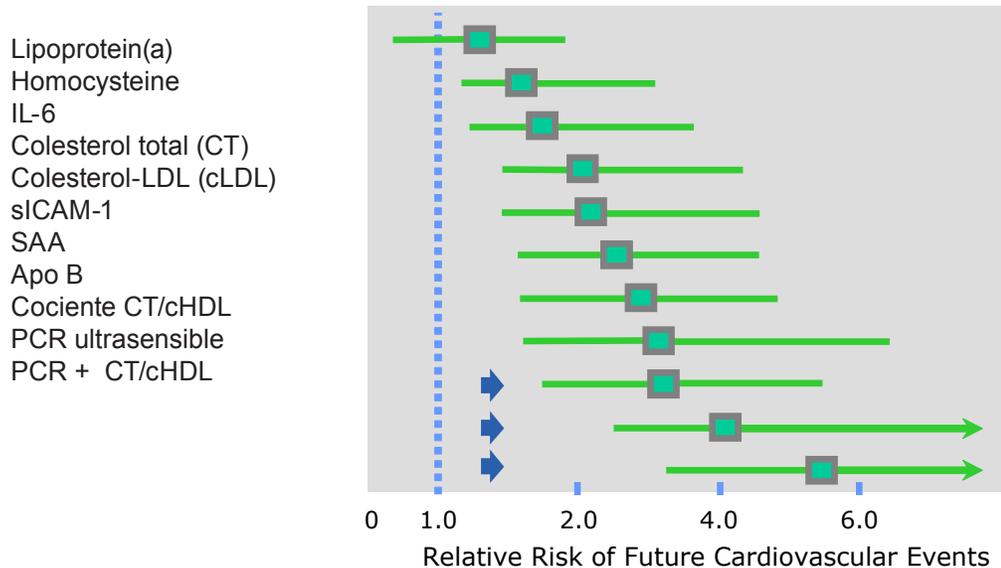
➡ OR = 2,001

➡ RR = 2,000

La OR sólo se aproxima al RR cuando el suceso es raro y ocurre en menos del 10% de los sujetos, en estos casos ambos estimadores son intercambiables.

# Importancia del intervalo de confianza en el RR

## - Risk Factors for Future Cardiovascular Events: WHS -



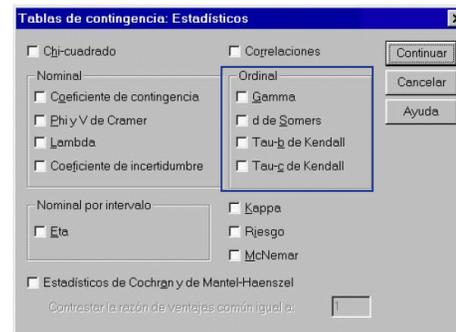
Ridker PM et al. N Engl J Med 2000;342:836-843.

## Datos categóricos:

### Cualitativa versus cualitativa. Ordinales

- *Test de  $\chi^2$  de tendencia lineal*
- Tau-B de Kendall o Stuart
- Tau-C de Kendall
- Coeficiente Gamma Y de Goodman y Kruskal
- D de Somers
- Correlación de Spearman

Estos estadísticos NO se pueden utilizar cuando alguna de las variables sea NOMINAL



## Pregunta de investigación

¿Existe relación entre el nivel cultural y el hábito tabáquico?

Variable PREDICTORA  
CUALITATIVA ordinal

Nivel cultural



Variable RESULTADO  
CUALITATIVA dicotómica

Hábito tabáquico

¿Cuál es el test estadístico más acertado?

## ¿Existe relación entre el nivel socioeconómico y el hábito tabáquico?

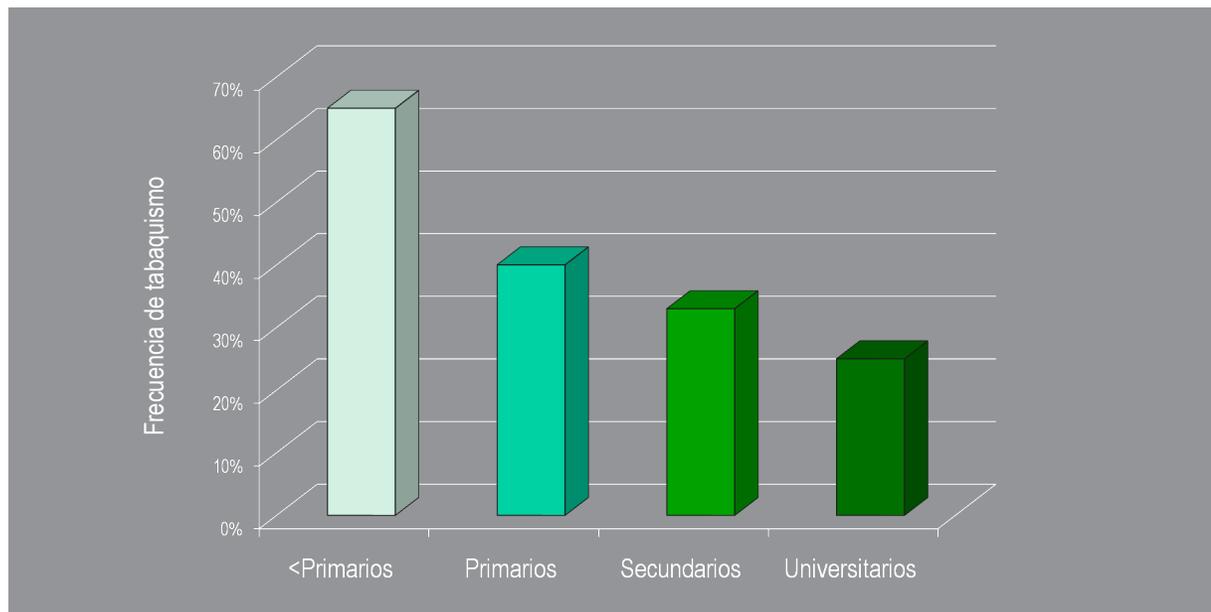
Tabla de contingencia que representa el nivel de estudios alcanzado frente al hábito tabáquico actual

ESTUDIOS	Fumador actual		Total	Porcentaje
	NO	SI		
<Primarios	7	13	20	65%
Primarios	30	20	50	40%
Secundarios	20	10	30	33%
Universitarios	75	25	100	25%
TOTAL	132	68	200	34%

Test de  $\chi^2$  de tendencia lineal

$$\chi^2_{\text{tendencia lineal}} = 11,56 \quad \rightarrow \quad p=0,001$$

## Gráfico que representa el porcentaje de fumadores según su nivel de instrucción



$\chi^2_{\text{tendencia lineal}} = 11,56 \rightarrow p = 0,001$

## Pregunta de investigación

¿Existe relación entre el nivel cultural y la realización de alguna mamografías de cribado en los 3 últimos años?

Variable PREDICTORA  
CUALITATIVA ordinal

Nivel educativo



Variable RESULTADO  
CUALITATIVA dicotómica

Mamografía

¿Cuál es el test estadístico más acertado?

## ¿Existe relación entre el nivel de instrucción y la realización de mamografías?

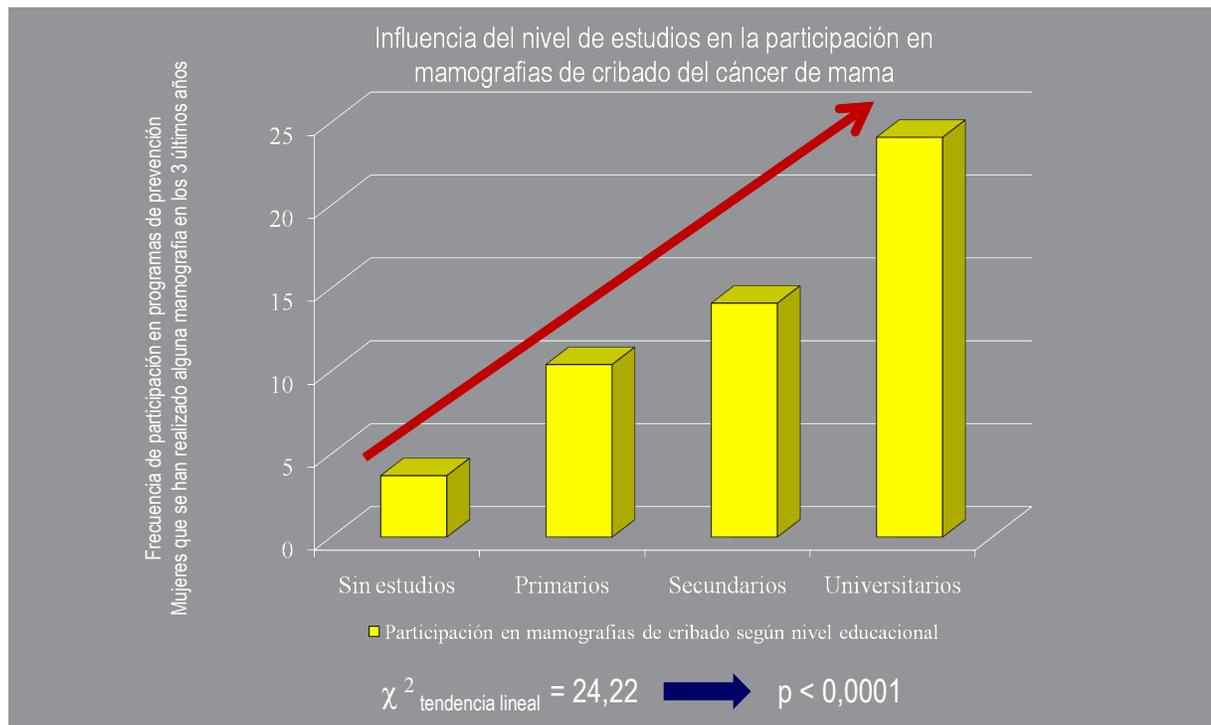
Tabla de contingencia que representa el nivel de estudios alcanzado frente a la realización de alguna mamografía de cribado en los últimos 3 años.

ESTUDIOS	Mamografía de cribado			OR	IC 95%
	SI	NO	%		
Sin estudios	9	233	3,7%	1,00	
Primarios	81	700	10,4%	3,00	1,48 - 6,06
Secundarios	23	140	14,1%	4,25	1,91 – 9.45
Universitarios	13	41	24,1%	8,21	3,30 – 20,44
TOTAL	126	1114			

Test de  $\chi^2$  de tendencia lineal

$$\chi^2_{\text{tendencia lineal}} = 24,22 \quad \rightarrow \quad p < 0,00001$$

## Gráfico que representa el porcentaje de fumadores según su nivel de instrucción



## Resultados (fragmento del texto)

“El conocimiento que las mujeres tienen de la mamografía como medida preventiva del cáncer de mama se encuentra asociado estadísticamente a la mayor utilización de la misma por parte de éstas y se detecta una clara tendencia creciente en la utilización de la mamografía de cribado a medida que aumenta el nivel educativo de las mujeres de la muestra”.

Simó Miñana J, Gaztambide Ganuza M, Morote Sanchís M, et al. Utilización de la mamografía de cribado y sus determinantes demográficos y de riesgo entre mujeres de 25 a 65 años. Med Clin (Barc) 1997; 108: 767-771.

## Discusión (fragmento del texto)

“Coincidente con otros datos publicados es la relación encontrada entre el nivel educativo y la utilización de la mamografía de cribado. Al igual que ocurre con otras prácticas de carácter preventivo, la mamografía de cribado es más utilizada por las mujeres con mayor nivel educativo. No hay que olvidar que este factor está relacionado con el nivel económico, variable que no ha sido recogida en nuestro trabajo y que también ha sido asociada a una mayor utilización de la mamografía de cribado”.

Simó Miñana J, Gaztambide Ganuza M, Morote Sanchís M, et al. Utilización de la mamografía de cribado y sus determinantes demográficos y de riesgo entre mujeres de 25 a 65 años. Med Clin (Barc) 1997; 108: 767-771.

## Pregunta de investigación

¿Existe relación entre el síndrome anémico gestacional y el grupo cultural en Ceuta? ¿Existen diferencias en la prevalencia durante los trimestres de la gestación?

### Variable PREDICTORA

Cualitativa dicotómica

- Cristianas
- Musulmanas



### Variable RESULTADO

Cualitativa ordinal

- Síndrome anémico gestacional
- Primer trimestre
- Segundo trimestre
- Tercer trimestre

Díaz Portillo J et al. «Prevalencia del síndrome anémico gestacional en Ceuta. Estudio de una población birracial». Ciencia Ginecológica 1998, 2 (1): 3-14.

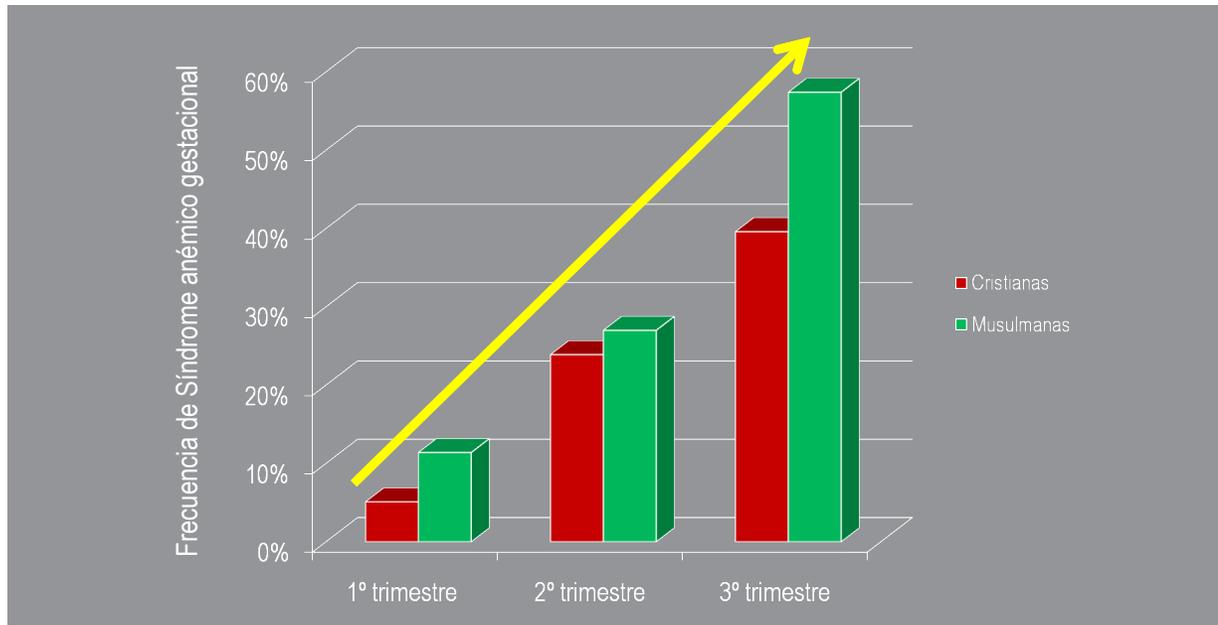
Tabla de contingencia que representa la prevalencia de síndrome anémico gestacional en los dos grupos culturales mayoritarios de Ceuta, y su incidencia en cada trimestre del embarazo.

Trimestre	Mamografía de cribado		Test $\chi^2$	p	Total
	Cristianas	Musulmanas			
1º trimestre	5,1%	11,4%	4,75	p = 0,029	7,9%
2º trimestre	23,9%	27,0%	NS	NS	25,5%
3º trimestre	39,6%	57,4%	19,91	p < 0,0001	49,2%
$\chi^2$ de tendencia lineal	34,3	56,7			71,8
Significación (p)	p < 0,001	p < 0,001			p < 0,001

Test de  $\chi^2$  de tendencia lineal

Test de  $\chi^2$  de *Pearson* de independencia

## Gráfico que representa la prevalencia de síndrome anémico gestacional en los dos grupos culturales mayoritarios de Ceuta



## Pregunta de investigación

¿La hipertrigliceridemia en sangre de cordón umbilical puede ser indicativa de episodios de sufrimiento fetal?

Variable PREDICTORA

SFA

Test de APGAR



Variable RESULTADO

Hipertrigliceridemia

Triglicéridos (sangre de cordón)

## Test de $\chi^2$ de tendencia lineal

Categorización

Variable RESULTADO  
Cuantitativa continua



Variable RESULTADO  
Cualitativa nominal

Variable PREDICTORA  
Cuantitativa discreta



Variable  
PREDICTORA  
Cualitativa ordinal

Tabla 4.- Relación entre la presencia de hipertrigliceridemia neonatal y el índice de apgar al minuto.

TEST DE	Hipertrigliceridemia (n=36)	Normotrigliceridemia (n=636)	OR
<b>APGAR</b>			
<b>Mayor de 7</b>	27 (2,1%)	1253 (97,9%)	1,00
<b>Igual a 7</b>	4 (11,1%)	81 (88,9%)	2,29
<b>Menor de 7</b>	5 (15,6%)	27 (84,4%)	8,57

OR *Odds Ratio*

Test de ji al cuadrado de tendencia ( $\chi^2$ ) = 20,85 ; p<0,0001

Díaz Portillo J. «Influencia del grupo étnico en el patrón lipoproteico de una determinada población pediátrica de Ceuta». Tesis doctoral . Facultad de Medicina. Universidad de Cádiz . Noviembre 1997.

## Resultados

La tabla 4 relaciona la presencia de hipertrigliceridemia neonatal con el índice de Apgar al minuto. Se observó una relación estadísticamente significativa con tendencia lineal entre ambas variables categóricas, con una asociación creciente entre la hipertrigliceridemia y la disminución del índice de Apgar. La probabilidad de detectar hipertrigliceridemia en neonatos con índices de Apgar inferiores a 7 fue ocho veces superior que en RN con Apgar fisiológico.

## Discusión

Nuestros resultados apoyan los obtenidos por Tsang et al. (12) y Andersen y Friis-Hansen (13), e indican una fuerte asociación entre los episodios de sufrimiento fetal y la hipertrigliceridemia neonatal.

Del mismo modo, la frecuencia de hipertrigliceridemia va aumentando conforme disminuye el índice de Apgar.

### Exposición

SFA



### Desenlace o efecto

Hipertrigliceridemia

OR = 7,23 p=0,003

Díaz Portillo J. «Influencia del grupo étnico en el patrón lipoproteico de una determinada población pediátrica de Ceuta». Tesis doctoral . Facultad de Medicina. Universidad de Cádiz . Noviembre 1997.

## Pregunta de investigación

¿La hipertrigliceridemia en sangre de cordón umbilical puede ser debida al consumo de tabaco durante la gestación?

### Variable PREDICTORA

Consumo de tabaco



### Variable RESULTADO

Hipertrigliceridemia  
Triglicéridos (sangre de cordón)

Jacobo Díaz Portillo, María Jesús Sánchez Del Pino, Ana Navarro Arévalo. INFLUENCIA DEL TABAQUISMO EN EL PERFIL LIPOPROTEICO Y GRADO DE PEROXIDACIÓN LIPÍDICA EN SANGRE DE CORDÓN.

Revista de Diagnóstico Biológico 2000; 49: 23-30.

## Test de $\chi^2$ de tendencia lineal



Influencia del tabaco	Hipertrigliceridemia neonatal	Normotrigliceridemia neonatal	OR
No fumadoras	1,7%	98,3%	1,00
Fumadoras moderadas	5,3%	94,7%	3,20
Grandes fumadoras	5,4%	94,6%	3,27

$\chi^2$  tendencia lineal = 5,68 p=0,017

“Utiliza la estadística como un borracho usa una farola, para aguantarse en lugar de para iluminarse”

Andrew Lang

## Pregunta de investigación

¿Existen diferencias en la prevalencia de infección tuberculosa en el colectivo de inmigrantes residentes en el campamento de Calamocarro en Ceuta según la edad?

Variable PREDICTORA  
CUALITATIVA policotómica

Inmigrantes  
< 25 años  
25-35 años  
>35



Variable RESULTADO  
CUALITATIVA dicotómica

Tuberculosis

¿Cuál es el test estadístico más adecuado?

Fernández MT, Díaz J, Sánchez JM, Pérez A, Vadillo J. Prevalencia de infección tuberculosa en la población de inmigrantes en Ceuta, España. Rev Esp Salud Pública 2001; 75: 551-8.

### Análisis del texto en el MATERIAL y MÉTODOS

“Empleamos el test de  $\chi^2$  para comparación de porcentajes, y la prueba de la t de Student para variables cuantitativas. La evaluación de las tendencias se realizó con la prueba de  $\chi^2$  de tendencia lineal”.

### Análisis del texto en los RESULTADOS

“Con independencia del punto de corte seleccionado, se observó mayor positividad en el sexo masculino que en el femenino, y la positividad también presentó una ligera tendencia a aumentar con la edad ( $p = 0,07$ )”.

### Análisis del texto en la DISCUSIÓN

¿ ?

## Test de $\chi^2$ de tendencia lineal



Grupo etario	Mantoux (+)	Mantoux (-)	N	Prevalencia
Menos de 25 años	221	721	942	28,2%
25 - 35 años	242	482	724	33,5%
> 35 años	15	38	53	27,8%

$\chi^2$  tendencia lineal = 5,2  p = 0,07

“Utiliza la estadística como un borracho usa una farola, para aguantarse en lugar de para iluminarse”

Andrew Lang

## Test de $\chi^2$ de Pearson de independencia



Grupo etario	Mantoux (+)	Mantoux (-)	N	Prevalencia
Menos de 25 años	221	721	942	28,2%
25 - 35 años	242	482	777	33,1%

$\chi^2 = 26,07$    $p=0,0000003$

“Utiliza la estadística como un borracho usa una farola, para aguantarse en lugar de para iluminarse”

Andrew Lang

## Pregunta de investigación

¿Existen diferencias en la prevalencia de infección tuberculosa en el colectivo de inmigrantes residentes en Toledo según la edad?

Variable PREDICTORA  
CUALITATIVA policotómica

Inmigrantes  
< 20 años  
25-29 años  
30-39 años



Variable RESULTADO  
CUALITATIVA dicotómica

Tuberculosis

¿Cuál es el test estadístico más adecuado?

Alonso Moreno FJ, García Bajo MC, Lougedo Calderón MJ, et al. Prevalencia de infección tuberculosa en las personas inmigrantes del área sur de Toledo. Rev Esp Salud Pública 2004; 78: 593-600.

### Análisis del texto en el MATERIAL y MÉTODOS

“En el análisis se emplearon los tests de la  $\chi^2$  de Pearson, para comparación de porcentajes, y la t de Student y análisis de varianza (ANOVA), para comparación de medias y varianzas. Se realizó también una regresión logística no condicional por el método stepwise backward (LR). Las variables independientes incluidas en el modelo fueron la edad, sexo, peso y origen subsahariano, mientras que la variable dependiente fue la IDRМ positiva”.

### Análisis del texto en los RESULTADOS

“La prevalencia de infección tuberculosa por grupos de edad se muestra en la tabla 2, en la que puede apreciarse cómo la mayor frecuencia (38,7%) apareció en el grupo de edad 30-39 años.”.

### Análisis del texto en la DISCUSIÓN

“En ésta última la prevalencia encontrada tanto en el grupo de edad de 20 a 29 años (12,1%), como en el de 30 a 39 años (22,7%), está en ambos casos por debajo de los resultados de nuestro estudio. Los resultados de nuestro trabajo aportan valores en la misma línea que los obtenidos por Durán y colaboradores<sup>13</sup> en inmigrantes recién llegados a Barcelona”.

Alonso Moreno FJ, García Bajo MC, Lougedo Calderón MJ, et al. Prevalencia de infección tuberculosa en las personas inmigrantes del área sur de Toledo. Rev Esp Salud Pública 2004; 78: 593-600.

## Test de $\chi^2$ de tendencia lineal



Grupo etario	Prueba de la tuberculina		N	Transversal	Casos- controles
	Mantoux (+)	Mantoux (-)		Prevalencia	OR
Menos de 25 años	5	61	64	7,8%	1,00
25 - 35 años	21	41	62	33,9%	6,25
> 35 años	36	57	93	38,7%	7,71

$\chi^2$  tendencia lineal = 17,4



p = 0,00003

## Pregunta de investigación

¿Qué relación existe entre la edad y la presencia de deterioro cognitivo moderado-severo en un grupo de enfermos mayores ( $\geq$  años) atendidos en los servicios de urgencia del hospital?

Variable PREDICTORA  
Cualitativa ORDINAL

Grupo etario  
- 60-64 años  
- 65-69 años  
- 70-74 años  
- 75-79 años  
- 80-84 años  
 $\geq$  85 años



Variable RESULTADO  
Cualitativa dicotómica

Deterioro cognitivo

¿Cuál es el test estadístico más adecuado?

Laguna P, Moya M, Gazapo R, Salgado R, Calabrese S. Detección de deterioro cognitivo en ancianos atendidos en un servicio de urgencias hospitalario. Rev Clín Esp 1997; 197: 336-340.

## Test de $\chi^2$ de tendencia lineal

Categorización

Variable RESULTADO  
Cuantitativa continua



Variable RESULTADO  
Cualitativa nominal ordinal

Grupo etario	Deterioro cognitivo			Transversal	Casos-contrroles	
	SI	NO	N	Prevalencia	OR	IC95%
60 - 64 años	8	41	49	16,3%	1,00	
65 - 69 años	14	56	70	20 %	1,28	0,45 - 3,72
70 - 74 años	19	64	83	22,9%	1,52	0,56 - 4,20
75 - 79 años	18	51	69	26,1%	1,81	0,66 - 5,09
80 - 84 años	18	33	51	35,3%	2,80	0,99 - 8,09
≥ 85 años	17	13	30	56,7%	6,70	2,11 – 22,07
Total	94	258	352	26,7%		

$\chi^2$  tendencia lineal = 16,3



$p = 0,0005$

## Mito I: “La sonrisa de la esfinge”

¿Por qué estratificas la variable edad en 6 categorías?

¿Por qué no categorizas la variable edad en sólo 3 estratos ordinales?

Aunque el cálculo de la prueba  $\chi^2_{\text{Pearson}}$  de tendencia lineal ha resultado significativa, el IC95% de los OR de los distintos grupos de edad incluyen el 1 (valor numérico de igualdad), con lo que perdemos información...

El test de  $\chi^2$  de tendencia lineal, aunque admite más de 3 categorías ordinales para la variable predictora, no es conveniente estratificar más allá a no ser que se disponga de un importante volumen muestral, de lo contrario los IC95% no son significativos.

## Test de $\chi^2$ de tendencia lineal

### Categorización

Variable RESULTADO  
Cuantitativa continua



Variable RESULTADO  
Cualitativa nominal ordinal

Grupo etario	Deterioro cognitivo			N	Transversal	Casos-contróles	
	SI	NO	Prevalencia		OR	IC95%	
60 - 69 años	22	97	119	18,5%	1,00		
70 - 79 años	37	115	152	32,2%	1,42	(1,12 - 1,83)	
≥ 80 años	35	46	81	43,2%	3,35	(2,05 - 4,96)	
Total	94	258	352	26,7%			

El análisis de tendencias demostró una relación estadísticamente significativa entre la edad del anciano y el deterioro cognitivo de grado moderado o severo ( $p < 0,001$ ).

$\chi^2$  tendencia lineal = 13,9   $p = 0,0002$

## Pregunta de investigación

¿Existe relación entre el hábito tabáquico y la presencia del ADN del Papiloma virus (HPV) en pacientes con lesiones escamosas intraepiteliales del cérvix uterino?

Variable PREDICTORA  
Cualitativa ORDINAL

Hábito tabáquico  
(n° cigarrillos/día)  
1. No fumador (0 )  
2. Moderado ( $\leq 20$ )  
3. Alto ( $> 20$ )



Variable RESULTADO  
Cualitativa dicotómica

ADN del HPV  
Human Papilloma Virus

¿Cuál es el test estadístico más acertado?

Laguna P, Moya M, Gazapo R, Salgado R, Calabrese S. Detección de deterioro cognitivo en ancianos atendidos en un servicio de urgencias hospitalario. Rev Clín Esp 1997; 197: 336-340.

## Test de $\chi^2$ de tendencia lineal

TABACO	ADN HPV			OR	IC 95%	p
	(+)	(-)	%			
No fumador	5	13	27,8	1,00		
Moderado	19	24	44,2	2,06	0,55 – 8,62	0,36
Alto	4	1	80,0	10,40	0,70 – 550,31	0,06
TOTAL	28	38		$\chi^2_{\text{tendencia lineal}} = 4,022$		0,045

Tabla de contingencia que representa el hábito tabáquico (nº de cigarrillos al día) y la presencia del ADN del papiloma virus (HPV) en pacientes con lesiones escamosas intraepiteliales del cérvix uterino.

## Resultados (fragmento del texto)

“No se observaron diferencias significativas entre la edad media y la edad media de inicio de las relaciones sexuales, y la positividad del HPV. Por el contrario, la asociación con el hábito tabáquico fue estadísticamente significativa, observándose una asociación creciente ente la detección del HPV y el consumo de cigarrillos (tabla)”.

Torné A, Ordi J, Lluís M, et al. Determinación del papiloma virus mediante hibridación in situ. Correlación clínicopatológica y virológica en pacientes con lesiones escamosas intraepiteliales del cérvix uterino. Med Clin (Barc) 1997; 109: 691-695.

## Discusión (fragmento del texto)

“Diversos estudios han destacado el efecto que el consumo de tabaco puede ejercer en interacción con el HPV facilitando y cooperando con los mecanismos oncogénicos. En el presente estudio, el hallazgo de una asociación creciente entre la detección del HPV y el consumo de cigarrillos puede relacionarse con el resultado de estudios experimentales que demuestran que los productos del tabaco causan un defecto inmunológico a nivel cervical, disminuyendo la capacidad de presentación de antígenos por las células de langerhans, fenómeno que finalmente favorece tanto la infección por HPV en el epitelio cervical”.

Torné A, Ordi J, Lluís M, et al. Determinación del papiloma virus mediante hibridación in situ. Correlación clinicopatológica y virológica en pacientes con lesiones escamosas intraepiteliales del cérvix uterino. Med Clin (Barc) 1997; 109: 691-695.

## Mito I: “La sonrisa de la esfinge”

Aunque el cálculo de la prueba  $\chi^2_{\text{Pearson}}$  de tendencia lineal ha resultado significativa ( $p=0,045$ ), sin embargo el IC95% de los OR de los distintos grupos de fumadoras (moderadas e intensas) incluyen el 1 (valor numérico de igualdad), con lo que perdemos información, además se pierde significación estadística ( $p>0,05$ )...

El test de  $\chi^2$  de tendencia lineal, aunque admite 3 o más categorías ordinales para la variable predictora, NO es conveniente que los casilleros contengan valores inferiores a 5, como ocurre en el grupo de las fumadoras intensas (> 20 cigarrillos/día) .

¿Sabes lo que ocurre si agrupas las fumadoras moderadas con las intensas y la enfrentas a las no fumadoras (no fumadoras versus fumadoras)?

¡¡¡OBSERVA LO QUE OCURRE!!!

Tabla de contingencia que representa la presencia del HPV con el hábito tabáquico.

	ADN HPV		
TABACO	SI (+)	NO (-)	Total
NO	5	13	18
SI	23	25	48
TOTAL	28	38	66

## Test de $\chi^2$ de Pearson de independencia

$$\chi^2 = 2,17$$



$$p=0,140$$

No existe asociación

## Mito I: “La sonrisa de la esfinge”

¿Por qué torturas a los datos con test estadísticos NO idóneos para que te digan lo que a ti te interesa?

¿Sabes lo que te ha ocurrido?

Te ha abrumado el enorme peso de la bibliografía que es muy numerosa en estudios experimentales y clínicos que relacionan el tabaco, el HPV y el cáncer de cérvix, y su interacción.

Ante tan bajo volumen muestral...

¿No crees que puedes haber cometido un error beta (bajo tamaño muestral)?

Síndrome de la CARGA bibliográfica

## Pregunta de investigación

¿Existe asociación entre el hábito tabáquico y la EPOC?

Variable PREDICTORA  
Cualitativa dicotómica

Hábito tabáquico



Variable RESULTADO  
Cualitativa dicotómica

EPOC

¿Cuál es el test estadístico más acertado?

Tabla de contingencia que representa la EPOC frente al hábito tabáquico.

	EPOC		
TABACO	SI	NO	Total
SI	26 (43%)	34	60
NO	7 (23%)	23	30
TOTAL	33	57	90

## Test de $\chi^2$ de Pearson de independencia

$$\chi^2 = 3,44$$



$$p=0,063$$

No existe asociación

## Pregunta de investigación

¿Existe relación entre la EPOC y el la intensidad del hábito tabáquico?

Variable PREDICTORA  
Cualitativa ordinal

Hábito tabáquico  
1. No fumador  
2. Moderado  
3. Alto



Variable RESULTADO  
Cualitativa dicotómica

EPOC

¿Cuál es el test estadístico más acertado?

Tabla de contingencia que representa la EPOC frente al hábito tabáquico.

	EPOC		
TABACO	SI	NO	Total
No fumador	7	23	30
Moderado	11	19	30
Alto	15	15	30
TOTAL	33	57	90

## Test de $\chi^2$ de Pearson de independencia

$$\chi^2 = 4,60$$



$$p=0,111$$

No existe asociación

## Mito I: “La sonrisa de la esfinge”

¿Por qué utilizas el test de  $\chi^2_{\text{Pearson}}$ ?

¿Se cumplen sus condiciones de aplicación?

En el cálculo de la prueba  $\chi^2_{\text{Pearson}}$  no se tiene en cuenta el orden de las categorías y de este modo perdemos una valiosa información..

El test de  $\chi^2$  de tendencia lineal, siempre tiene un grado de libertad, e incluye una modificación para su cálculo respecto a la expresión de la prueba  $\chi^2_{\text{Pearson}}$ , de modo se tiene en cuenta el orden de las categorías.

Tabla de contingencia que representa la EPOC frente al hábito tabáquico.

TABACO	EPOC		Total	Caso-control
	SI	NO		OR
No fumador	7	23	30	1,00
Moderado	11	19	30	1,90
Alto	15	15	30	3,29
TOTAL	33	57	90	

## Test de $\chi^2$ de tendencia lineal

$$\chi^2 = 4,54$$



$$p = 0,033$$

Existe asociación

## Pregunta de investigación

¿Existe relación entre la gravedad de la EPOC y la intensidad del hábito tabáquico?

Variable PREDICTORA  
Cualitativa ordinal

Hábito tabáquico  
1. No fumador  
2. Moderado  
3. Alto



Variable RESULTADO  
Cualitativa ordinal

Gravedad EPOC  
1. Leve  
2. Moderada  
3. Grave

¿Cuál es el test estadístico más acertado?

Tabla de contingencia que representa la gravedad de la EPOC frente al hábito tabáquico

TABACO	EPOC			Total
	Leve	Moderado	Grave	
No Fumador	20	8	2	30
Moderado	8	10	12	30
Alto consumo	2	8	20	30
TOTAL	30	26	34	90

Test de  $\chi^2$  de tendencia lineal

Existe asociación

Coefficiente Gamma  $\Upsilon$

$$\Upsilon = 0,718$$

Fuerte asociación positiva entre ambas variables

## Pregunta de investigación

¿Existe asociación entre el nivel de instrucción (cultural) y el nivel económico?

¿A mayor nivel cultural habrá mayor nivel económico?

Variable PREDICTORA  
Cualitativa ordinal

Nivel instrucción



Variable RESULTADO  
Cualitativa ordinal

Nivel económico

¿Cuál es el test estadístico más acertado?

## ¿Existe asociación entre el nivel de instrucción y el nivel económico?

Tabla de contingencia que representa el nivel de estudios frente al nivel económico

ESTUDIOS	Nivel económico			TOTAL
	Alto	Medio	Bajo	
Universitarios	19	5	1	25
Secundarios	14	22	1	37
Primarios	7	36	7	50
TOTAL	40	63	9	112

$$\chi^2_{\text{Pearson}} = 30,15$$



$$p < 0,000001$$

SI

¿Existe relación gradual entre el nivel de instrucción y el nivel económico?

Tau-b de Kendall

¿Existe relación gradual entre el nivel de instrucción y el nivel económico?

$H_0$ : A mayor nivel de instrucción no existe mayor nivel económico.

$H_a$ : A mayor nivel de instrucción mayor nivel económico.

## Tau-b de Kendall

Ordinal por ordinal	Valor del estadístico	Significación
Tau-b de Kendall	+ 0,448	0,000
Tau-c de Kendall	+ 0,399	0,000

$H_a$ : A mayor nivel de instrucción mayor nivel económico.

SI

## Tau-b de Kendall

- Condiciones de aplicación.
  - Tablas cuadradas (3 x 3), (4 x 4)...
- Interpretación:
  - $\tau = 0$  independencia.
  - $\tau > 0$  asociación positiva.
  - $\tau < 0$  asociación negativa.
- Intervalo de Tau : (-1, +1).
- Información:
  - Medida de la intensidad de la asociación de variables ordinales.
  - Sentido de la misma (direccionalidad).

Estadístico	Tipo Variables	Tipo tabla
Tau-b de Kendall	Ordinal X Ordinal	K = F
Tau-c de Kendall	Ordinal X Ordinal	K ≠ F

## Tau-b de Kendall

- Condiciones de aplicación: número de filas distinto al número de columnas (Tablas NO cuadradas).

## Pregunta de investigación

¿Existe asociación entre el nivel de instrucción (cultural) y el nivel económico?

¿A mayor nivel cultural habrá mayor nivel económico?

Variable PREDICTORA  
Cualitativa ordinal

Nivel instrucción



Variable RESULTADO  
Cualitativa ordinal

Nivel económico

¿Cuál es el test estadístico más acertado?

## ¿Existe asociación entre el nivel de instrucción y el nivel económico?

Tabla de contingencia que representa el nivel de estudios frente al nivel económico

ESTUDIOS	Nivel económico			TOTAL
	Alto	Medio	Bajo	
Universitarios	19	5	1	25
Secundarios	14	22	1	37
Primarios	7	36	7	50
TOTAL	40	63	9	112

$$\chi^2_{\text{Pearson}} = 30,15$$



$$p < 0,000001$$

SI

¿Existe relación gradual entre el nivel de instrucción y el nivel económico?

Coeficiente Gamma  $\Upsilon$

¿Existe relación gradual entre el nivel de instrucción y el nivel económico?

$H_o$ : A mayor nivel de instrucción no existe mayor nivel económico.

$H_a$ : A mayor nivel de instrucción mayor nivel económico.

## Coeficiente Gamma $\Upsilon$

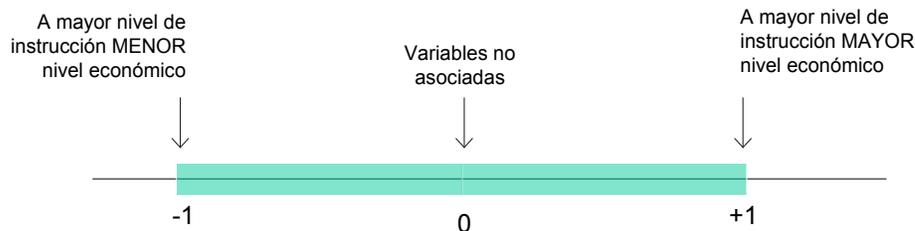
Ordinal por ordinal	Valor de $\Upsilon$	EEY	$T_{\text{aprox}}$	Significación
Gamma	+ 0,687	0,098	5,804	0,000
Nº de casos válidos	112			

$H_a$ : A mayor nivel de instrucción mayor nivel económico.

SI

## Coeficiente Gamma $\Upsilon$

- Test de contraste de hipótesis.
- Test de asociación para variables ordinales.
- Interpretación:
  - $\Upsilon = 0$             independencia
  - $\Upsilon > 0$             asociación positiva
  - $\Upsilon < 0$             asociación negativa
- Intervalo de  $\Upsilon$  : (-1, +1).
- Información:
  - Medida de la intensidad de la asociación de variables ordinales.
  - Sentido de la misma (direccionalidad).



## Pregunta de investigación

¿Existe asociación entre el nivel de instrucción (cultural) y el nivel económico?

¿A mayor nivel cultural habrá mayor nivel económico?

Variable PREDICTORA  
Cualitativa ordinal

Nivel instrucción



Variable RESULTADO  
Cualitativa ordinal

Nivel económico

¿Cuál es el test estadístico más acertado?

## ¿Existe asociación entre el nivel de instrucción y el nivel económico?

Tabla de contingencia que representa el nivel de estudios frente al nivel económico

ESTUDIOS	Nivel económico			TOTAL
	Alto	Medio	Bajo	
Universitarios	19	5	1	25
Secundarios	14	22	1	37
Primarios	7	36	7	50
TOTAL	40	63	9	112

$$\chi^2_{\text{Pearson}} = 30,15$$



$$p < 0,000001$$

SI

$H_a$ : El nivel económico (variable resultado) es dependiente del nivel de instrucción (variable predictora).

¿Existe relación gradual entre el nivel de instrucción y el nivel económico?

D de Somers

¿Existe relación gradual entre el nivel de instrucción y el nivel económico?

$H_o$ : A mayor nivel de instrucción no existe mayor nivel económico.

$H_a$ : A mayor nivel de instrucción mayor nivel económico.

## D de Somers

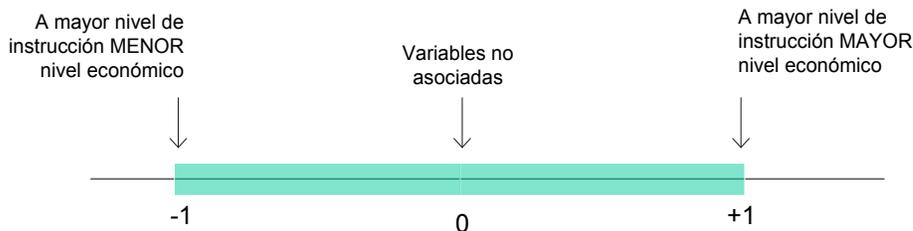
	Valor del estadístico D	Significación
Simétrica (no se sabe cual es la resultado)	0,447	0,000
Nivel de instrucción resultado (dependiente)	0,484	0,000
Nivel económico resultado (dependiente)	0,415	0,000

$H_a$ : A mayor nivel de instrucción mayor nivel económico.

SI

## D de Somers

- Test de contraste de hipótesis.
- Test de asociación para variables ordinales.
- Interpretación:
  - D = 0            independencia
  - D > 0           asociación positiva
  - D < 0           asociación negativa
- Intervalo de D de Somers: (-1, +1).
- Información:
  - Medida de la intensidad de la asociación de variables ordinales.
  - Sentido de la misma (direccionalidad).



## Pregunta de investigación

¿Existe asociación entre el nivel de instrucción (cultural) y el nivel económico?

¿A mayor nivel cultural habrá mayor nivel económico?

Variable PREDICTORA  
Cualitativa ordinal

Nivel instrucción



Variable RESULTADO  
Cualitativa ordinal

Nivel económico

## RESUMEN

¿Cuál es los tests estadísticos más acertados?

	Nivel económico				
ESTUDIOS	Alto	Medio	Bajo	Muy bajo	TOTAL
Titulado superior	12	32	6	0	50
Titulado medio	5	21	136	19	181
Secundarios	0	1	94	95	190
Primarios	0	0	24	29	53
TOTAL	17	54	260	143	474

¿Existe relación gradual entre el nivel de instrucción y el nivel económico?

D de Somers

Gamma  $\Upsilon$

Tau-b de Kendall

Tau-c de Kendall

D de Sommers (medidas direccionales)	Valor del estadístico	EED	T <sub>aprox</sub>	Muy bajo
Simétrica (no se sabe cual es la variable resultado)	0,557	0,029	16,075	0,000
Nivel económico variable resultado (dependiente)	0,525	0,030	16,075	0,000
Nivel de instrucción variable resultado (dependiente)	0,593	0,030	16,075	0,000

Ordinal por ordinal	Valor del estadístico	EE	T <sub>aprox</sub>	Significación
Gamma	0,798	0,031	16,075	0,000
Tau-b de Kendall	0,558	0,029	16,075	0,000
Tau-c de Kendall	0,469	0,029	16,075	0,000
Nº de casos válidos	474			

H<sub>a</sub>: A mayor nivel de instrucción mayor nivel económico.

## Capítulo 7

Publicación Bioestadística: Análisis bivariante.  
Test de contraste de hipótesis.  
Cuantitativa versus cuantitativa.

Dr. Jacobo Díaz Portillo  
Coordinador de Formación Continuada  
Hospital Universitario de Ceuta

# Materiales y métodos

## Estructura

1. Tipo de diseño (estudio).
2. Sujetos incluidos (criterios de selección de la muestra)
3. Fuentes de información.
4. Descripción de la intervención y definición del punto final de la investigación (estudios experimentales).
5. Instrumentos y procedimientos de medida (cuestionarios).
6. Definición de las variables y covariables incluidas en el estudio (factor de estudio y criterio de evaluación).
7. Limitaciones y consideraciones éticas de la investigación.
8. Pruebas estadísticas a utilizar (programa estadístico).

## Fases del análisis estadístico

1. ANÁLISIS DESCRIPTIVO



2. ANÁLISIS BIVARIANTE



3. ANÁLISIS MULTIVARIANTE

## Decisiones que se pueden tomar en un test de contraste de hipótesis después de conocer el valor de significación estadística $p$

$p < 0,05$ $A \neq B$	$p \geq 0,10$ $A = B$
Se rechaza la hipótesis nula	No se puede rechazar la hipótesis nula
No parece que el azar pueda explicarlo todo	No se puede descartar que el azar lo explique todo
El EFECTO observado es mayor que el ERROR	El EFECTO observado es similar que el ERROR
Hay diferencias estadísticamente significativas	NO hay diferencias estadísticamente significativas
Existen evidencias a favor de la hipótesis alternativa $H_a$ , y por tanto rechazamos la hipótesis nula $H_0$	NO existen evidencias a favor de la hipótesis alternativa $H_a$ , y por tanto aceptamos la hipótesis nula $H_0$

Los límites de 0,05 y 0,1 pueden ser considerados en cierto modo arbitrarios y aproximados

## Análisis estadístico

Tipos de test analíticos estadísticos de contraste de hipótesis según el tipo de variables implicadas en el estudio

Variable dependiente

Variable resultado

		Dicotómica	Continua
Variable Predictora	Dicotómica	Test de $\chi^2$	t de Student
Variable Independiente	Continua	Regresión logística	Regresión lineal

## Análisis estadístico

ANÁLISIS ESTADÍSTICO		Variable resultado	
		Dicotómica	Continua
Variable Predictora	Dicotómica	Test de $\chi^2$	t de Student
Variable Independiente	Continua	Regresión logística	Regresión lineal

REGRESIÓN LINEAL SIMPLE

## Pruebas estadísticas de contraste de hipótesis más aplicables a cada situación

Variable Independiente o PREDICTORA	Variable Dependiente o RESULTADO	Pruebas de contraste de hipótesis	Observaciones y/o limitaciones del test
Cualitativa	Cualitativa	Ji cuadrado Prueba exacta de Fisher Test de McNemar Q de Cochran REGRESIÓN LOGÍSTICA	Si el tamaño muestral es grande Si el tamaño muestral es pequeño Dos grupos de muestras repetidas o apareadas Más de dos grupos de muestras repetidas o apareadas Varias variables predictoras cualitativas
Cualitativa	Cuantitativa	t de Student Análisis de la Varianza Test de Mann-Whitney Test de Wilcoxon Test de Kruskal-Wallis Test de Friedman	Compara medias entre dos grupos independientes Compara medias entre más de dos grupos independientes Dos grupos independientes No paramétricos Dos grupos medidas repetidas No paramétrico Más de dos grupos No paramétrico, medidas repetidas Más de dos grupos No paramétrico medidas repetidas
Cuantitativa	Cualitativa	Regresión logística simple Regresión logística multivariante	Una sola variable predictora cuantitativa Varias variable predictoras cuantitativa o cualitativas
Cuantitativa	Cuantitativa	Regresión lineal simple Correlación de Pearson Correlación de Sperman REGRESIÓN MÚLTIPLE	Paramétrico Paramétrico No paramétrico Varias variables predictoras cuantitativas
Cualitativa	Supervivencia	Kaplan-Meier Long-Rank Regresión de COX	Curvas de supervivencia Compara curvas de supervivencia Análisis Multivariante

## Tipos de test analíticos estadísticos bivariantes de contraste de hipótesis más frecuentes

Variable PREDICTORA o Independiente (factor de estudio)	Variable RESULTADO o Dependiente (Criterio de evaluación)			
	Cualitativa nominal dos categorías	Cualitativa nominal > 2 categorías	Cualitativa ordinal o Cuantitativa NO paramétrica	Cuantitativa paramétrica
Cualitativa dos grupos independientes	Chi cuadrado Z comparación de proporciones Test exacto de Fisher	Chi cuadrado	U de Mann-Whitney	T de Student (prueba de Welch)
Cualitativa dos grupos apareados	Test de McNemar	Q de Cochran	Prueba de los rangos de Wilcoxon	T de Student para datos emparejados
Cualitativa > 2 grupos independientes	Chi cuadrado	Chi cuadrado	Prueba de Kruskal-Wallis	Análisis de la Varianza (ANOVA)
Cualitativa > 2 grupos apareados	Q de Cochran	Q de Cochran	Prueba de Friedman	Análisis de la Varianza de dos vías
Cuantitativa	Regresión logística	Regresión logística	Correlación de Spearman Tau de Kendall	Regresión lineal Correlación de Pearson

## Cualitativa versus cuantitativa

¿Cuáles son las pruebas estadísticas que tenemos que emplear?

Variable PREDICTORA  
CUANTITATIVA

- Continua
- Discreta



Variable RESULTADO  
CUANTITATIVA

- Continua
- Discreta

## Datos cuantitativos:

### Cuantitativa versus cuantitativa

Regresión lineal simple:

- Describe como es la relación entre la variable predictora ( $X$ ) y la variable resultado ( $Y$ )
- Busca la línea que mejor se ajusta a los puntos  $(x, y)$ , y por tanto busca la mejor predicción de  $Y$  a partir de  $X$

Correlación:

- Mide el grado o fuerza de asociación entre dos variables cuantitativas ( $X, Y$ ).

## Datos cuantitativos:

### Cuantitativa versus cuantitativa

Regresión lineal simple.

- Coeficiente de regresión  $\beta$
- Covarianza

Correlación paramétrica

- Coeficiente de correlación de Pearson ( $r$ )
- Coeficiente de determinación  $r^2$

Correlación NO paramétrica

- Coeficiente de correlación de Spearman ( $\rho$ )

Concordancia

- Coeficiente de correlación intraclase

## Datos cuantitativos:

### Cuantitativa versus cuantitativa

#### Comparación de métodos analíticos

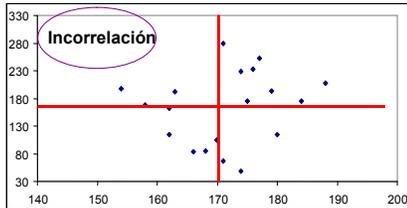
##### Regresión paramétrica

- Regresión lineal simple
  - Coeficiente de regresión  $\beta$
- Coeficiente de correlación intraclase

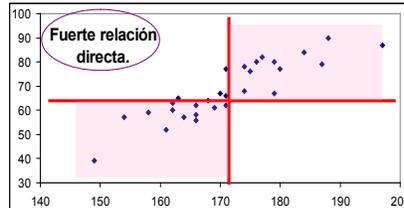
##### Regresiones NO paramétricas

- Regresión de Theil
- Regresión de Passing - Bablok
- Regresión de Bland - Altman
- Regresión de Tholen - Ludbrook

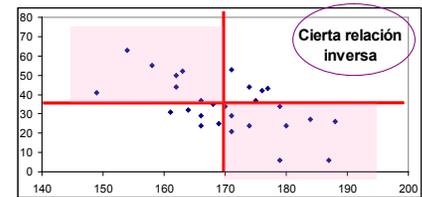
## Relación directa e inversa



Para valores de  $X$  por encima de la media tenemos valores de  $Y$  por encima y por debajo en proporciones similares. **Incorrelación.**



Para los valores de  $X$  mayores que la media le corresponden valores de  $Y$  mayores también. Para los valores de  $X$  menores que la media le corresponden valores de  $Y$  menores también. Esto se llama **relación directa.**



Para los valores de  $X$  mayores que la media le corresponden valores de  $Y$  menores. Esto es **relación inversa** o decreciente.

## Pregunta de investigación

¿Qué relación existe entre el grado de IRC y el grado de anemización del enfermo?

Variable PREDICTORA (X)  
Cuantitativa continua

IRC

Creatinemia

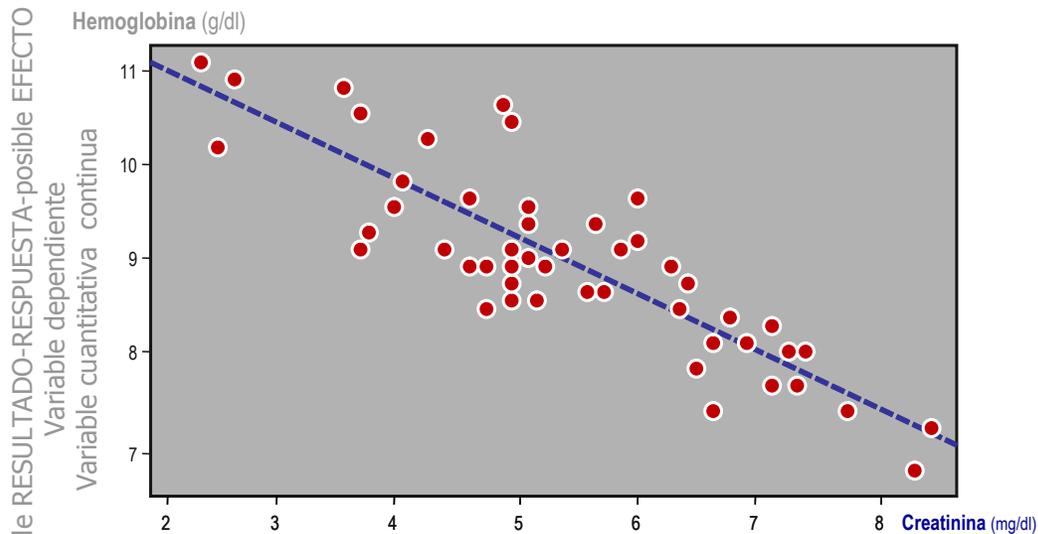


Variable RESULTADO (Y)  
Cuantitativa continua

Anemia

Hemoglobinemia

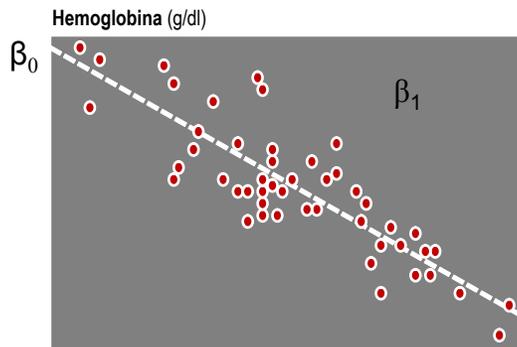
## Relación entre variables cuantitativas continuas



Representación gráfica de los niveles de creatinina y hemoglobina en 50 pacientes con fallo renal crónico

Variable PREDICTORA-EXPOSICIÓN-posible CAUSA  
Variable independiente, fija controlada por el investigador  
Variable cuantitativa continua

## Regresión lineal simple. Coeficiente de regresión



Representación gráfica de los niveles de creatinina y hemoglobina en 50 pacientes con fallo renal crónico

Estimaciones de los parámetros y sus errores estándar para en la relación de los niveles de creatinina y hemoglobina en 50 pacientes con fallo renal crónico

Variable	Estimación	Error estándar
Constante	12,25	0,2984
Creatinina	-0,5708	0,05312

$$E(Y) = \beta_0 + \beta_1 * X$$

Coeficiente de regresión  $\beta_1$

$$Hb (Y) = 12,25 - 0,5708 * [Creatinina]$$

## Regresión Lineal Simple. Coeficiente de regresión

¿El valor de  $\beta_1$  es estadísticamente distinto de cero y por tanto podemos decir que existe dependencia lineal entre los niveles de Hb y los de Cr en los pacientes con IRC?

Z	Valor	$\alpha$	Error $\alpha$
$Z_{90\%}$	1,64	90%	10%
$Z_{95\%}$	1,96	95%	5%
$Z_{99\%}$	2,56	99%	1%
$Z_{99,5\%}$	3,29	99,5%	0,5%

Estimaciones de los parámetros y sus errores estándar para en la relación de los niveles de creatinina y hemoglobina en 50 pacientes con fallo renal crónico

Variable	Estimación	Error estándar
Constante	12,25	0,2984
Creatinina	- 0,5708	0,05312

$$\frac{\beta_1}{EE\beta_1} = \frac{0,5708}{0,05312} = 10,74 > Z_{95\%}$$

Podemos decir con error  $\alpha$  inferior al 5% que el nivel de Hb está asociado o depende del nivel de Cr.

## Cálculo del Intervalo de Confianza (IC) del coeficiente de regresión

¿Cuál es el IC al 95% de confianza para el cambio en el nivel de Hb por unidad de aumento en el nivel de Cr en los pacientes con IRC?

Z	Valor	$\alpha$
$Z_{95\%}$	1,96	95%
$Z_{99\%}$	2,56	99%
$Z_{99,5\%}$	3,29	99,5%

$$\beta_1 \pm Z_{95\%} \times EE\beta_1$$

$$-0,5708 \pm (1,96) \times 0,05312 = (-0,67, -0,47)$$

Podemos decir con error  $\alpha$  inferior al 5% , es decir con una confianza del 95%, que el descenso medio del nivel de Hb será como mínimo 0,47 g/dl y como máximo 0,67 g/dl por aumento en 1 mg/dl en el nivel de Cr. El hecho de que el IC95% no contenga al cero esta de acuerdo con el resultado del contraste de hipótesis realizado ( $p < 0,05$ ).

## Covarianza de dos variables X e Y

- La covarianza entre dos variables,  $S_{xy}$ , nos indica si la posible relación entre dos variables es directa o inversa.

Directa:  $S_{xy} > 0$

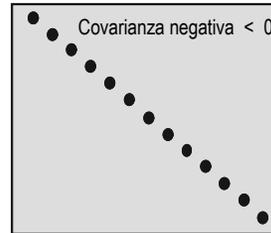
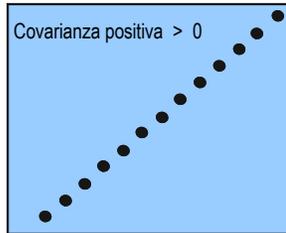
Inversa:  $S_{xy} < 0$

Incorreladas:  $S_{xy} = 0$

$$S_{xy} = \frac{1}{n} \sum_i (x_i - \bar{x})(y_i - \bar{y})$$

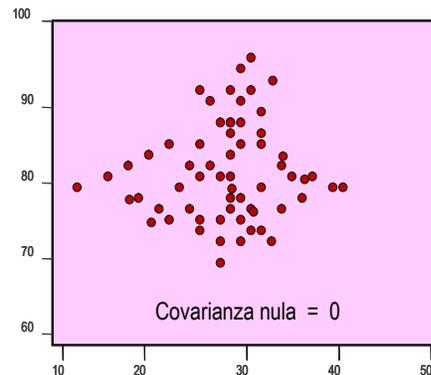
- El signo de la covarianza nos dice si el aspecto de la nube de puntos es creciente o no, pero no nos dice nada sobre el grado de relación entre las variables.

## Regresión lineal simple: Covarianza



$$S_{xy} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

Valores de la COVARIANZA dependiendo de la variación al unísono de x e y



$$r_{xy} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n \cdot S_x \cdot S_y}$$

$$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y}$$

El valor de la COVARIANZA es cero cuando no hay variación al unísono de x e y

## Covarianza y coeficiente de correlación de Pearson ( $r$ )

$$r_{xy} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n \cdot S_x \cdot S_y}$$

$$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y}$$

$$r = \frac{\text{covarianza}(xy)}{\sqrt{SC_x SC_y}}$$

$$r = \frac{SP_{xy}}{\sqrt{SC_x SC_y}}$$

$$r = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}}$$

## Coeficiente de correlación lineal de Pearson

$$r = \frac{S_{xy}}{S_x S_y}$$

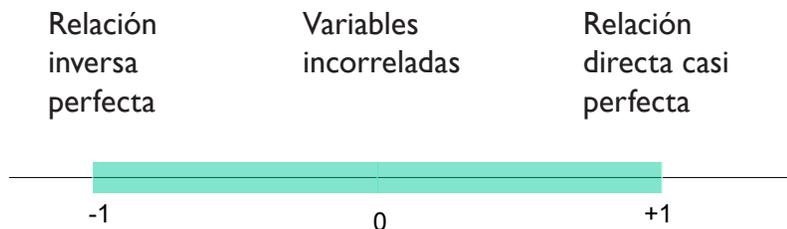
- La coeficiente de correlación lineal de Pearson de dos variables,  $r$ , nos indica si los puntos tienen una tendencia a disponerse alineadamente.
  - El coeficiente tiene el mismo signo que la covarianza  $S_{xy}$  por tanto su signo nos informa del tipo de relación, directa o inversa.
- $r$  es útil para determinar si hay relación lineal entre dos variables, pero no servirá para otro tipo de relaciones (cuadrática, logarítmica, ...)
- El coeficiente de correlación no se debe extrapolar más allá del rango de valores observado de las variables en estudio ya que la relación existente entre  $X$  e  $Y$  puede cambiar fuera de dicho rango.
  - La correlación no implica causalidad. La causalidad es un juicio de valor que requiere más información que un simple valor cuantitativo de un coeficiente de correlación (\*).
  - No debe utilizarse para comparar dos métodos que intentan medir el mismo evento (variable), como por ejemplo dos instrumentos que miden la tensión arterial. El coeficiente  $r$  mide el grado de asociación entre dos cantidades pero no mira el nivel de acuerdo o concordancia. Si los instrumentos de medida miden sistemáticamente cantidades diferentes uno del otro, la correlación puede ser  $r \approx 1$  y su concordancia ser nula (\*\*).

(\*) Pita Fernández S. Correlación frente a causalidad JANO 1996; (1174): 59-60.

(\*\*) Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1: 307-310.

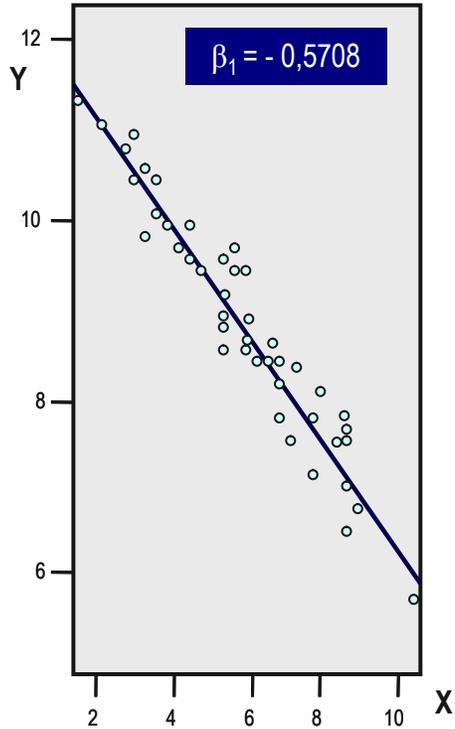
## Propiedades de $r$

- Es adimensional
- Sólo toma valores en  $[-1,1]$
- Las variables son incorreladas  $\leftrightarrow r=0$
- Relación lineal perfecta entre dos variables  $\leftrightarrow r=+1$  o  $r=-1$ 
  - Excluimos los casos de puntos alineados horizontal o verticalmente.
- Cuanto más cerca esté  $r$  de  $+1$  o  $-1$  mejor será el grado de relación lineal.
  - Siempre que no existan observaciones anómalas.

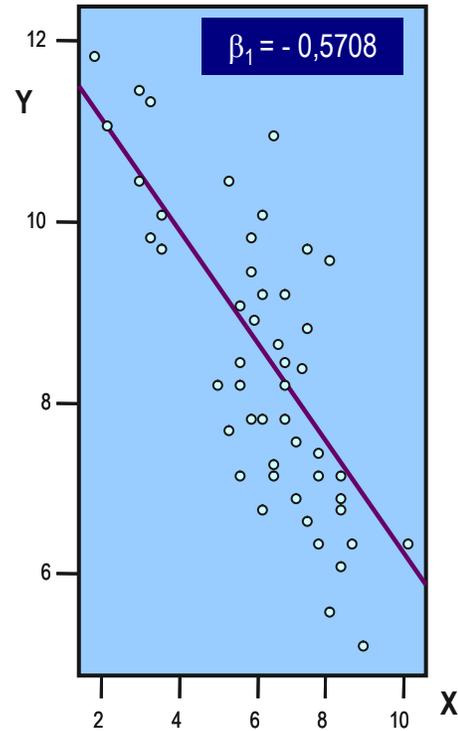


Horbar JD, Onstad L, Wright E. Predicting mortality risk for infants weighing 501 to 1500 grams at birth: The National Institutes of Health Neonatal Research Network report. Crit Care Med 1993; 21:12-18.

$$Hb (Y) = 12,25 - 0,5708 * [Creatinina]$$

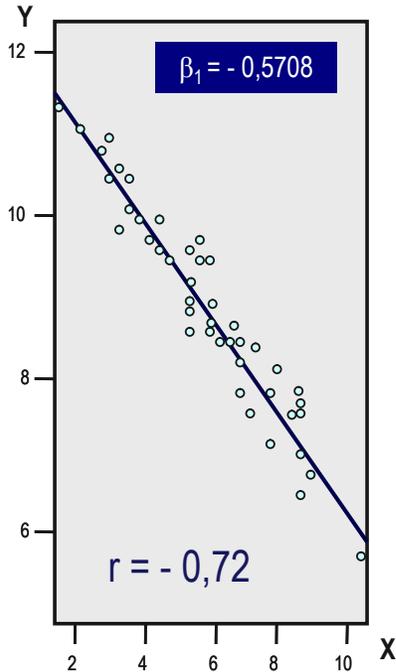


$$Hb (Y) = 12,25 - 0,5708 * [Creatinina]$$

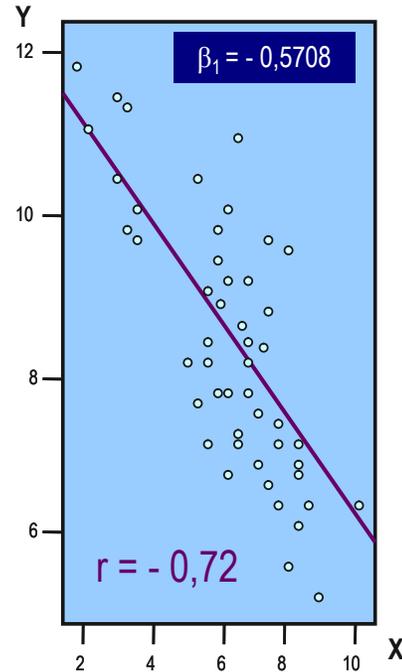


Nube de puntos con distinta fuerza de asociación pero con la misma regresión  $\beta_1$

# Coeficiente de correlación de Pearson

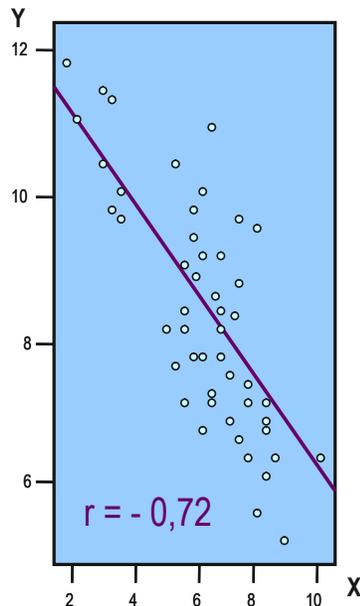


Alta fuerza de asociación  
Alto valor de  $r$



Menor fuerza de asociación  
Menor valor de  $r$

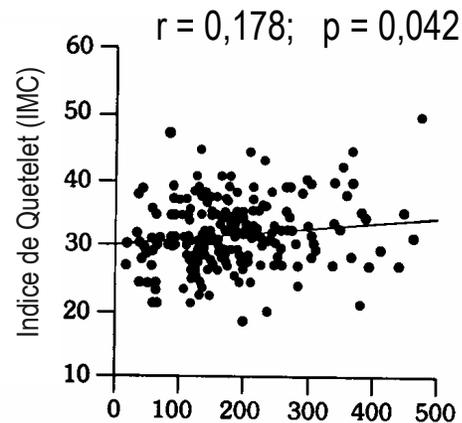
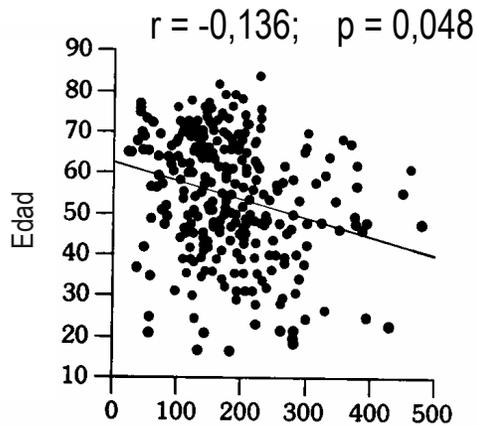
## Coeficiente de correlación de Pearson



- El valor absoluto de  $r$  es una medida de la fuerza de asociación entre dos variables cuantitativas “distintas”.
  - $r > 0,70$ : magnitud intensa
  - $0,3 < r < 0,7$ : magnitud moderada
  - $r < 0,3$ : magnitud débil
- Valores: -1 y 1
- Cálculo del tamaño muestral
- Mide la asociación pero NO confundir la asociación estadística con una relación CAUSA-EFECTO.

Valor de  $r$  indica FUERZA de asociación

## Coeficiente de correlación de Pearson



Bajo valor de  $r$  ➡ Baja fuerza de asociación

## Ejemplo del cálculo del tamaño muestral para el cálculo del coeficiente de correlación entre dos variables cuantitativas

Supongamos que se desea estudiar la asociación entre la edad y el nivel de colesterol entre los pacientes de un centro de salud. Se diseña un estudio piloto en el que se determinará mediante una analítica el colesterol en una muestra aleatoria de los pacientes durante un periodo de tiempo prefijado, de los que también se registrará su edad. Se obtiene un valor del coeficiente de correlación lineal de Pearson entre la edad y el colesterol de  $r=0,3$ . Aplicando las fórmulas, con un NIVEL DE CONFIANZA del 95% (error  $\alpha=5\%$ ) y un PODER ESTADÍSTICO del 80% (error  $\beta=20\%$ ), se obtiene:

$$n = \left( \frac{z_{1-\alpha/2} + z_{1-\beta}}{\frac{1}{2} \ln \left( \frac{1+r}{1-r} \right)} \right)^2 + 3 = \left( \frac{1,96 + 0,84}{\frac{1}{2} \ln \left( \frac{1+0,3}{1-0,3} \right)} \right)^2 + 3 \approx 85$$

En este caso, se necesitaría incluir a 85 pacientes para llevar a cabo el estudio.

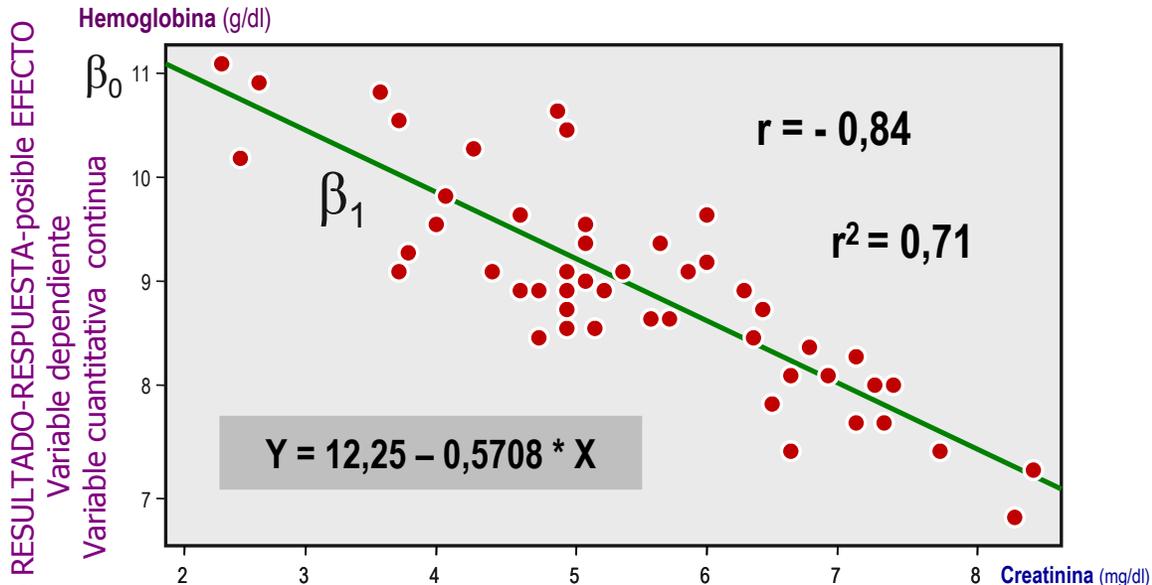
Si, además, en este último caso se prevé un 20% de posibles pérdidas de información durante la ejecución del estudio, el tamaño muestral se recalcula según la expresión:

$$n' = \frac{n}{1-L} = \frac{85}{1-0,2} = 106,25 \approx 107$$

Es decir, se necesitaría una muestra de 107 pacientes para llevar a cabo la investigación.

Pértegas Díaz S, Pita Fernández S. Determinación del tamaño muestral para calcular la significación del coeficiente de correlación lineal. Cad Aten Primaria 2001; 9: 209-211.

## Regresión lineal simple. Coeficiente de determinación



Representación gráfica de los niveles de creatinina y hemoglobina en 50 pacientes con fallo renal crónico

Variable PREDICTORA-EXPOSICIÓN-posible CAUSA  
Variable independiente, fija controlada por el investigador  
Variable cuantitativa continua

## Coeficiente de determinación $r^2$

- Indica el porcentaje de la variabilidad total de la variable dependiente o resultado (Y) que puede ser explicada por la variable independiente o predictora (X)

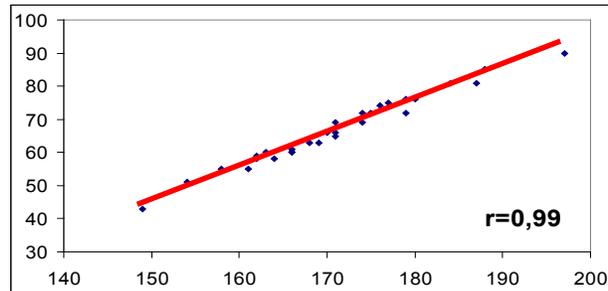
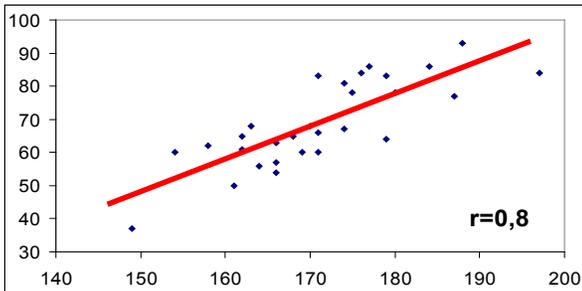
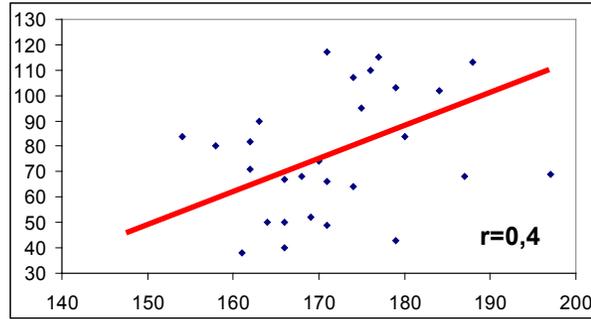
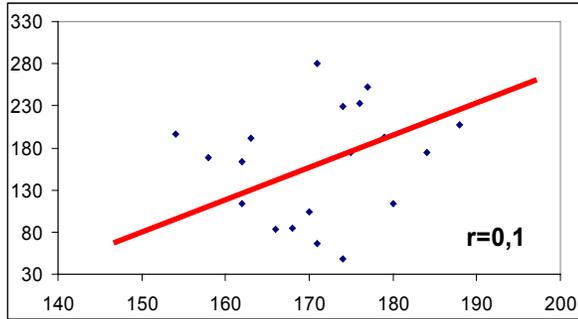
- Se puede afirmar que el 71% de la variabilidad de las Hb (variable resultado) de los 50 pacientes con IRC puede ser explicada por la distinta Cr de ellos ( $r^2 = 0,71$ ), resta por explicar un 29% que es achacable a otras variables predictoras distintas de la Cr.

- Valores: -1 y 1

$$R^2 = \frac{scE}{scG} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad \text{verifica} \quad 0 \leq R^2 \leq 1.$$

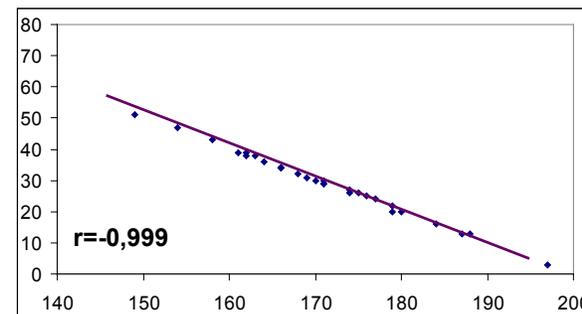
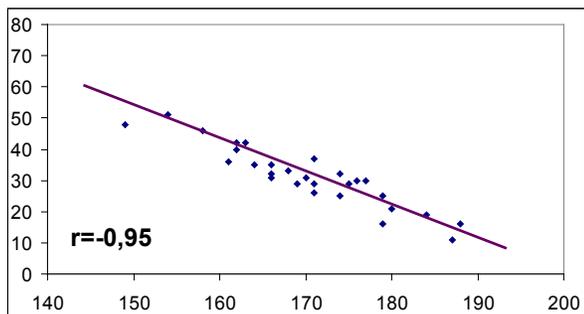
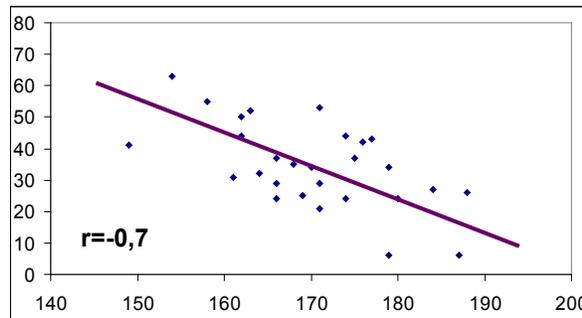
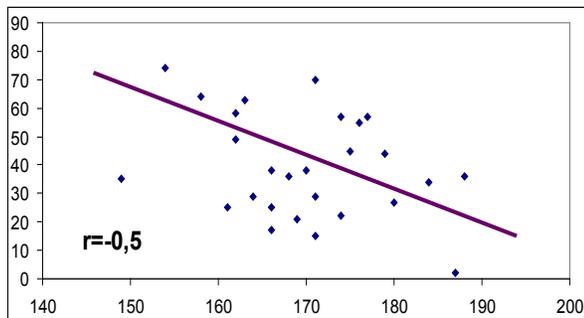
- El coeficiente de correlación  $r$  de Pearson decía que la magnitud de la asociación entre la Hb y la Cr es del 84% ( $r=0,84$ ) de la máxima posible.

## Tipos de correlaciones: positivas ( $r > 0$ )



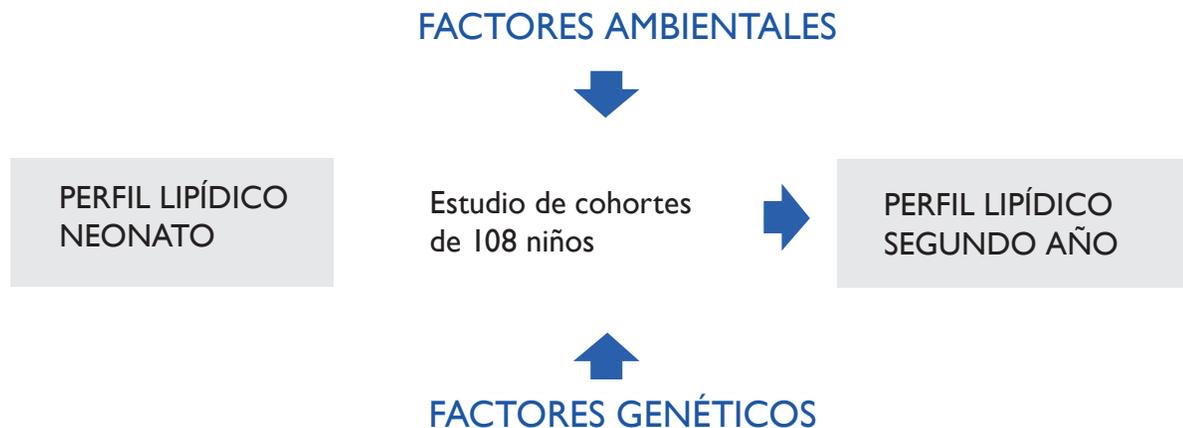
$$\beta_1 = \text{constante (+)}$$

## Tipos de correlaciones: negativas ( $r < 0$ )

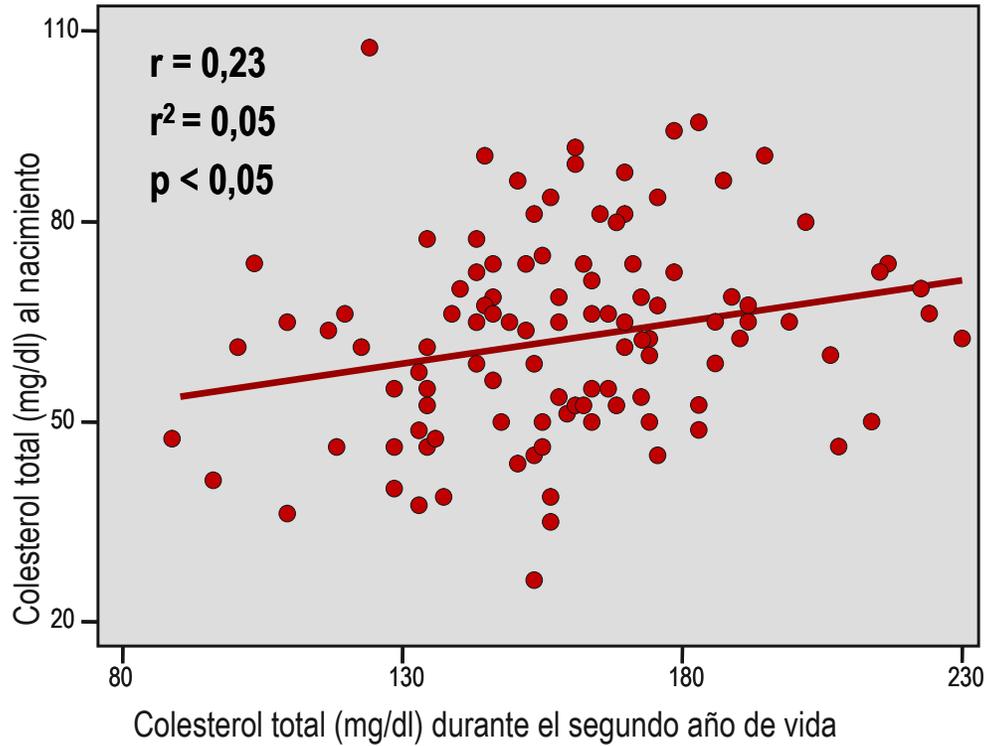


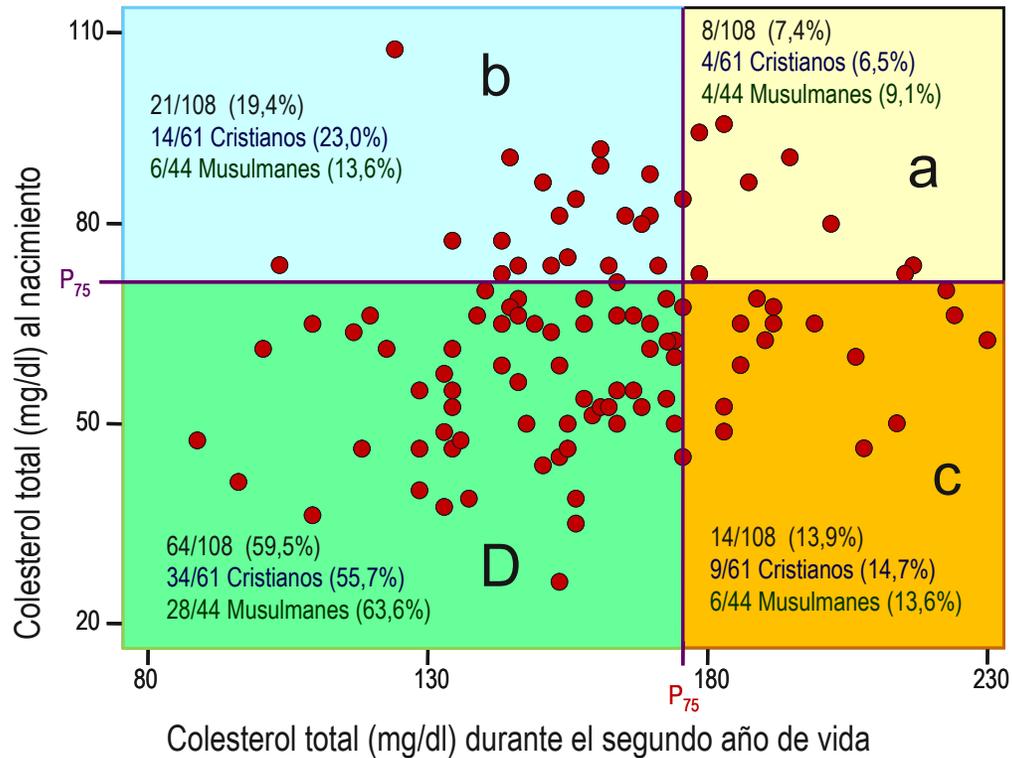
$\beta_1 = \text{constante (-)}$

## Fenómeno “Tracking”



Díaz Portillo J. “Influencia del grupo étnico en el patrón lipoproteico de una determinada población pediátrica de Ceuta”. Tesis doctoral (Noviembre 1997). Facultad de Medicina de la Universidad de Cádiz.





## Test de McNemar para datos apareados

(Prueba de la significación de los cambios)

McNemar	DESPUÉS del curso	
NEONATOS	CT > P <sub>75</sub>	CT ≤ P <sub>75</sub>
CT > P <sub>75</sub>	A	B
CT ≤ P <sub>75</sub>	C	D

$$\chi^2_{\text{McMemar}} = \frac{[(B - C) - 1]^2}{B + C}$$

$$Z_{\text{McNemar}} = \frac{(B - C) - 1}{\sqrt{B + C}}$$

## Test de McNemar para datos apareados

(Prueba de la significación de los cambios)

McNemar	DESPUÉS del curso	
NEONATOS	CT > P <sub>75</sub>	CT ≤ P <sub>75</sub>
CT > P <sub>75</sub>	8	21
CT ≤ P <sub>75</sub>	14	64

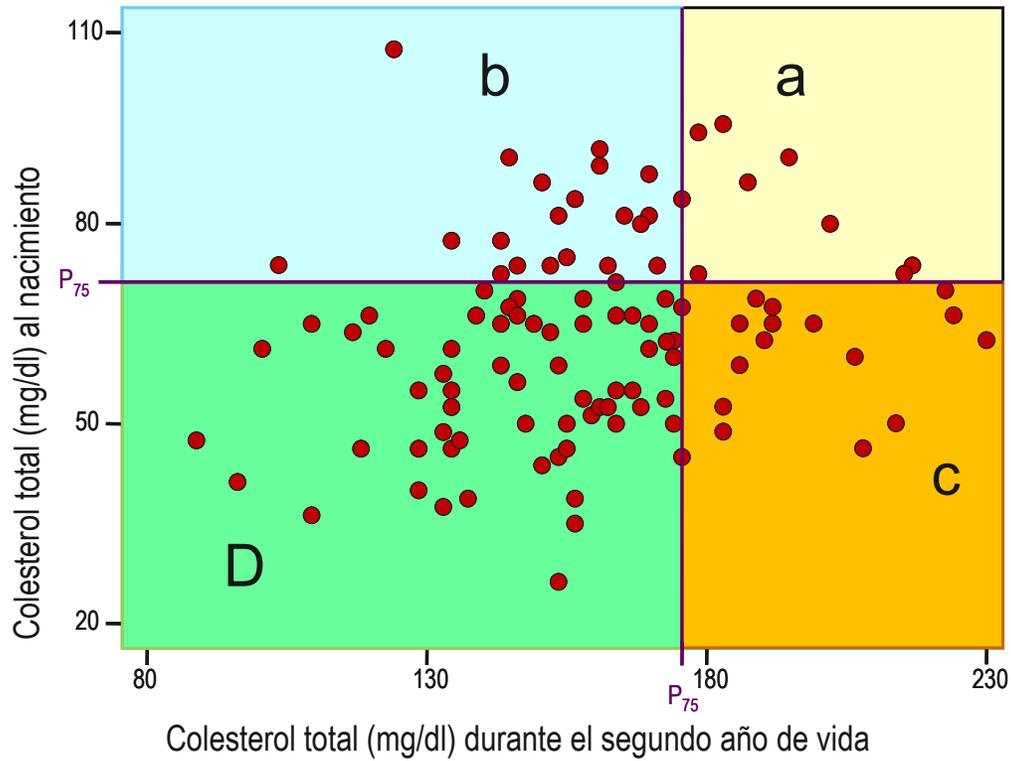
$$\chi^2_{\text{McMemar}} = \frac{[(21 - 14) - 1]^2}{21 + 14} = 1,03$$

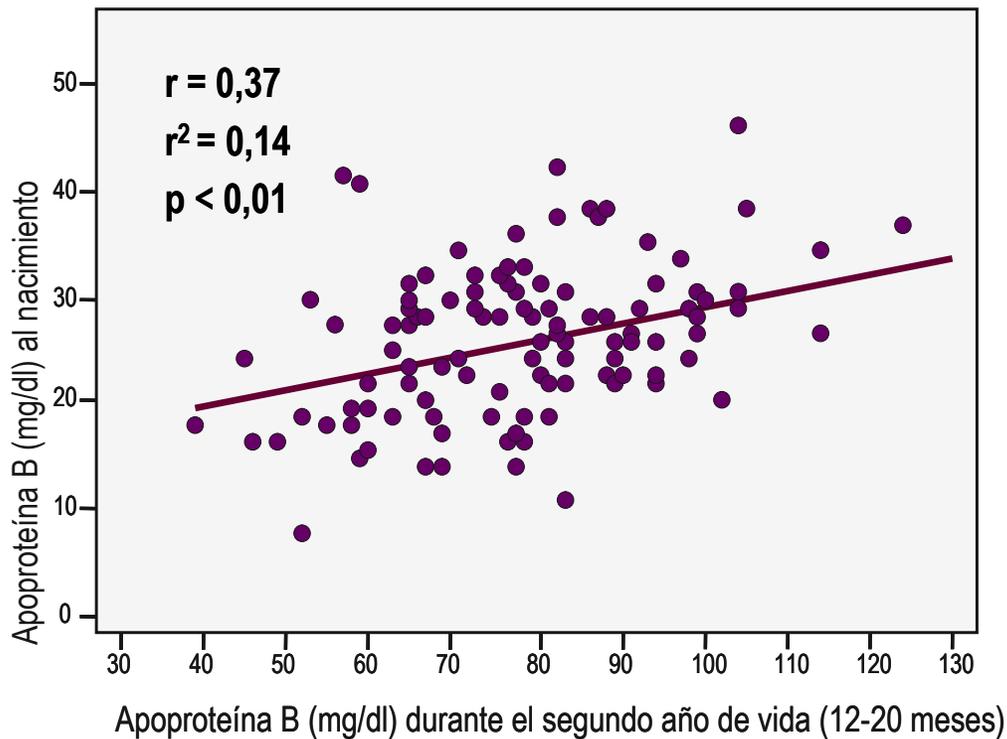
Z	Valor	$\alpha$
$Z_{95\%}$	1,96	0,05
$Z_{98\%}$	2,32	0,02
$Z_{99\%}$	2,56	0,01
$Z_{99,5\%}$	3,29	0,005

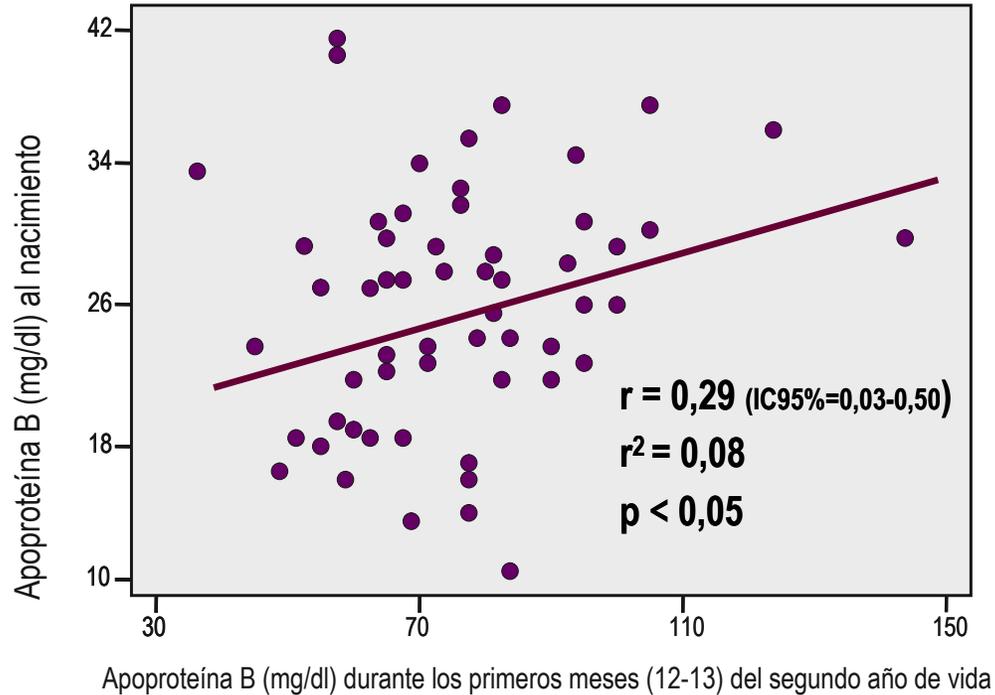
$$Z_{\text{McNemar}} = \frac{(B - C) - 1}{\sqrt{B + C}}$$

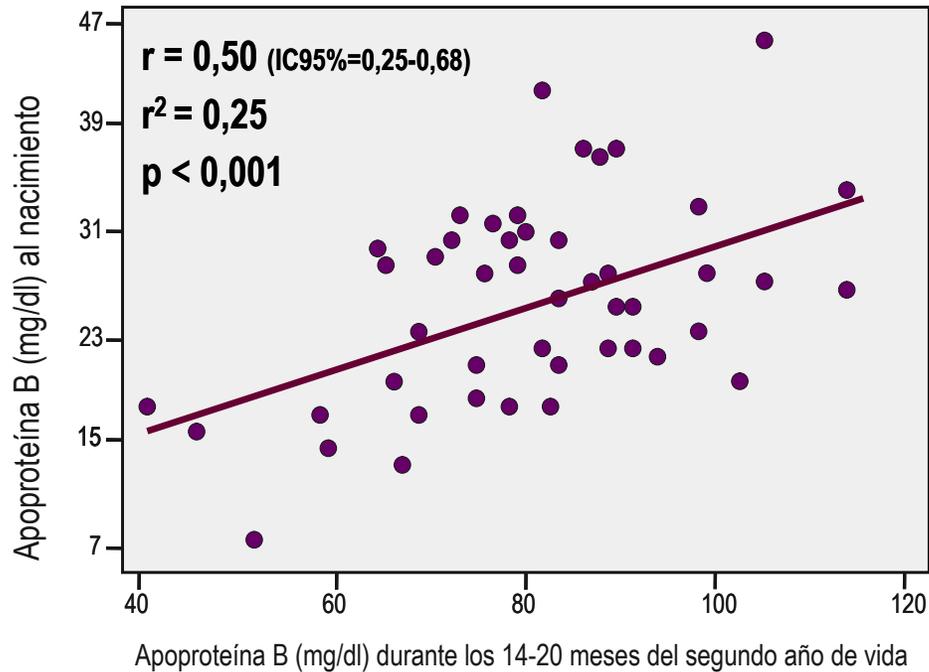
$$Z_{\text{McNemar}} = \frac{(21 - 14) - 1}{\sqrt{21 + 14}} = 1,014$$

**Conclusión:** Como el estadístico Z calculado 1,014 es INFERIOR a Z tabulado para un error alfa de 0,05 (1,96), podemos decir que las dos variables evaluadas son independientes, es decir NO están correlacionadas, y por tanto no hay cambios importantes en la distribución del colesterol total al pasar del nacimiento al segundo año de vida. Este comportamiento también se repite en los dos grupos culturales (cristianos y musulmanes) de Ceuta.



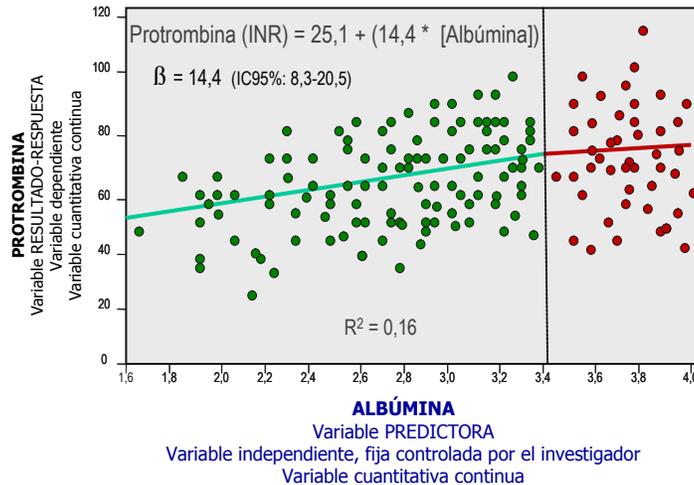






# Limitaciones del coeficiente de correlación de Pearson

Representación gráfica de la actividad de protrombina (INR) y la albúmina sérica en pacientes con hepatocarcinoma.



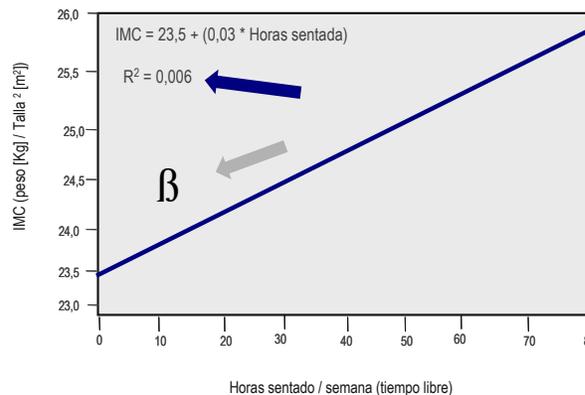
Sólo con la variable predictora incluida en el modelo (ALBÚMINA sérica) sólo se puede explicar el 16% de la variabilidad de la variable resultado (PROTROMBINA), queda por explicar un 84% de la variabilidad que es achacable a otras variables predictoras distintas no incluidas.

## Limitaciones del coeficiente de correlación de Pearson

El modelo (coeficiente  $\beta$ ) indica que por cada 10 horas que pasa sentada una mujer a la semana aumenta su peso en unos 800 gr.

Sólo con la variable predictora incluida en el modelo (horas sentado) sólo se puede explicar el 0,6% de la variabilidad de la variable resultado (IMC), queda por explicar un 99,4% de la variabilidad que es achacable a otras variables predictoras distintas no incluidas.

Variable RESULTADO-RESPUESTA-posible EFECTO  
Variable dependiente  
Variable cuantitativa continua

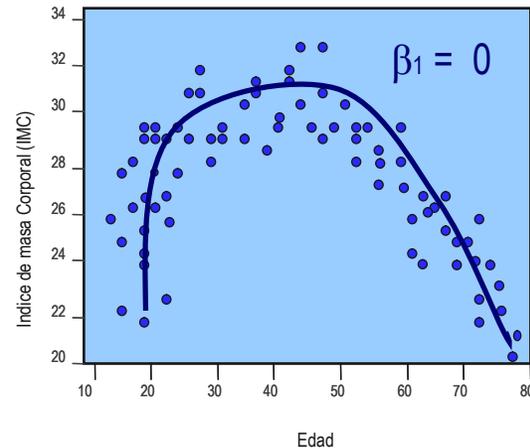
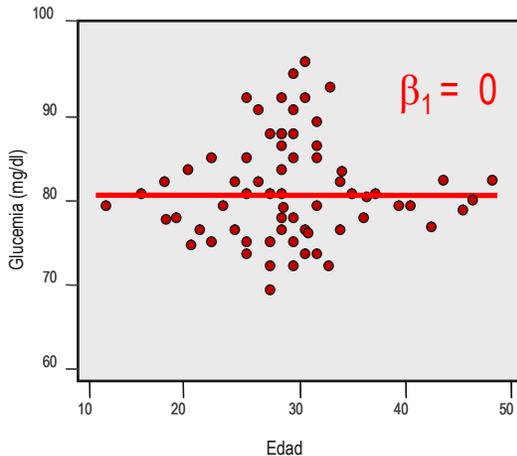


Relación entre el índice de masa corporal (IMC) y las horas de estar sentado a la semana en el tiempo libre.

Muestra representativa (n=7.926) mujeres de la Unión Europea

Variable PREDICTORA-EXPOSICIÓN-posible CAUSA  
Variable independiente, fija controlada por el investigador  
Variable cuantitativa continua

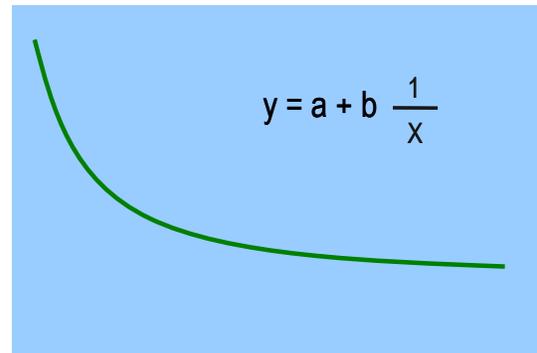
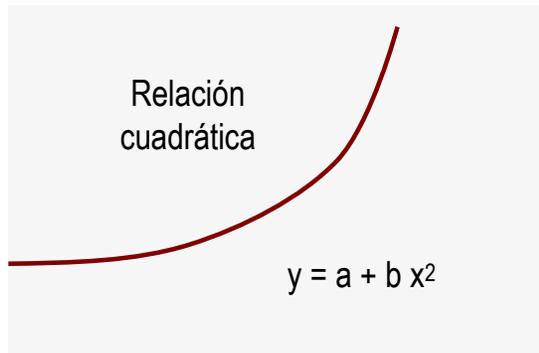
## Limitaciones del coeficiente de correlación de Pearson



Dos nubes de puntos con la misma pendiente  $\beta_1$  de valor cero que muestran dos situaciones muy diferentes. A la izquierda ausencia de relación. A la derecha relación NO lineal.

NO se puede emplear el coeficiente de correlación de Pearson

## Limitaciones del coeficiente de correlación de Pearson



NO se puede emplear el coeficiente de correlación de Pearson

Podemos encontrar cualquier tipo de función matemática con claro apartamiento de la linealidad.

## Otros coeficientes de correlación

- Cuando las variables en vez de ser numéricas son ordinales, ¿es posible preguntarse sobre si hay algún tipo de correlación entre ellas?
- Disponemos para estos casos de tres test estadísticos:
  - $\rho$  ('rho') de Spearman
  - $\tau$  ('tau') de Kendall
  - Coeficiente de asociación ordinal de Goodman-Kruscall (coeficiente gamma)
- Son estadísticos análogos al coeficiente de correlación lineal de Pearson  $r$ , que aparecen en estudios donde las variables cuantitativas no son continuas y/o no siguen una distribución normal.
- El coeficiente de correlación de Spearman es recomendable utilizarlo cuando los datos presentan valores externos ya que dichos valores afectan mucho el coeficiente de correlación de Pearson, o ante distribuciones NO normales.

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

## Coefficientes de correlación

### PEARSON

- Ambas variables deben seguir la distribución NORMAL
- Sólo detecta asociaciones lineales
- Las observaciones han de ser independientes

### SPEARMAN

- No requiere supuestos previos de normalidad
- Detecta asociaciones distintas de la lineal
- Es un método no paramétrico pero permite la construcción de intervalos de confianza
- Estima el grado de asociación de cualquier tipo con independencia de la función matemática que relaciona ambas variables

## Pregunta de investigación

¿Existe relación entre la edad del paciente y la duración de la estancia en UCI?

Variable PREDICTORA I  
CUANTITATIVA continua

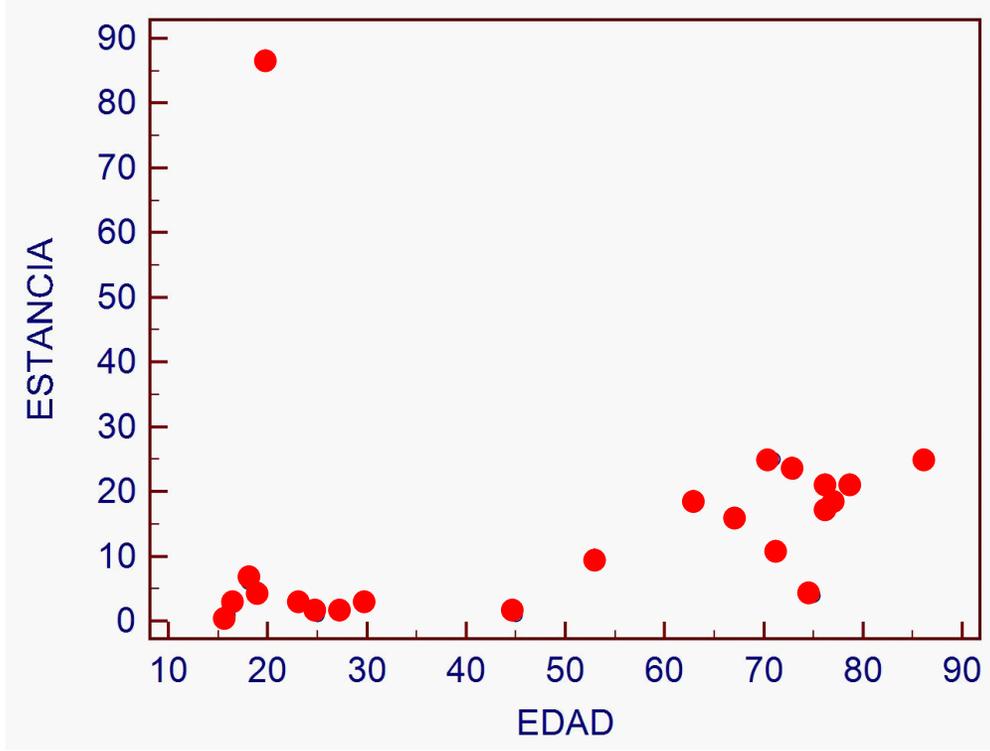
Edad  
- Años

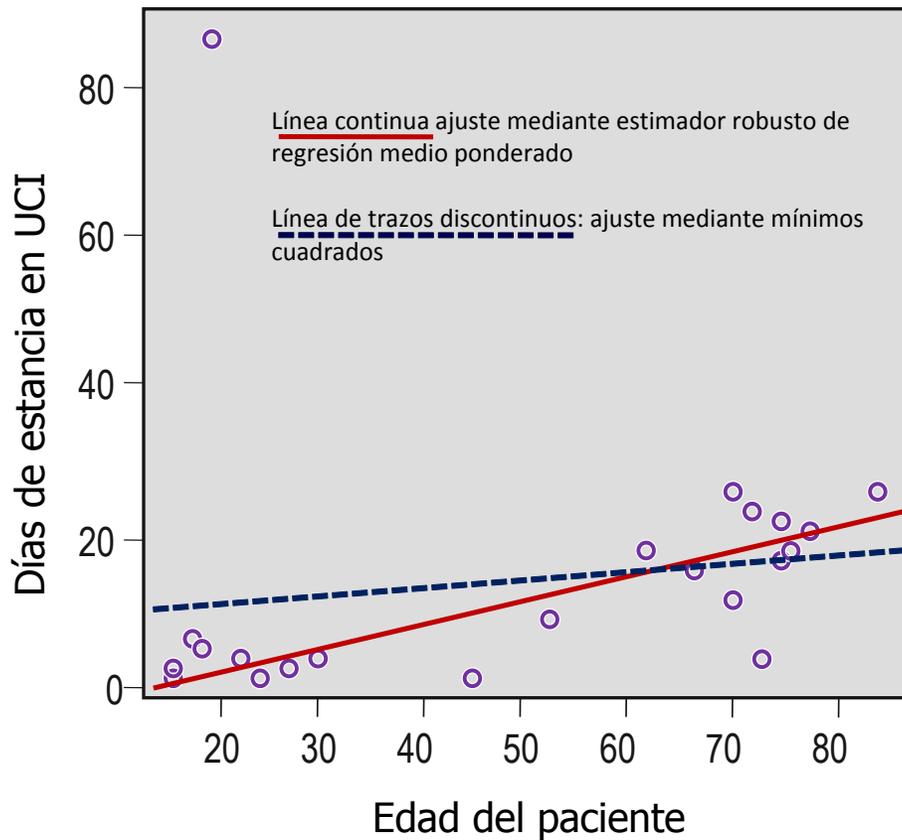


Variable RESULTADO  
CUANTITATIVA discreta

Estancia  
- Días

¿Cuál es el test estadístico más acertado?





## Resultados

### Coeficiente de correlación de Spearman

Variable	N	Correlación (r)	Significación
Edad del paciente versus Estancia en UCI	22	0,614 (**)	0,002

### Tau-b de Kendall

Variable	N	Correlación (r)	Significación
Edad del paciente versus Estancia en UCI	22	0,470 (**)	0,003

\*\* La correlación es significativa al nivel 0,01 (bilateral).

$p < 0,01$   Rechazamos la hipótesis nula

## Pregunta de investigación

¿Existe relación entre la intensidad del estudio y el rendimiento académico en el examen del curso de Bioestadística?

Variable PREDICTORA I  
Cuantitativa continua

Horas de estudio



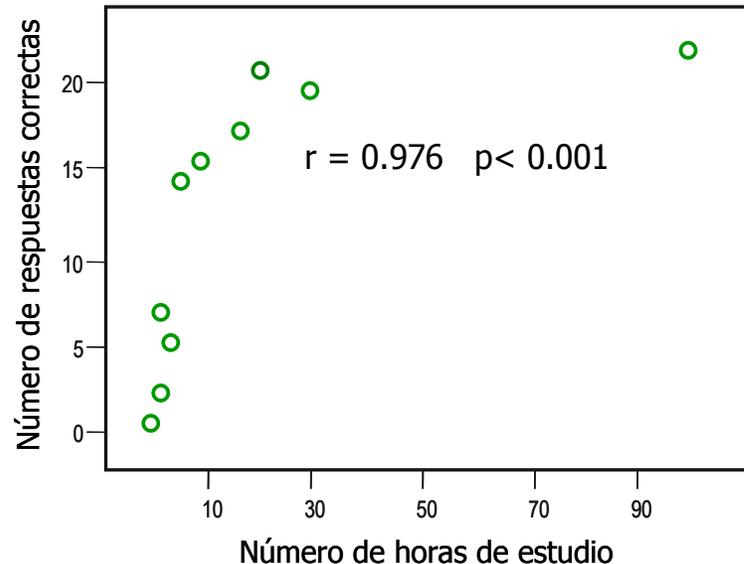
Variable RESULTADO  
Cuantitativa discreta

Rendimiento académico

¿Cuál es el test estadístico más acertado?

Pregunta de investigación: ¿Existe relación entre la intensidad del estudio y el rendimiento académico en el examen del curso de Bioestadística?

Nº del Alumno	Horas de estudio ( $X_1$ )	Nº respuestas correctas (Y)
1ª	5	6
2ª	9	16
3ª	17	18
4ª	1	1
5ª	2	3
6ª	21	21
7ª	3	7
8ª	29	20
9ª	7	15
10ª	100	22



¿Cuál es el test estadístico más acertado?

## Resultados

### Coeficiente de correlación de Spearman

Variable	N	Correlación (r)	Significación
Horas de estudio versus Respuestas correctas	10	0,976 (**)	0,000

### Tau-b de Kendall

Variable	N	Correlación (r)	Significación
Horas de estudio versus Respuestas correctas	10	0,911 (**)	0,000

\*\* La correlación es significativa al nivel 0,01 (bilateral).

$p < 0,01$   Rechazamos la hipótesis nula

## Mito I: “La sonrisa de la esfinge”

He observado que el individuo nº 20 se aparta bastante de la tendencia del grupo. De hecho si se eliminara se podría considerar una supuesta relación lineal.

¿Estás seguro que ese dato nº 10 es correcto? ¿Está bien tomado? ¿Cómo estás seguro de que ese individuo ha estudiado 100 horas? ¿Qué medidas has tomado para considerar la credibilidad de la variable predictora (horas de estudio)?

¿No podría considerarse a ese individuo nº 10 como un periférico o OUTLIER?

## Pregunta de investigación

¿Existe relación entre el coeficiente intelectual de los alumnos y el rendimiento académico en el examen del curso de Bioestadística?

Variable PREDICTORA 2  
Cuantitativa discreta

Coeficiente intelectual



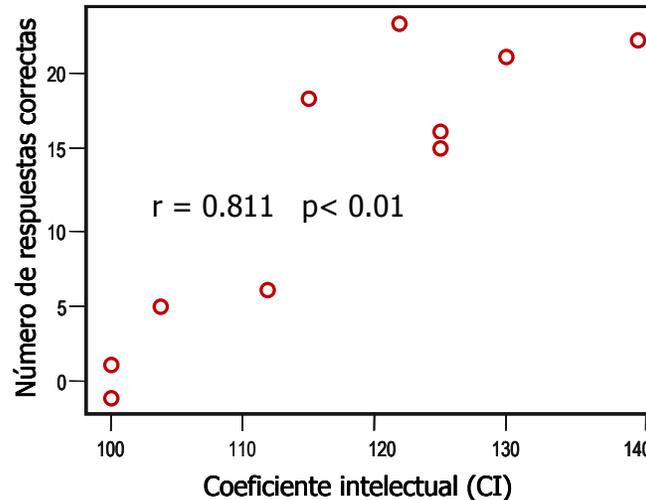
Variable RESULTADO  
Cuantitativa discreta

Rendimiento académico

¿Cuál es el test estadístico más acertado?

Pregunta de investigación: ¿Existe relación entre el coeficiente intelectual del alumno y el rendimiento académico en el examen del curso de Bioestadística?

Nº del Alumno	Horas de estudio ( $X_2$ )	Nº respuestas correctas (Y)
1º	104	6
2º	125	16
3º	115	18
4º	100	1
5º	100	3
6º	140	21
7º	112	7
8º	130	20
9º	125	15
10º	122	22



¿Cuál es el test estadístico más acertado?

## Resultados

### Coeficiente de correlación de Spearman

Variable	N	Correlación (r)	Significación
Coeficiente intelectual versus respuestas correctas	10	0,811 (**)	0,004

### Tau-b de Kendall

Variable	N	Correlación (r)	Significación
Coeficiente intelectual versus respuestas correctas	10	0,705 (**)	0,005

\*\* La correlación es significativa al nivel 0,01 (bilateral).

$p < 0,01$   Rechazamos la hipótesis nula

## Pregunta de investigación

¿Existe relación entre las horas de asistencia a clase y el rendimiento académico en el examen del curso de Bioestadística?

Variable PREDICTORA 3  
Cuantitativa continua

Horas de asistencia a clases



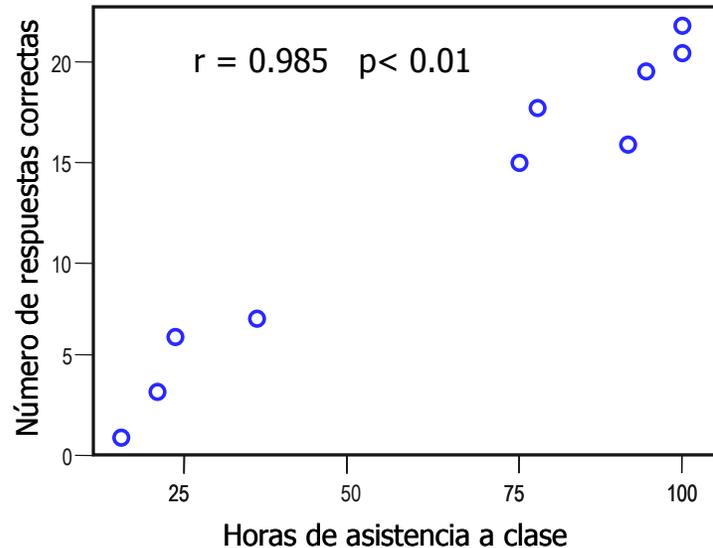
Variable RESULTADO  
Cuantitativa discreta

Rendimiento académico

¿Cuál es el test estadístico más acertado?

Pregunta de investigación: ¿Existe relación entre las horas de asistencia a clase y el rendimiento académico en el examen del curso de Bioestadística?

Nº del Alumno	Horas de estudio ( $X_3$ )	Nº respuestas correctas (Y)
1º	23	6
2º	92	16
3º	78	18
4º	15	1
5º	20	3
6º	100	21
7º	35	7
8º	95	20
9º	75	15
10º	100	22



¿Cuál es el test estadístico más acertado?

## Resultados

### Coeficiente de correlación de Spearman

Variable	N	Correlación (r)	Significación
Asistencia a clase versus respuestas correctas	10	0,985 (**)	0,000

### Tau-b de Kendall

Variable	N	Correlación (r)	Significación
Asistencia a clase versus respuestas correctas	10	0,944 (**)	0,000

\*\* La correlación es significativa al nivel 0,01 (bilateral).

$p < 0,01$   Rechazamos la hipótesis nula

## Resumen de los resultados de correlación bivariante

Variables predictoras	Rho Spearman	Significación	Tau de Kendall	Significación
Horas de estudio	0,976	0,000	0,911	0,000
Coefficiente intelectual	0,811	0,004	0,705	0,005
Asistencia a clase	0,985	0,000	0,944	0,000

Variable resultado: Rendimiento académico (número de respuestas correctas)

¿Existe relación entre las distintas variables predictoras?

¿COLINEALIDAD?

## Pregunta de investigación

¿Existe relación entre las horas dedicadas al estudio por el alumno y el coeficiente intelectual de los alumnos del curso de Bioestadística?

Variable PREDICTORA 1  
Cuantitativa continua

Horas de estudio



Variable PREDICTORA 2  
Cuantitativa discreta

Coeficiente intelectual

¿Cuál es el test estadístico más acertado?

## Resultados

### Coeficiente de correlación de Spearman

Variable	N	Correlación (r)	Significación
Horas de estudio versus coeficiente intelectual	10	0,787 (**)	0,007

\*\* La correlación es significativa al nivel 0,01 (bilateral).

### Tau-b de Kendall

Variable	N	Correlación (r)	Significación
Horas de estudio versus coeficiente intelectual	10	0,614 (*)	0,015

\* La correlación es significativa al nivel 0,05 (bilateral).

$p < 0,05$   Rechazamos la hipótesis nula

## Pregunta de investigación

¿Existe relación entre las horas de estudio del alumno y las horas de asistencia a clase del curso de Bioestadística?

Variable PREDICTORA 1  
Cuantitativa continua

Horas de estudio



Variable PREDICTORA 3  
Cuantitativa continua

Horas de asistencia a clases

¿Cuál es el test estadístico más acertado?

## Resultados

### Coeficiente de correlación de Spearman

Variable	N	Correlación (r)	Significación
Horas de estudio versus asistencia a clase	10	0,954 (**)	0,000

\*\* La correlación es significativa al nivel 0,01 (bilateral).

### Tau-b de Kendall

Variable	N	Correlación (r)	Significación
Horas de estudio versus asistencia a clase	10	0,854 (**)	0,001

\*\* La correlación es significativa al nivel 0,01 (bilateral).

$p < 0,05$   Rechazamos la hipótesis nula

## Pregunta de investigación

¿Existe relación entre el coeficiente intelectual del alumno y las horas de asistencia a clase del curso de Bioestadística?

Variable PREDICTORA 2  
Cuantitativa discreta

Coeficiente intelectual



Variable PREDICTORA 3  
Cuantitativa continua

Horas de asistencia a clases

¿Cuál es el test estadístico más acertado?

## Resultados

### Coeficiente de correlación de Spearman

Variable	N	Correlación (r)	Significación
Coeficiente intelectual versus asistencia a clase	10	0,869 (**)	0,001

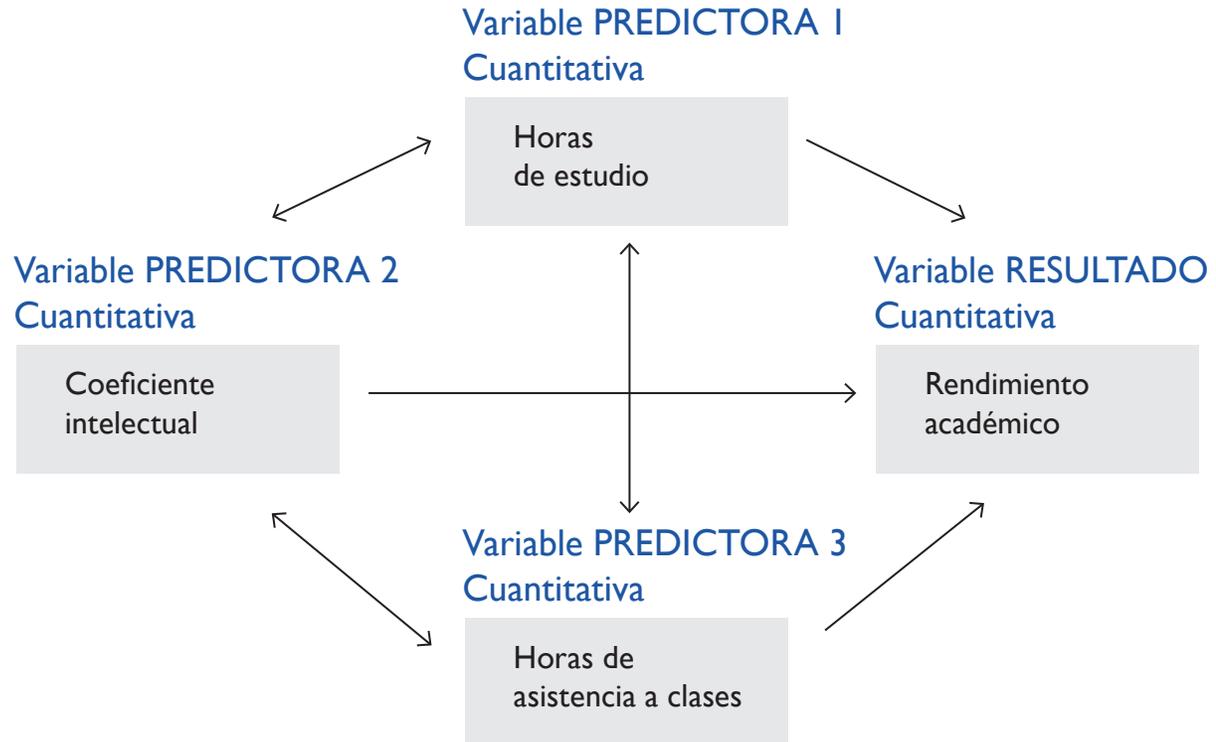
\*\* La correlación es significativa al nivel 0,01 (bilateral).

### Tau-b de Kendall

Variable	N	Correlación (r)	Significación
Coeficiente intelectual versus asistencia a clase	10	0,782 (**)	0,002

\*\* La correlación es significativa al nivel 0,01 (bilateral).

$p < 0,05$   Rechazamos la hipótesis nula



Rev Lab Clin. 2009;2(3):115–123



## Revista del Laboratorio Clínico

[www.elsevier.es/LabClin](http://www.elsevier.es/LabClin)



ORIGINAL

### Determinación de plomo y cadmio en sangre y su relación con fuentes de exposición. Estudio PESA, 2008

Montserrat González-Estecha<sup>a,\*</sup>, Elena Trasobares Iglesias<sup>a</sup>, Sara Cano Escudero<sup>b</sup>, Pilar Oliván Osambela<sup>a</sup>, Manuel Fuentes Ferrer<sup>b</sup>, Cristina Fernández Pérez<sup>b</sup>, María José Martínez García<sup>c</sup>, Ángel Luis García González<sup>c</sup>, María Jesús Gaspar Blázquez<sup>d</sup>, Joaquín González Revaldería<sup>d</sup>, María Carmen Barciela Alonso<sup>e</sup>, Paloma Herbello Hermelo<sup>e</sup>, Pilar Bermejo Barrera<sup>e</sup>, José Jesús Guillén Pérez<sup>c</sup>, Elena Miravalles González<sup>d</sup>, Manuel Arroyo Fernández<sup>a</sup>, en nombre del grupo PESA <sup>♦</sup>

¿Qué relación existe entre el consumo de cigarrillos y los niveles de cadmio en sangre?

### Cuantitativa discreta

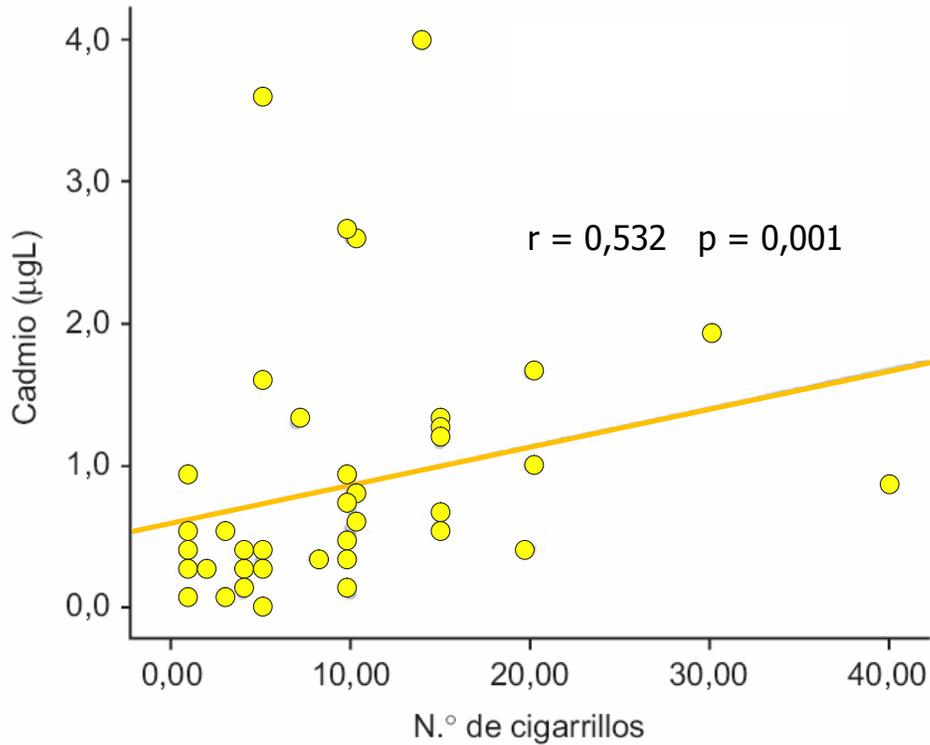
Consumo de cigarrillos  
Número de cigarrillos  
consumidos por día



### Cuantitativa continua

Cadmio en sangre  
Niveles séricos de cadmio  
( $\mu\text{g/L}$ )

¿Cuál es el test estadístico más acertado?



**Figura 3** Correlación entre la concentración de cadmio en sangre ( $\mu\text{g/l}$ ) y el número de cigarrillos.

## ***ORIGINAL***

# **Precisión diagnóstica de la calprotectina fecal para predecir una colonoscopia patológica**

María del Valle García Sánchez<sup>a</sup>, Raúl González<sup>b</sup>, Eva Iglesias Flores<sup>a</sup>, Federico Gómez Camacho<sup>a</sup>, Luis Casais Juanena<sup>a</sup>, Antonio Cerezo Ruiz<sup>a</sup>, Manuel Montero Pérez-Barquero<sup>c</sup>, Jordi Muntané<sup>b</sup> y Juan Francisco de Dios Vega<sup>a</sup>

<sup>a</sup>Unidad Clínica Aparato Digestivo. Hospital Universitario Reina Sofía. Córdoba.

<sup>b</sup>Unidad de Investigación. Hospital Universitario Reina Sofía. Córdoba.

<sup>c</sup>Medicina Interna. Hospital Universitario Reina Sofía. Córdoba. España.

**Med Clin (Barc) 2006;127:41-6**

**FUNDAMENTO Y OBJETIVO:** El dolor abdominal y la diarrea son síntomas comunes en la población general que no permiten discriminar enfermedad intestinal orgánica. La colonoscopia es la prueba de referencia, pero es cara, invasiva, no puede repetirse frecuentemente y supone una gran carga asistencial. La calprotectina fecal (CPF) es un marcador fácil y rápido de medir, no invasivo y barato, que podría detectar la enfermedad orgánica en la zona del colon. El objetivo de nuestro estudio es valorar la utilidad de este marcador para predecir una colonoscopia patológica y correlacionar sus valores con el grado de actividad en la enfermedad inflamatoria intestinal (EII).

**PACIENTES Y MÉTODO:** Se determinó la calprotectina en muestras fecales (mediante enzimoanálisis) en 190 personas sometidas a una colonoscopia. Los diagnósticos endoscópicos fueron: 117 normales, 28 pólipos colónicos, 20 cáncer colorrectal (CCR) y 25 EII. El análisis de los datos se realizó con el programa SPSS 11.0.

**RESULTADOS:** La media (DE) de CPF estaba aumentada en CCR (726,6 [533] mg/kg) y EII (2.171,1 [2.133,6] mg/kg). No hubo diferencias en pacientes con pólipos (158,3 [15,8] mg/kg). En pacientes con EII, los valores se correlacionaron con el grado de actividad de la enfermedad. El valor de corte fue 217 mg/kg, y se obtuvo una sensibilidad de la CPF del 85% y un valor predictivo negativo del 93%. En el grupo control, la única variable que se relacionó con un aumento de CPF fue la medicación previa con ácido acetilsalicílico (AAS) y/o antiinflamatorios no esteroideos (AINE).

**CONCLUSIONES:** La CPF es una prueba no invasiva, barata y sensible para detectar enfermedad orgánica en la zona del colon, que puede ser útil para seleccionar a pacientes que deben someterse a una colonoscopia. Se correlaciona con el grado de actividad en la EII. La toma de AAS y/o AINE podría aumentar la tasa de falsos positivos.

*Palabras clave:* Calprotectina fecal. Colonoscopia. Inflamación intestinal.

## Pregunta de investigación

¿Qué relación y que grado de asociación existe entre el índice de actividad clínica de las enfermedades inflamatorias intestinales (EII) y los niveles de calprotectina fecal?

Variable PREDICTORA  
Cuantitativa continua

Calprotectina fecal



Variable RESULTADO  
Cuantitativa discreta

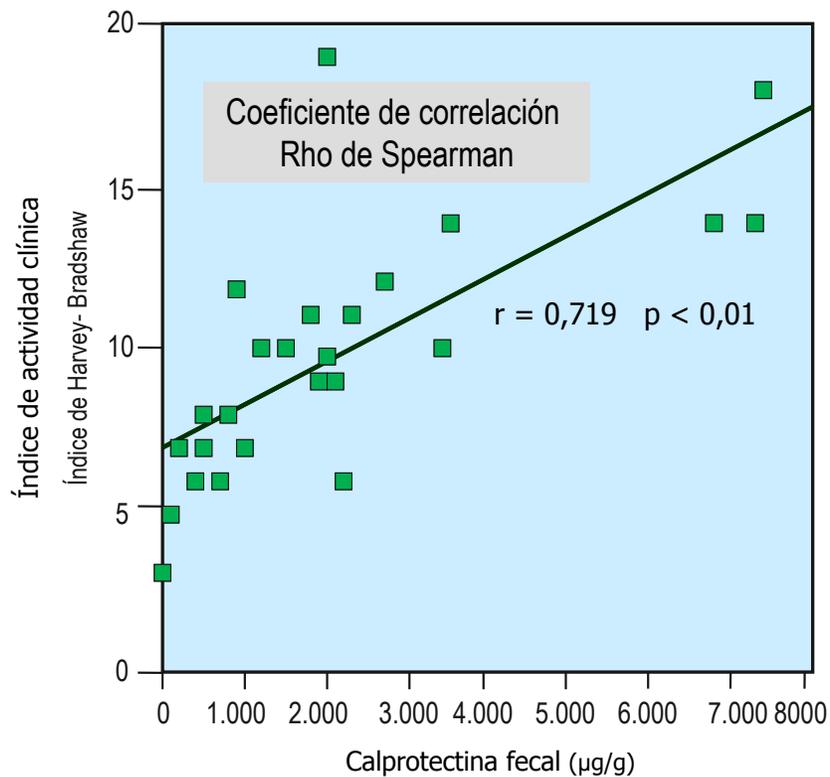
Índice de actividad de la enfermedad inflamatoria intestinal (Harvey-Bradshaw modificado para colitis ulcerosa y enfermedad de Crohn)

## Índice de Harvey-Bradshaw

<i>Variable</i>	<i>Puntuación</i>	
A. Número diario de deposiciones	1 por deposición	
B. Dolor abdominal	Ninguno	0
	Suave	1
	Moderado	2
	Muy intenso	3
C. Estado general	Bueno	0
	Regular	1
	Malo	2
	Muy malo	3
	Terrible	4
D. Masa abdominal	Ausente	0
	Dudosa	1
	Definida	2
	Definida y dolorosa	3
E. Manifestaciones extradigestivas	1 por manifestación	

*Índice: A + B + C + D + E; brote leve: 0-5; brote moderado: 6-10; y brote grave: > 10.*

¿Qué tipo de variable es el índice de Harvey-Bradshaw?



¿Cuál es la prueba estadística más adecuada?

## Análisis estadístico (fragmento del texto)

El análisis de los datos se llevó a cabo mediante el programa estadístico SPSS 11.0. El valor de la CPF se presenta como **media** y con un intervalo de confianza (IC) del 95%. Se utilizó el test de **Kruskal-Wallis** para evaluar las diferencias entre los valores de CPF entre los diferentes grupos y el **coeficiente de Spearman** para valorar la correlación entre este marcador y el índice de actividad de la enfermedad en el grupo de pacientes con enfermedad inflamatoria intestinal (EII).

## Resultados (fragmento del texto)

“En la figura 2 se muestra la correlación positiva entre la concentración de la CPF y el índice de actividad de la enfermedad en pacientes con enfermedades inflamatoria intestinales (EII) (enfermedad de Crohn (EC) y colitis ulcerosa (CU)) ( $r = 0,719$ ,  $p < 0,01$ )”.

## Discusión (fragmento del texto)

“El presente trabajo muestra que la concentración de CPF está significativamente elevada en pacientes con procesos orgánicos en colon, fundamentalmente en el cáncer colorrectal y enfermedad inflamatoria intestinal (EII). Además, los valores más altos se alcanzan en pacientes con EII, y sus valores se correlacionan con el grado de actividad de la enfermedad”.

## Pregunta de investigación

¿Qué relación y qué grado de asociación existe entre el nivel de conocimientos médicos actualizados sobre el mejor tratamiento posible y el tiempo transcurridos desde la graduación?

Variable PREDICTORA  
Cuantitativa continua

Años desde la graduación

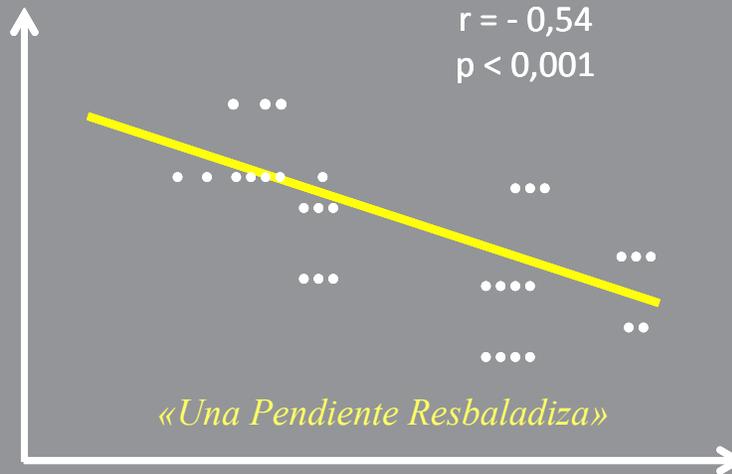


Variable RESULTADO  
Cuantitativa discreta

Conocimientos actualizados  
sobre el mejor tratamiento posible  
(medidos según escala de MBE)

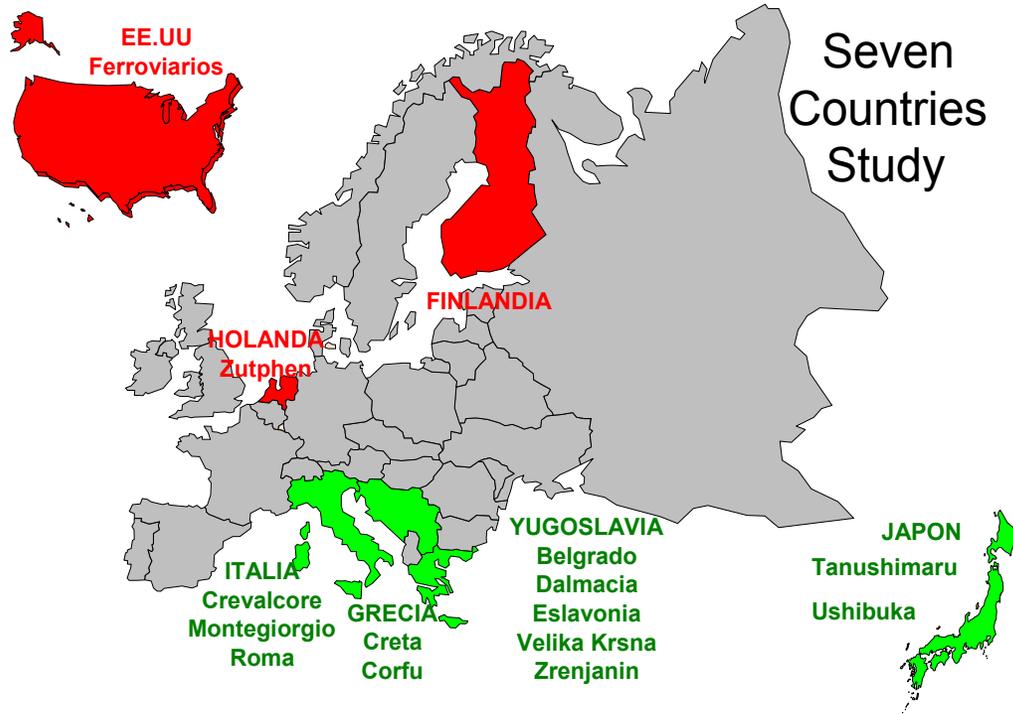
## Coefficiente de correlación Rho de Spearman

Conocimientos actualizados sobre el mejor  
tratamiento posible (niveles de evidencia)



Años desde la graduación

# Estudios ecológicos



## Estudio ecológico

Variable PREDICTORA  
Factor de exposición (F)

Grasas saturadas

Ácidos grasos saturados  
de la dieta (% de aporte  
calórico)



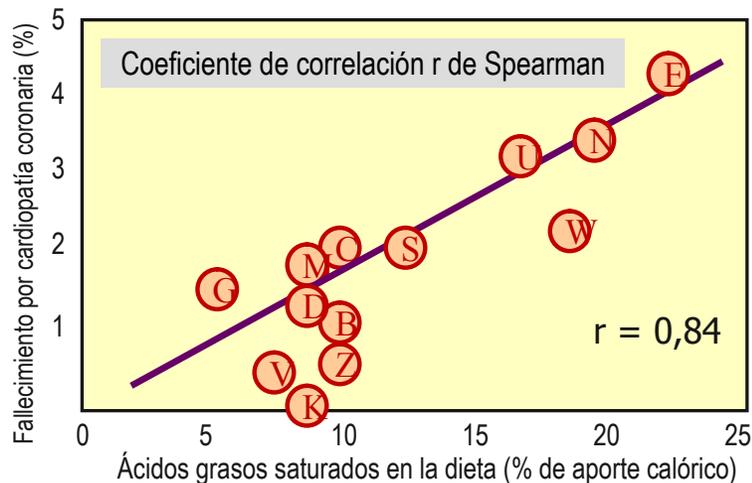
Variable RESULTADO  
Efecto o Enfermedad (E)

Cardiopatía coronaria

Tasas de mortalidad por  
cardiopatía coronaria

¿Cuál es el test estadístico más acertado?

## Estudio ecológico



Tasa de mortalidad por cardiopatía coronaria en función del contenido en ácidos grasos saturados de la dieta (% del aporte calórico). Ambas variables están significativamente correlacionadas según un amplio estudio (Seven Countries Study) llevado a cabo en diferentes grupos de población. B = Belgrado, C= Crevalcore, D= Dalmacia, E = este de Finlandia, G = Corfú, K = Creta, N = Zutphen, M = Montegiorgio, S = Eslavonia, T = Tanushimaru, U = empleados de ferrocarriles de EEUU, V = Velika Krsna, W = oeste de Finlandia, Z = Zrenjarin.

## Estudio ecológico

Variable PREDICTORA  
Factor de exposición (F)

Grasas saturadas

Ácidos grasos saturados  
de la dieta (% de aporte  
calórico)



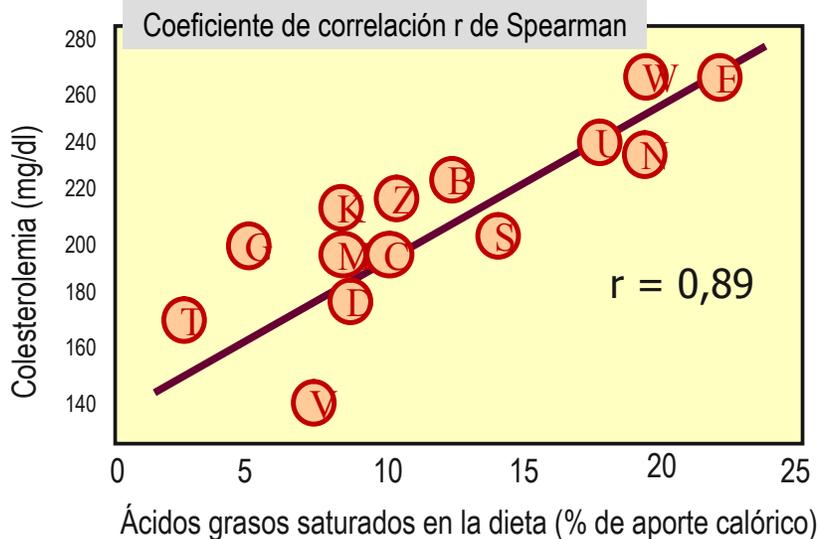
Variable RESULTADO  
Efecto o Enfermedad (E)

Colesterolemia

Tasas de colesterol en  
sangre (mg/dl)

¿Cuál es el test estadístico más acertado?

## Estudio ecológico



Valores de colesterol en sangre en relación con el contenido en ácidos grasos saturados de la dieta (% del aporte calórico). Ambas variables están significativamente correlacionadas según un amplio estudio (Seven Countries Study) llevado a cabo en diferentes grupos de población. B = Belgrado, C= Crevalcore, D= Dalmacia, E = este de Finlandia, G = Corfú, K = Creta, N = Zutphen, M = Montegiorgio, S = Eslavonia, T = Tanushimaru, U = empleados de ferrocarriles de EEUU, V = Velika Krsna, W = oeste de Finlandia, Z = Zrenjarin.

## Capítulo 8

Publicación Bioestadística: Análisis bivariante.  
Test de contraste de hipótesis.  
Cualitativa dicotómica versus cuantitativa.

Dr. Jacobo Díaz Portillo  
Coordinador de Formación Continuada  
Hospital Universitario de Ceuta

# Materiales y métodos

## Estructura

1. Tipo de diseño (estudio).
2. Sujetos incluidos (criterios de selección de la muestra)
3. Fuentes de información.
4. Descripción de la intervención y definición del punto final de la investigación (estudios experimentales).
5. Instrumentos y procedimientos de medida (cuestionarios).
6. Definición de las variables y covariables incluidas en el estudio (factor de estudio y criterio de evaluación).
7. Limitaciones y consideraciones éticas de la investigación.
8. Pruebas estadísticas a utilizar (programa estadístico).

## Fases del análisis estadístico

1. ANÁLISIS DESCRIPTIVO



2. ANÁLISIS BIVARIANTE



3. ANÁLISIS MULTIVARIANTE

## Decisiones que se pueden tomar en un test de contraste de hipótesis después de conocer el valor de significación estadística $p$

$p < 0,05$ $A \neq B$	$p \geq 0,10$ $A = B$
Se rechaza la hipótesis nula	No se puede rechazar la hipótesis nula
No parece que el azar pueda explicarlo todo	No se puede descartar que el azar lo explique todo
El EFECTO observado es mayor que el ERROR	El EFECTO observado es similar que el ERROR
Hay diferencias estadísticamente significativas	NO hay diferencias estadísticamente significativas
Existen evidencias a favor de la hipótesis alternativa $H_a$ , y por tanto rechazamos la hipótesis nula $H_0$	NO existen evidencias a favor de la hipótesis alternativa $H_a$ , y por tanto aceptamos la hipótesis nula $H_0$

Los límites de 0,05 y 0,1 pueden ser considerados en cierto modo arbitrarios y aproximados

## Análisis estadístico

Tipos de test analíticos estadísticos de contraste de hipótesis según el tipo de variables implicadas en el estudio

Variable dependiente

Variable resultado

		Dicotómica	Continua
Variable Predictora	Dicotómica	Test de $\chi^2$	t de Student
Variable Independiente	Continua	Regresión logística	Regresión lineal

## Análisis estadístico

ANÁLISIS ESTADÍSTICO		Variable resultado	
		Dicotómica	Continua
Variable Predictora	Dicotómica	Test de $\chi^2$	t de Student
Variable Independiente	Continua	Regresión logística	Regresión lineal

t de Student

## Pruebas estadísticas de contraste de hipótesis más aplicables a cada situación

Variable Independiente o PREDICTORA	Variable Dependiente o RESULTADO	Pruebas de contraste de hipótesis	Observaciones y/o limitaciones del test
Cualitativa	Cualitativa	Ji cuadrado Prueba exacta de Fisher Test de McNemar Q de Cochran REGRESIÓN LOGÍSTICA	Si el tamaño muestral es grande Si el tamaño muestral es pequeño Dos grupos de muestras repetidas o apareadas Más de dos grupos de muestras repetidas o apareadas Varias variables predictoras cualitativas
Cualitativa	Cuantitativa	t de Student Análisis de la Varianza Test de Mann-Whitney Test de Wilcoxon Test de Kruskal-Wallis Test de Friedman	Compara medias entre dos grupos independientes Compara medias entre más de dos grupos independientes Dos grupos independientes No paramétricos Dos grupos medidas repetidas No paramétrico Más de dos grupos No paramétrico, medidas repetidas Más de dos grupos No paramétrico medidas repetidas
Cuantitativa	Cualitativa	Regresión logística simple Regresión logística multivariante	Una sola variable predictora cuantitativa Varias variable predictoras cuantitativa o cualitativas
Cuantitativa	Cuantitativa	Regresión lineal simple Correlación de Pearson Correlación de Sperman REGRESIÓN MÚLTIPLE	Paramétrico Paramétrico No paramétrico Varias variables predictoras cuantitativas
Cualitativa	Supervivencia	Kaplan-Meier Long-Rank Regresión de COX	Curvas de supervivencia Compara curvas de supervivencia Análisis Multivariante

# Tipos de test analíticos estadísticos bivariantes de contraste de hipótesis más frecuentes

Variable PREDICTORA o Independiente (factor de estudio)	Variable RESULTADO o Dependiente (Criterio de evaluación)			
	Cualitativa nominal dos categorías	Cualitativa nominal > 2 categorías	Cualitativa ordinal o Cuantitativa NO paramétrica	Cuantitativa paramétrica
Cualitativa dos grupos independientes	Chi cuadrado Z comparación de proporciones Test exacto de Fisher	Chi cuadrado	U de Mann-Whitney	T de Student (prueba de Welch)
Cualitativa dos grupos apareados	Test de McNemar	Q de Cochran	Prueba de los rangos de Wilcoxon	T de Student para datos emparejados
Cualitativa > 2 grupos independientes	Chi cuadrado	Chi cuadrado	Prueba de Kruskal-Wallis	Análisis de la Varianza (ANOVA)
Cualitativa > 2 grupos apareados	Q de Cochran	Q de Cochran	Prueba de Friedman	Análisis de la Varianza de dos vías
Cuantitativa	Regresión logística	Regresión logística	Correlación de Spearman Tau de Kendall	Regresión lineal Correlación de Pearson

## Cualitativa versus cuantitativa

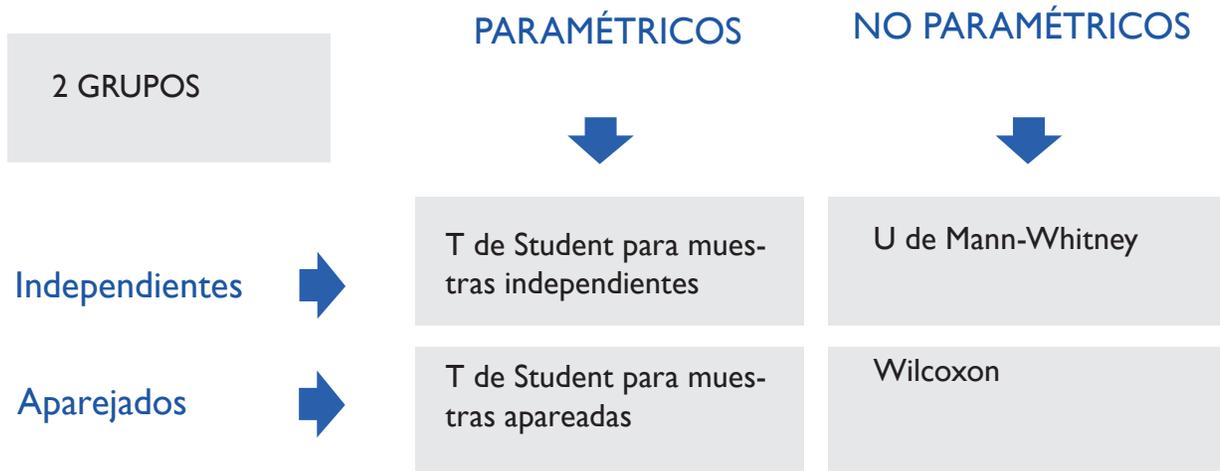
¿Cuáles son las pruebas estadísticas que tenemos que emplear?

Variable PREDICTORA  
CUALITATIVA dicotómica



Variable RESULTADO  
CUANTITATIVA

## Cualitativa dicotómica versus cuantitativa



Variable predictora Cualitativa versus variable resultado Cuantitativa.

# Test de la T de Student para dos muestras independientes

## Condiciones de aplicación (I)

- $n^1$  y  $n^2 \geq 30$
- Normalidad (Requisitos necesarios)
  - Comprobar que el máximo y el mínimo quedan dentro del intervalo definido por 3 DE por encima y por debajo de la media ( $\text{media} \pm 3 \text{ DE}$ )
  - Que la asimetría (en valor absoluto) sea menor que dos veces el error estándar [asimetría]  $< 2 \text{ EE de asimetría}$
  - Que la curtois (en valor absoluto) sea menor que 2 veces su EE [Curtois]  $< 2 \text{ EE de curtois}$

Test de Kolmogorov-Smirnov

Test de Shapiro-Wilks

## Test de la T de Student para dos muestras independientes

### Condiciones de aplicación (2)

Homogeneidad de varianzas (homoscedasticidad)

- Prueba F de Snedecor
- Test de Bartlett
- Test de Levene



$$F = \frac{\text{Varianza mayor}}{\text{Varianza menor}}$$

Ho: Las varianzas de ambos grupos son iguales ➔ Test T de Student

Ha: Existen diferencias entre las varianzas ➔ Test de Welch

$F_{\text{calculada}} < F_{\text{tabulada}}$        $p > 0,05$  (aceptamos la Ho)

$F_{\text{calculada}} > F_{\text{tabulada}}$        $p < 0,05$  (rechazamos la hipótesis nula)

## Pregunta de investigación

¿El peso al nacer de los varones difiere del peso al nacer de las mujeres?

Variable PREDICTORA  
CUALITATIVA dicotómica

Sexo



Variable RESULTADO  
CUANTITATIVA continua

Peso (Kg)

¿Cuál es la prueba estadística más adecuada?

## Test de la T de Student:

### Muestras independientes

Hipótesis nula  $H_0$ : El peso al nacer de los varones NO difiere del peso al nacer de las mujeres

Hipótesis alternativa  $H_a$ : El peso al nacer de los varones difiere del peso al nacer de las mujeres

	Peso al nacer del RN	
	Varones	Mujeres
Media	3380,3	3246,8
Desviación estándar (S)	417,3	388,0
n	833	789

$$t_{\text{calculada}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n-1)S_1^2 + (m-1)S_2^2}{(n+m-2)} \left( \frac{1}{n} + \frac{1}{m} \right)}}$$

$$t_{\text{calculada}} = 6,666$$

## Test de la T de Student para muestras independientes

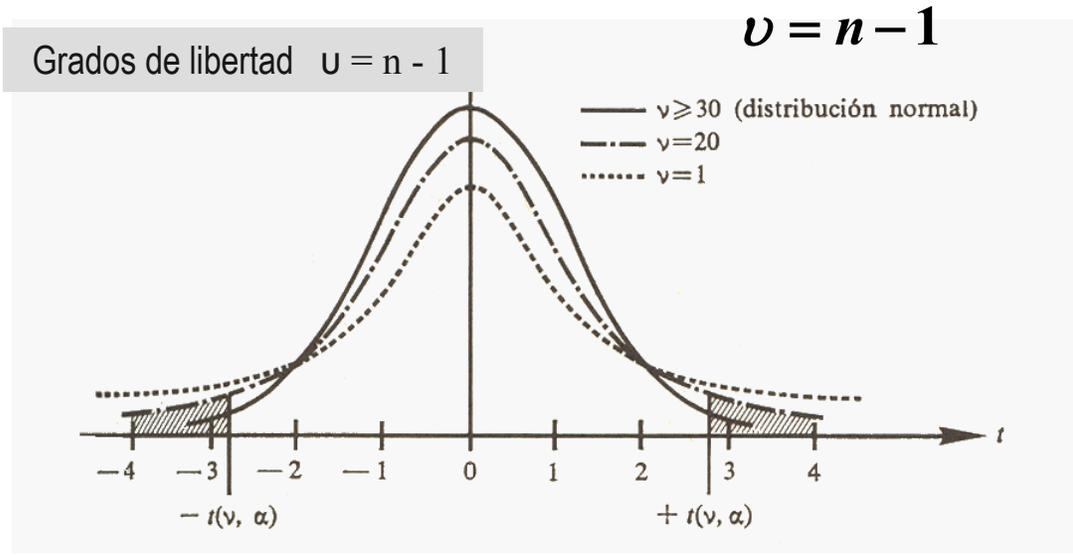
Grados de libertad (g.l) =  $n + m - 2$        $gl = 883 + 789 - 2 = 1620$

Peso del RN	Varones	Mujeres
n	883	789
Media	3380,34	3246,79
Desviación estándar	417,28	388,0
EEM	14,46	13,81
t calculada	6,666	
Grados de libertad	1620	
Significación (bilateral)	0,000	

$$t_{\text{calculada}} = 6,666 > t_{\text{tabulado}} = 1,96; \quad p < 0,001$$

$H_a$ : El peso al nacer de los varones difiere del peso al nacer de las mujeres

# Distribución de la t de Student

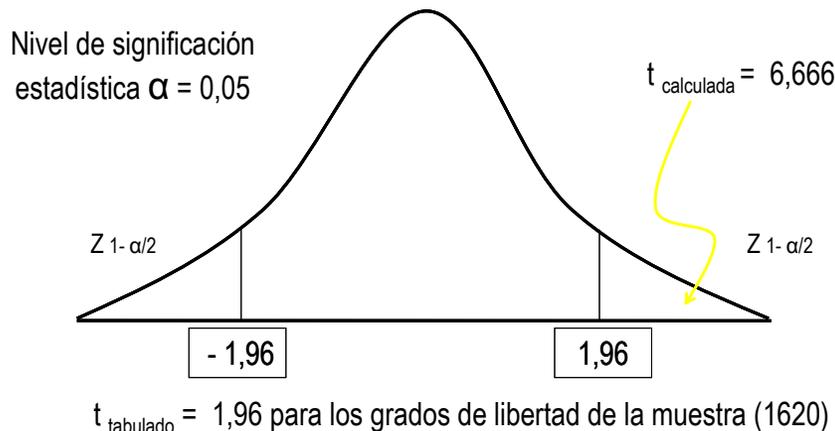


**valores "t":**  $t = \frac{(x - \mu)}{s}$

## Distribución de la t de Student ( $gl \geq 30$ )

El test de la T de Student para muestras de más de 30 individuos presenta una distribución muy similar a la distribución normal tipificada Z.

El test de la T de Student es una prueba BILATERAL, y por tanto NO especifica la dirección de la diferencia ( $A \neq B$ ).



## Interpretación del valor de la t de Student ( $t=6,666$ )

- La diferencia observada ( $t_{\text{calculada}} = 6,666$ ) es muy grande en relación a la variabilidad aleatoria esperada ( $t_{\text{tabulada}}$ ), y por tanto existen diferencias significativas ( $p < 0,001$ )
- El efecto observado (diferencia entre el peso de los dos sexos) es muy superior a la variabilidad biológica esperada.
- El efecto observado (diferencia entre el peso de los dos sexos) es importante, es decir mucho mayor de lo esperado por el azar.
- El valor de la  $t_{\text{tabulada}}$  para un error alfa del 5% es muy inferior al encontrado  
 $t_{\text{tabulado}} < t_{\text{calculada}}$
- El valor de la p ( $p < 0,001$ ) indica que existe una probabilidad muy baja ( $< 0,1\%$ ) de observar al menos estas diferencias en los pesos por efecto del AZAR y no porque los varones tengan distinto peso que las mujeres.

## Pregunta de investigación

¿Tiene influencia el tipo de dieta (pescado/huevos) seguida en el peso del individuo?

Variable PREDICTORA  
CUALITATIVA dicotómica

Tipo de dieta



Variable RESULTADO  
CUANTITATIVA continua

Peso (Kg)

# Test de la T de Student:

## Muestras independientes

Hipótesis nula  $H_0$  (ambas dietas tienen igual efecto)

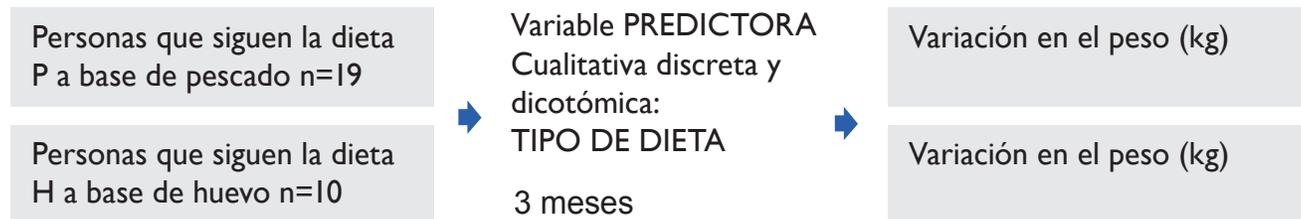
Hipótesis alternativa  $H_1$  (las dietas tienen distinto efecto)

Variable PREDICTORA

Cualitativa discreta y dicotómica: TIPO DE DIETA (pescado o huevos)

Variable RESULTADO

Cuantitativa continua: Kg adelgazados durante los 3 meses que duró la dieta



Personas que siguen la dieta  
P a base de pescado n=19

ID	DIETA	REDUCCION PESO (Kg)
1	PESCADO	19
2	PESCADO	28
3	PESCADO	17
4	PESCADO	15
5	PESCADO	16
6	PESCADO	21
7	PESCADO	27
8	PESCADO	18
9	PESCADO	15
10	PESCADO	10
11	PESCADO	10
12	PESCADO	2
13	PESCADO	12
14	PESCADO	13
15	PESCADO	19
16	PESCADO	9
17	PESCADO	29
18	PESCADO	35
19	PESCADO	38
Media		18,579
Desviación estándar		9,26
Varianza		85,8

Variable RESULTADO  
Cuantitativa continua

Personas que siguen la dieta  
H a base de huevo n=10

ID	DIETA	REDUCCION PESO (Kg)
20	HUEVOS	23
21	HUEVOS	23
22	HUEVOS	5
23	HUEVOS	14
24	HUEVOS	22
25	HUEVOS	20
26	HUEVOS	4
27	HUEVOS	12
28	HUEVOS	18
29	HUEVOS	33
30	HUEVOS	19
Media		17,545
Desviación estándar		8,43
Varianza		71,1

Variable RESULTADO  
Cuantitativa continua

$$t_{\text{calculada}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n-1)S_1^2 + (m-1)S_2^2}{(n+m-2)}} \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

$$t_{\text{calculada}} = 0,3$$

$$t_{\text{calculada}} = 0,3 < t_{\text{tabulado}} \quad p = 0,76$$

# Test de la T de Student:

## Muestras independientes

### Test de la t de Student

$t_{\text{calculada}} < t_{\text{tabulada}}$   
 $p > 0,05$

Hipótesis nula  $H_0$

Las medias de la reducción del peso en los dos grupos comparados es independiente de la intervención dietética

$t_{\text{calculada}} > t_{\text{tabulada}}$   
 $p < 0,05$

Hipótesis alternativa  $H_a$

Las medias de la reducción del peso en los dos grupos comparados son diferentes entre sí, es decir NO pueden considerarse iguales

$$t_{\text{calculada}} = 0,3 < t_{\text{tabulado}} \rightarrow p = 0,76$$

¿Tienen el mismo efecto las dietas ensayadas para reducir el peso corporal?

Sí

## Interpretación del valor de la t de Student ( $t=0,30$ )

- La diferencia observada ( $t_{\text{calculada}} = 0,3$ ) es pequeña en relación a la variabilidad aleatoria esperada ( $t_{\text{tabulada}}$ ), y por tanto no existen diferencias significativas ( $p > 0,05$ )
- El efecto observado (diferencia entre las dos dietas) es sólo del 30% de la variabilidad biológica esperada.
- El efecto observado (diferencia entre las dos dietas) es insignificante, es decir mucho menor de lo esperado por el azar.
- El valor de la  $t_{\text{tabulada}}$  para un error alfa del 5% es superior al encontrado  
 $t_{\text{tabulado}} > t_{\text{calculada}}$
- El valor de la p ( $p = 0,76$ ) indica que existe una probabilidad muy alta (76%) de observar al menos estas diferencias en la reducción del peso si la muestra procediese de una población donde las dos dietas tienen el mismo efecto (las diferencias observadas son producidas por el AZAR y no porque las dietas tengan distinto efecto).

# Test de la T de Student para dos muestras independientes

## Condiciones de aplicación (I)

- $n^1$  y  $n^2 \geq 30$
- Normalidad (Requisitos necesarios)
  - Comprobar que el máximo y el mínimo quedan dentro del intervalo definido por 3 DE por encima y por debajo de la media ( $\text{media} \pm 3 \text{ DE}$ )
  - Que la asimetría (en valor absoluto) sea menor que dos veces el error estándar [asimetría]  $< 2 \text{ EE de asimetría}$
  - Que la curtois (en valor absoluto) sea menor que 2 veces su EE [Curtois]  $< 2 \text{ EE de curtois}$

Test de Kolmogorov-Smirnov

Test de Shapiro-Wilks

## Test de la T de Student para dos muestras independientes

### Condiciones de aplicación (2)

Homogeneidad de varianzas (homoscedasticidad)

- Prueba F de Snedecor
- Test de Bartlett
- Test de Levene



$$F = \frac{\text{Varianza mayor}}{\text{Varianza menor}}$$

Ho: Las varianzas de ambos grupos son iguales ➔ Test T de Student

Ha: Existen diferencias entre las varianzas ➔ Test de Welch

$F_{\text{calculada}} < F_{\text{tabulada}}$        $p > 0,05$  (aceptamos la Ho)

$F_{\text{calculada}} > F_{\text{tabulada}}$        $p < 0,05$  (rechazamos la hipótesis nula)

## Mito I: “La sonrisa de la esfinge”

La t de Student es un test paramétrico, es decir para usarlo se debe cumplir la normalidad de la variable cuantitativa. ¿Se cumple? ¿Como lo has comprobado?

He observado que en los dos grupos de estudio comparados sólo tienen 19 y 11 sujetos respectivamente. Ante tan bajo volumen muestral ( $n < 30$ ),

¿Cómo justificas esa supuesta normalidad? ¿Que test has empleado para comprobar esa normalidad de la reducción del peso? ¿Estás seguro que se cumplen los criterios para poder usar la t de Student ? ¿Se cumplen todos los requisitos para su utilización?

## Realidad I: “Respuesta de Edipo”

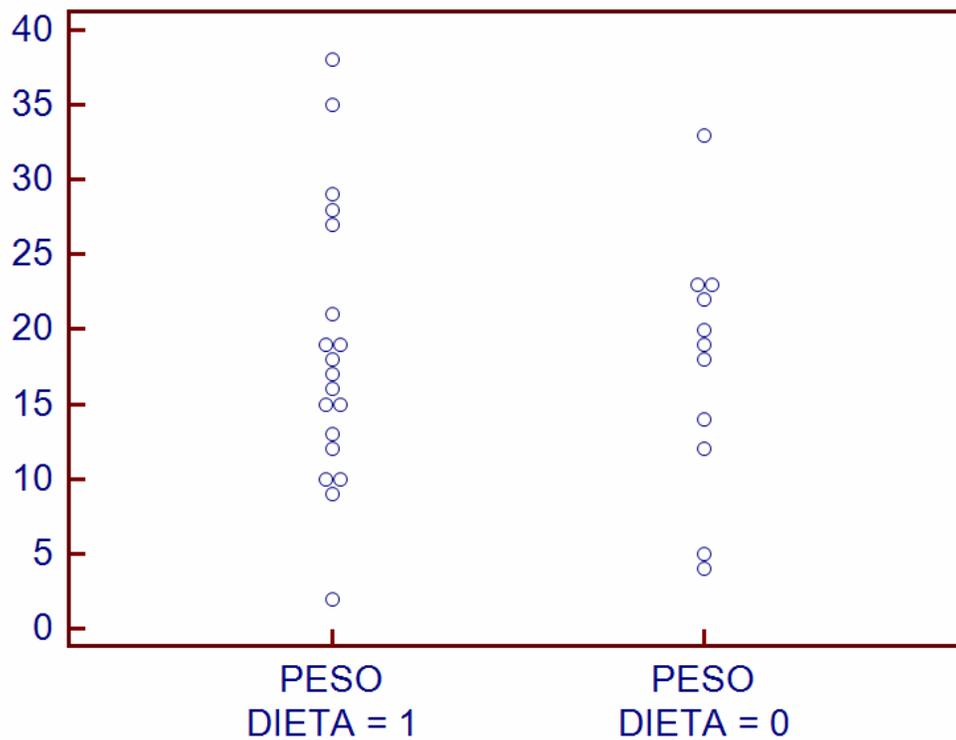
Test de Kolmogorov-Smirnov				
	Tipo de dieta	Estadístico	g.l	Significación
Reducción del peso	Pescado	0,166	19	0,178
	Huevos	0,168	11	0,200

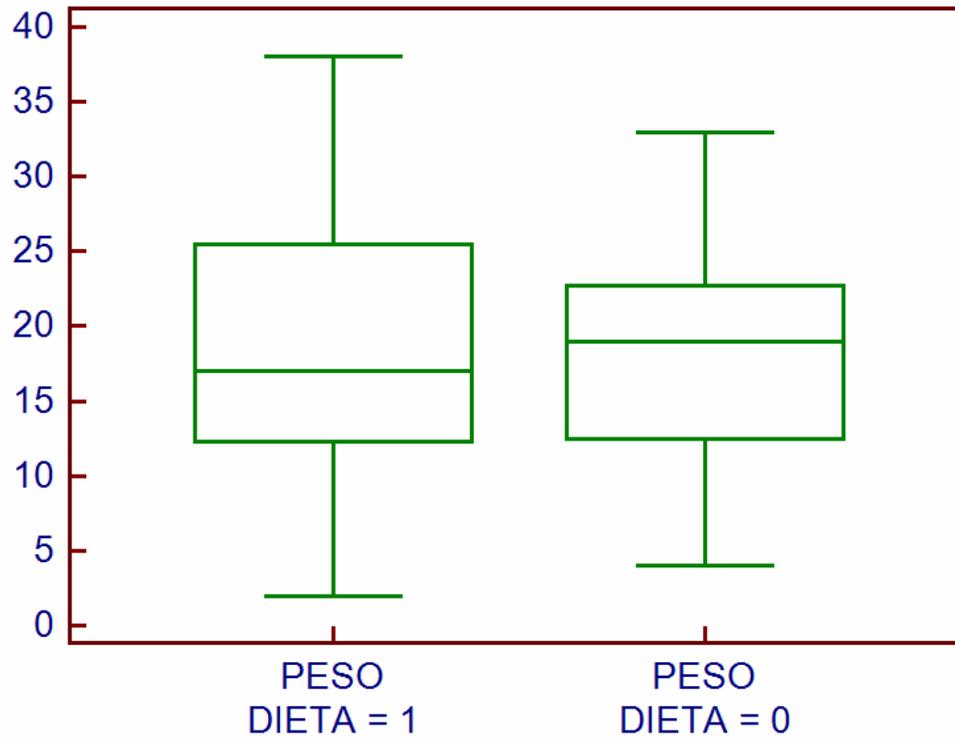
Test de Shapiro-Wilk				
	Tipo de dieta	Estadístico	g.l	Significación
Reducción del peso	Pescado	0,956	19	0,493
	Huevos	0,950	11	0,615

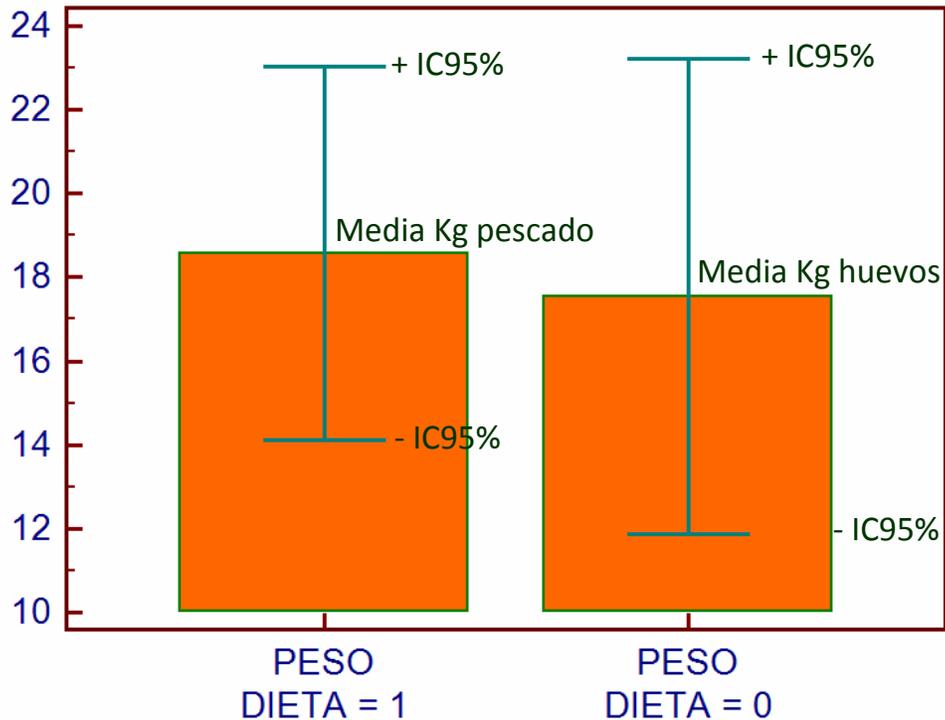
$H_0: p > 0,05$  test paramétricos

$H_a: p < 0,05$  test NO paramétricos

Los test de normalidad ensayados dan un valor de p NO significativo ( $p > 0,05$ ) en ambos grupos (tipos de dieta). Por tanto no hay evidencias suficientes para rechazar la hipótesis nula de normalidad de la variable reducción del peso en ambas muestras.







## Pregunta de investigación

¿Existen diferencias en los valores del VCM en sangre en función de los depósitos de hierro normales o ferropénicos en médula ósea?

Variable PREDICTORA  
CUALITATIVA dicotómica

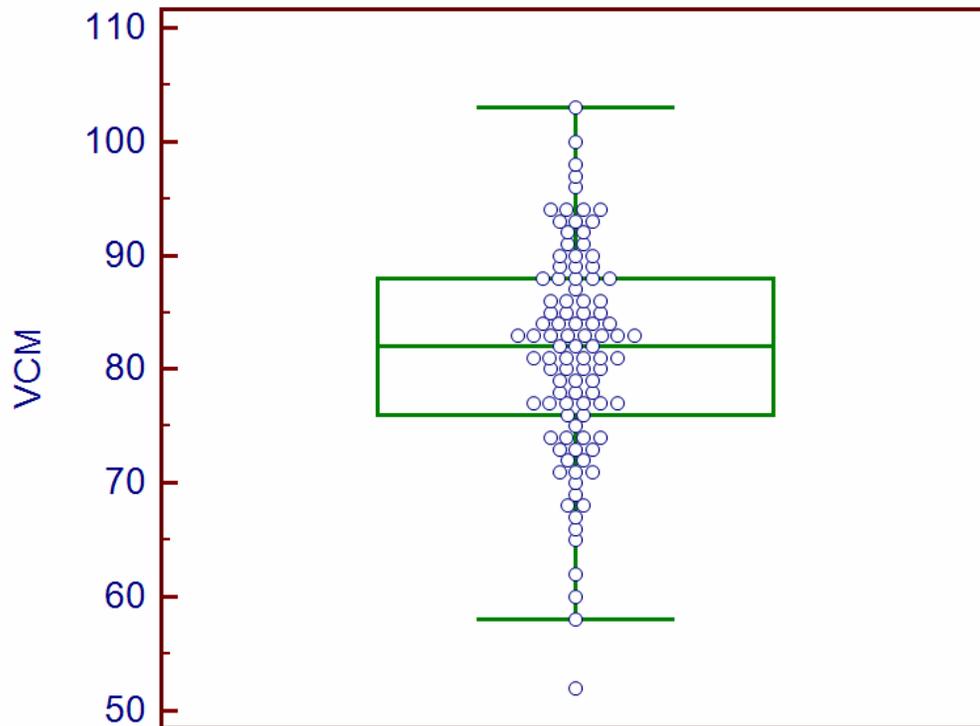
Depósitos de Fe en médula ósea  
Normal = 1  
Déficit = 0

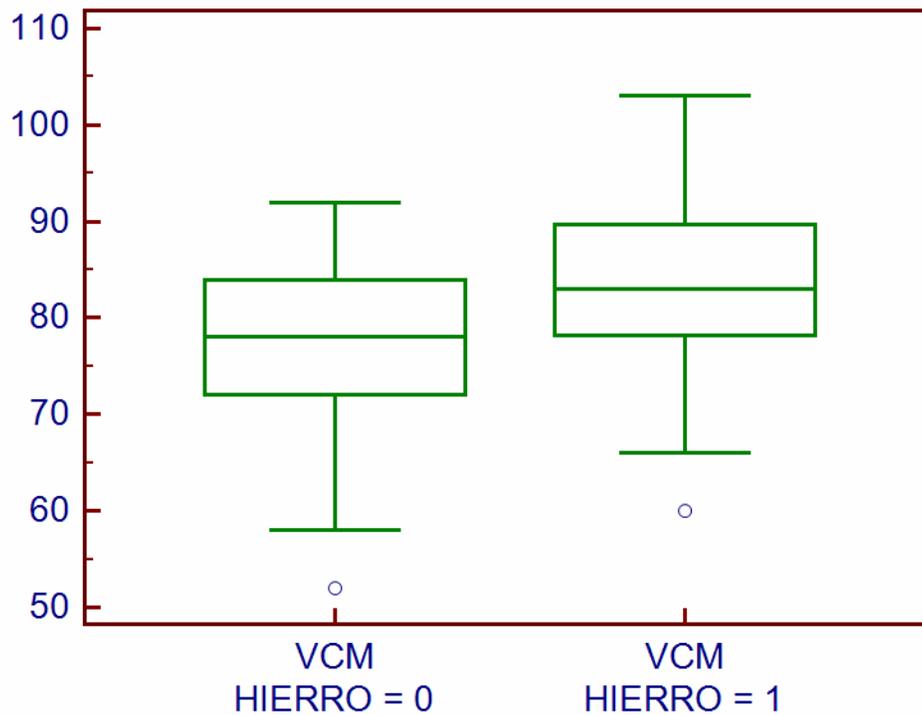


Variable RESULTADO  
CUANTITATIVA continua

VCM  
(hemograma)

¿Cuál es el test estadístico más acertado?





## Test de la T de Student para muestras independientes

Grados de libertad (g.l) =  $n + m - 2$

$$gl = 67 + 33 - 2 = 98$$

VCM	Medula normal	Déficit de hierro
n	67	33
Media	83,56	76,89
Desviación estándar	8,588	9,323
EEM	1,049	1,623
$t_{\text{calculada}}$		3,560
Grados de libertad		98
Significación (bilateral)		0,001

$$t_{\text{calculada}} = 3,560 > t_{\text{tabulado}} = 1,96; \quad p < 0,001$$

$H_a$ : El VCM es depende de los depósitos medulares de hierro.

## Pregunta de investigación

¿Existen diferencias en los valores de la ferritina en sangre en función de los depósitos de hierro normales o ferropénicos en médula ósea?

Variable **PREDICTORA**  
**CUALITATIVA** dicotómica

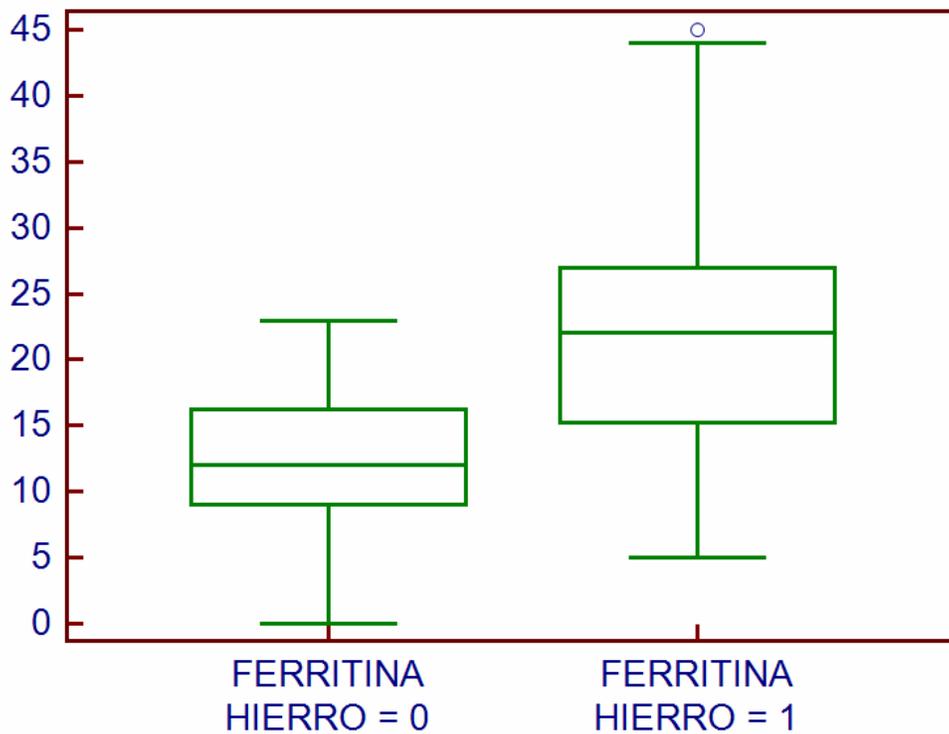
Depósitos de Fe en médula ósea  
Normal = 1  
Déficit = 0



Variable **RESULTADO**  
**CUANTITATIVA** continua

FERRITINA sérica

¿Cuál es el test estadístico más acertado?



## Test de Welch para desigualdad de varianzas

FERRITINA sérica	Médula normal	Déficit de hierro
n	67	33
Media	21,76	12,61
Desviación estándar	9,165	5,528
EEM	1,120	0,962
T <sub>calculada</sub>		6,201
Grados de libertad		93,876
Significación (bilateral)		0,000

$$t_{\text{calculada}} = 3,560 > t_{\text{tabulado}} = 1,96; \quad p < 0,001$$

H<sub>a</sub>: El ferritina sérica depende de los depósitos medulares de hierro.

## Pregunta de investigación

¿Qué relación existe entre el grupo cultural/religioso y los niveles de anticuerpos frente al virus de la Rubeola en la población gestante inmunizada de Ceuta en el año 1994?

Variable PREDICTORA

Cualitativa dicotómica

Cristianas inmunizadas  
Musulmanas inmunizadas



Variable RESULTADO

Cuantitativa continua

Virus de la Rubeola

¿Cuál es el test estadístico más adecuado?

## Test de t de Student para muestras independientes

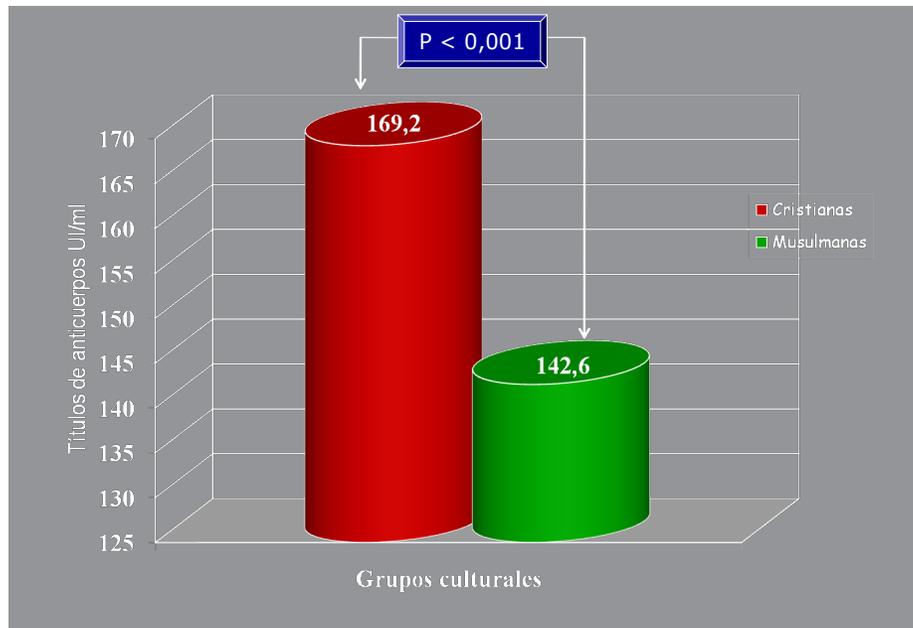


Figura 1.- Títulos de anticuerpos frente al virus de la Rubeola en gestantes inmunizadas de Ceuta en 1994. Distribución por grupos culturales.

## Pregunta de investigación

¿Qué relación existe entre el grupo cultural/religioso y los niveles de anticuerpos frente al virus de la Rubeola en la población gestante inmunizada de Ceuta en el año 2010?

Variable PREDICTORA  
Cualitativa dicotómica

Cristianas inmunizadas  
Musulmanas inmunizadas



Variable RESULTADO  
Cuantitativa continua

Virus de la Rubeola

¿Cuál es el test estadístico más adecuado?

## Test de la T de Student para muestras independientes

Grados de libertad (g.l) =  $n + m - 2$

$gl = 234 + 487 - 2 = 719$

IgG-anti-RUBEOLA	Cristianas	Musulmanas
n	234	487
Media	133,20	94,90
Desviación estándar	197,26	100,69
EEM	12,89	4,56
t calculada		3,452
Grados de libertad		719
Significación (bilateral)		0,001

$t_{\text{calculada}} = 3,452 > t_{\text{tabulado}} = 1,96; \quad p < 0,001$

$H_a$ : Los niveles de IgG anti-Rubeola depende del grupo cultural, siendo distintos en cristianas y musulmanas.

## Test de t de Student para muestras independientes (varianzas desiguales)

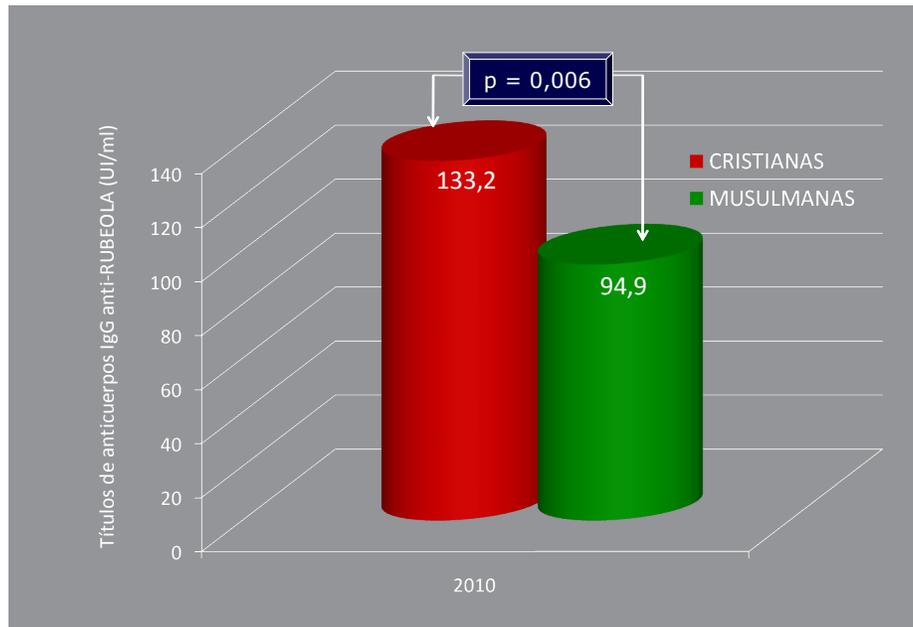


Figura 2.- Títulos de anticuerpos frente al virus de la Rubeola en gestantes inmunizadas de Ceuta en 2010. Distribución por grupos culturales.

## Mito I: “La sonrisa de la esfinge”

Para poder usar la  $t$  de Student para muestras independientes es necesario, entre otras condiciones, la homogeneidad de las varianzas (homocedasticidad) de la variable cuantitativa.

¿La variable cuantitativa niveles de anticuerpos IgG anti-RUBEOLA (UI/ml) en la población gestante inmunizada de Ceuta cumple la homocedasticidad?

¿Que test has empleado para comprobar que se cumple la homogeneidad de las varianzas?

## Test de Levene para homocedasticidad

Test de Levene para la igualdad de varianzas	IgG-anti-RUBEOLA
F	14,498
Significación	0,000

$H_0$ : Las varianzas de ambos grupos son iguales.

$H_a$ : Existen diferencias entre las varianzas.

$F_{\text{calculada}} > F_{\text{tabulada}} \Rightarrow p < 0,05$  (rechazamos la hipótesis nula)

Test de Welch

## Test de Welch para desigualdad de varianzas

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

IgG-anti-RUBEOLA	Cristianas	Musulmanas
n	234	487
Media	133,20	94,90
Desviación estándar	197,26	100,69
EEM	12,89	4,56
T calculada		2,800
Grados de libertad		292,8
Significación (bilateral)		0,005

$$t_{\text{calculada}} = 2,800 > t_{\text{tabulado}} = 1,96; \quad p < 0,01$$

$H_a$ : Los niveles de IgG anti-Rubeola depende del grupo cultural, siendo distintos en cristianas y musulmanas.

## Pregunta de investigación

¿Qué relación existe entre el grupo cultural/religioso y los niveles de anticuerpos frente a *Toxoplasma* en la población gestante inmunizada de Ceuta en el año 1994?

Variable PREDICTORA  
Cualitativa dicotómica

Cristianas inmunizadas  
Musulmanas inmunizadas



Variable RESULTADO  
Cuantitativa continua

Virus de la Rubeola

¿Cuál es el test estadístico más adecuado?

## Test de t de Student para muestras independientes

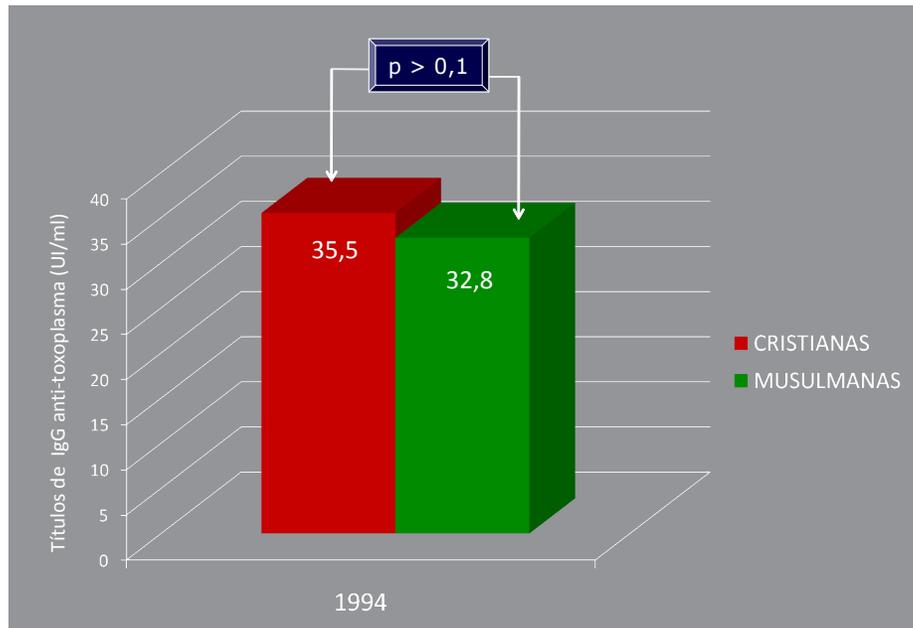


Figura 1.- Títulos de anticuerpos frente a Toxoplasma en gestantes inmunizadas de Ceuta en 1994. Distribución por grupos culturales.

## Pregunta de investigación

¿Qué relación existe entre el grupo cultural/religioso y los niveles de anticuerpos frente a *Toxoplasma* en la población gestante inmunizada de Ceuta en el año 2010?

Variable PREDICTORA  
Cualitativa dicotómica

Cristianas inmunizadas  
Musulmanas inmunizadas



Variable RESULTADO  
Cuantitativa continua

*Toxoplasma*

¿Cuál es el test estadístico más adecuado?

## Test de t de Student para muestras independientes

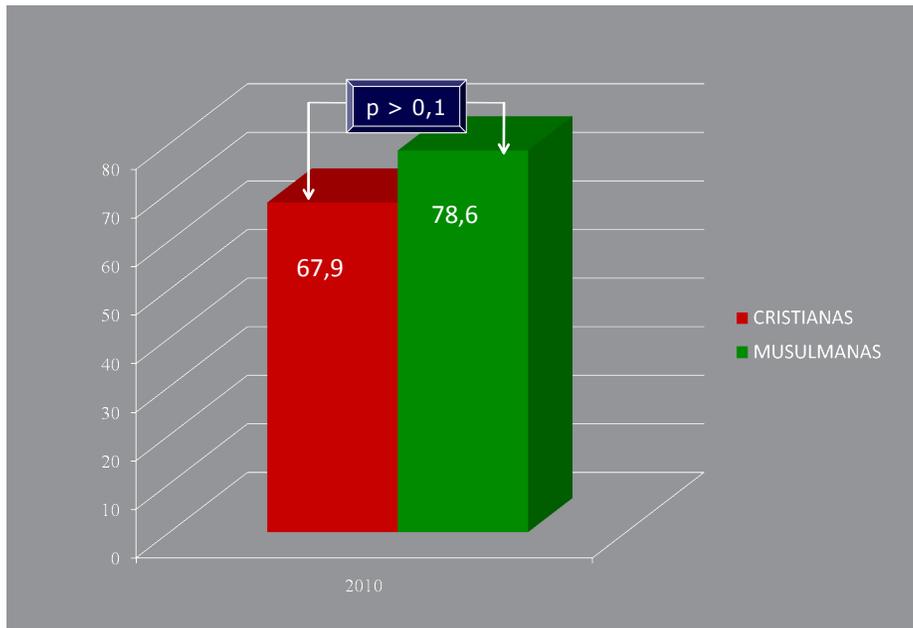


Figura 2.- Títulos de anticuerpos frente a Toxoplasma en gestantes inmunizadas de Ceuta en 2010. Distribución por grupos culturales.

## Mito I: “La sonrisa de la esfinge”

La t de Student para muestras independientes es un test paramétrico, es decir para usarlo se debe cumplir la normalidad de la variable cuantitativa.  
¿La variable cuantitativa niveles de anticuerpos (UI/ml) en la población gestante inmunizada de Ceuta sigue una distribución normal?

No encuentro datos estadísticos de dispersión (desviación típica)...  
¿Cómo justificas esa supuesta normalidad? ¿Qué test has empleado para comprobar que se cumple la normalidad (campana de Gauss)?

Test de normalidad	Test de Kolmogorov-Smirnov				
Año 2010	N	Media	DE	Estadístico Z	Significación
Niveles de anticuerpos IgG anti-RUBEOLA	721	107,3	140,6	6,870	0,000

Test de normalidad	Test de Kolmogorov-Smirnov					
Año 2010	Tipo de cultura	N	Media	DE	Estadístico Z	Significación
Niveles de anticuerpos IgG anti-RUBEOLA	Cristianas	234	133,2	197,3	4,211	0,000
	Musulmanas	487	94,9	100,7	4,723	0,000

$H_0: p > 0,05$  test paramétricos

$H_a: p < 0,05$  test NO paramétricos

Los test de normalidad ensayados dan un valor de p significativo ( $p < 0,05$ ) en ambos grupos culturales. Por tanto hay evidencias suficientes para rechazar la hipótesis nula de normalidad de la variable reducción del peso en ambas muestras. NO PODEMOS USAR TEST PARAMÉTRICOS.

Test de normalidad	Test de Kolmogorov-Smirnov				
Año 2010	N	Media	DE	Estadístico Z	Significación
Niveles de anticuerpos IgG anti-RUBEOLA	297	77,5	133,6	5,321	0,000

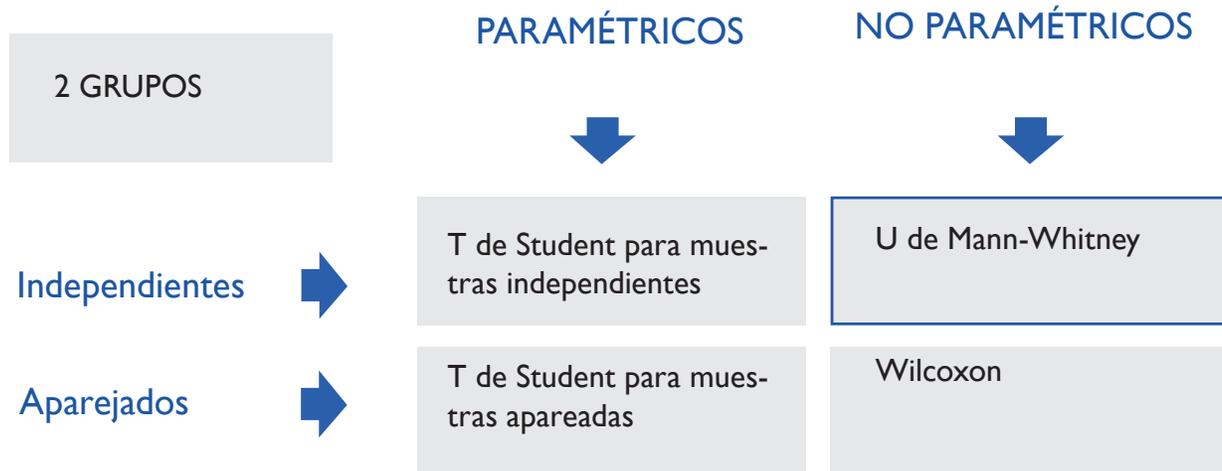
Test de normalidad	Test de Kolmogorov-Smirnov					
Año 2010	Tipo de cultura	N	Media	DE	Estadístico Z	Significación
Niveles de anticuerpos IgG anti-RUBEOLA	Cristianas	32	67,9	58,6	1,260	0,084
	Musulmanas	265	78,62	139,9	5,110	0,000

$H_0: p > 0,05$  test paramétricos

$H_a: p < 0,05$  test NO paramétricos

Los test de normalidad ensayados dan un valor de p significativo ( $p < 0,05$ ) en ambos grupos culturales. Por tanto hay evidencias suficientes para rechazar la hipótesis nula de normalidad de la variable reducción del peso en ambas muestras. NO PODEMOS USAR TEST PARAMÉTRICOS.

## Cualitativa dicotómica versus cuantitativa



Variable predictora Cualitativa versus variable resultado Cuantitativa.

## Pregunta de investigación

¿Existen diferencias de la tensión arterial sistólica (TAS) de los 20 alumnos del curso de Metodología de investigación en función del sexo?

Variable PREDICTORA  
Cualitativa dicotómica

Sexo



Variable RESULTADO  
Cuantitativa

TAS

¿Cuál es el test estadístico más adecuado?

## Test de la U de Mann-Whitney

Condiciones de aplicación:

- Si alguna de las dos muestras contiene menos de 30 observaciones y No se puede asumir la normalidad.
- Variable medida en escala ORDINAL.
- Si la muestra es de muy pequeño tamaño (<10 observaciones en alguno de los dos grupos).

Inconvenientes:

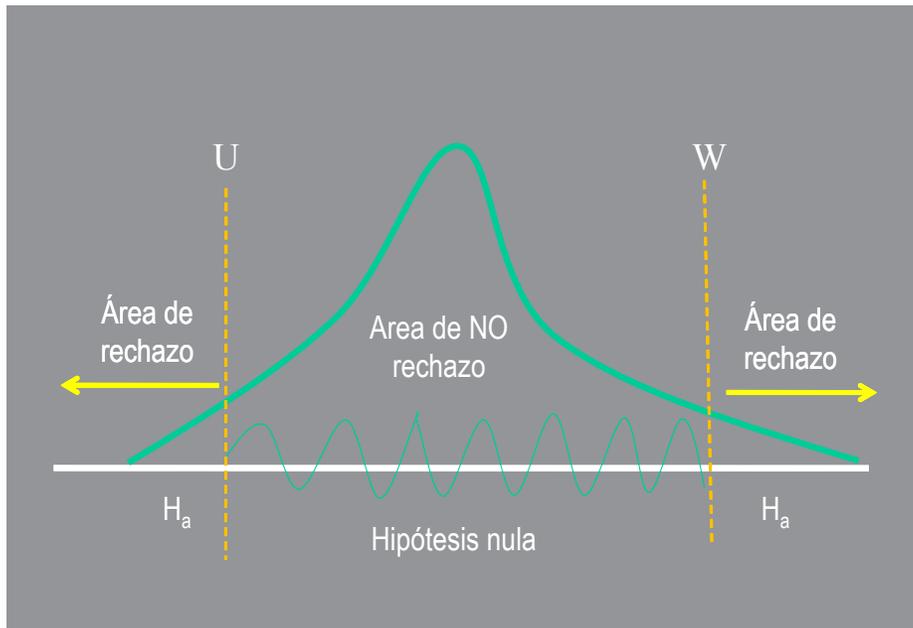
- Se pierde potencia para detectar diferencias significativas (para una comparación que resulte significativa con 95 individuos en la t de Student, se necesitarían 100 individuos para que la misma comparación resulte significativa con el test de Mann-Whitney).
- No permite construir intervalos de confianza.

Alumnos	TAS Alumnos	Rango 1 (R1)	Orden	Alumnas	TAS Alumnas	Rango 2 (R2)	Orden
1º	110	9,5	7	1ª	90	1	1
2º	120	16	15	2ª	95	2,5	2
3º	100	5	4	3ª	110	9,5	9
4º	115	13,5	13	4ª	110	9,5	10
5º	110	9,5	8	5ª	120	16	17
6º	100	5	5	6ª	115	13,5	14
7º	120	16	16	7ª	110	9,5	11
8º	125	18,5	18	8ª	100	5	6
9º	125	18,5	19	9ª	95	2,5	3
10º	125	20	20	10ª	100	9,5	12
10	Total			10		78,5	

Hipótesis nula  $H_0$ : La TAS es igual en alumnos y en alumnas.

Hipótesis alternativa  $H_a$ : La TAS es diferente en alumnos y en alumnas.

## U de Mann-Whitney



Los valores de estadístico U y W calculados por el test se comparan con los tabulados para un contraste bilateral con un error alfa 0,05.

## U de Mann-Whitney

	SEXO	N	Mean Rank	Sum of Ranks
TAS	Alumna	10	7,85	78,50
	Alumno	10	13,15	131,50
	Total	20		

Test estadísticos	TAS
Mann-Whitney U	23,50
Wilcoxon W	78,50
Zu	- 2,030
Significación a dos colas	0,041
Significación exacta a una cola	0,043

$$\text{Mean rank} = \frac{\text{Sum of ranks}}{N}$$

$$Z_u = \frac{U - R}{\frac{\sigma}{\sqrt{n}}}$$

El ranking promedio de la TAS en mujeres es 7,85 y en los hombres 13,15, es decir existen diferencias entre alumnos y alumnas, y por tanto como  $p < 0,05$  se rechaza la  $H_0$ .

Conclusión: La TAS es diferente en alumnos y en alumnas, siendo mayor en varones.

## Ejemplo para muestras pequeñas

- Un profesor de bioestadística utiliza dos tipos de modelos de cursos para enseñar la misma materia a un grupo de 10 alumnos médicos con un nivel previo similar de conocimientos en bioestadística. El experimentador quiere demostrar que el procedimiento clásico de asistencia a clase (PRESENCIAL) es más efectivo que el curso online (a DISTANCIA) que le propone la dirección que presupone que es más eficiente; para ello, mide el nivel de conocimientos adquiridos por los alumnos en bioestadística en ambos modelos de curso mediante el mismo examen de conocimientos realizado al final de cada tipo de curso.
- El plan experimental preliminar consiste en elegir al azar tanto una muestra total de 10 médicos, 5 para cada método.

## Planteamiento de la hipótesis

- Hipótesis nula ( $H_0$ ). Las diferencias observadas entre las calificaciones del examen de conocimientos mediante los dos tipos de cursos se deben al azar.
- Hipótesis alternativa ( $H_a$ ). Las calificaciones del examen de conocimientos, según el método de enseñanza PRESENCIAL son más altas y diferentes que las observadas en el método A DISTANCIA.

¿Cuál es la prueba estadística más adecuada?

Tipo de Curso	Calificaciones de los alumnos				
Nº del alumno	1º	2º	3º	4º	5º
A DISTANCIA	80	85	25	70	90

Tipo de Curso	Calificaciones de los alumnos				
Nº del alumno	6º	7º	8º	9º	10º
PRESENCIAL	95	100	93	110	45

- Nivel de significación: Para todo valor de probabilidad igual o menor que  $p = 0,05$ , se acepta  $H_a$  y se rechaza  $H_0$ .
- Zona de rechazo: Para todo valor de probabilidad mayor que  $0,05$ , se acepta  $H_0$  y se rechaza  $H_a$ .
- Dos cursos diferentes aplicados en dos grupos de alumnos médicos.

## Test de la U de Mann-Whitney (paso I)

- Aplicación de la prueba estadística.  
De acuerdo con los pasos, las observaciones se deben ordenar en rangos del menor al mayor.
- Rangos de lectura de la tabla anterior.

Tipo de Curso	Calificaciones de los alumnos					$\Sigma R (S)$
A distancia Grupo 1	80 (4)	85 (5)	25 (1)	70 (3)	90 (6)	19
Presencial Grupo 2	95 (8)	100 (9)	93 (7)	110 (10)	45 (2)	36

## Test de la U de Mann-Whitney (paso 2)

Calculamos las U de Mann-Whitney ( $U_1$  y  $U_2$ ) con las siguientes fórmulas:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - S_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - S_2$$

## Test de la U de Mann-Whitney (paso 2)

Calculamos las U de Mann-Whitney ( $U_1$  y  $U_2$ )

$$U_1 = n_1 n_2 + \frac{n_1 (n_1 + 1)}{2} - \Sigma R_1 = 5 \times 5 + \frac{5 (5 + 1)}{2} - 19 = 21$$

$$U_2 = n_1 n_2 + \frac{n_2 (n_2 + 1)}{2} - \Sigma R_2 = 5 \times 5 + \frac{5 (5 + 1)}{2} - 36 = 4$$

## Test de la U de Mann-Whitney (paso 3)

- De los dos valores de U calculados ( $U_1$  y  $U_2$ ), se elige el valor más pequeño ( $U_2 = 4$ ) y se comparan con los valores tabulados críticos de la U Mann-Whitney

- En caso de que el valor de U calculado no se localice en las tablas correspondientes, se obtendrá de la fórmula siguiente;

$$U_{\text{calculado}} = [n_1 * n_2] - U' \quad U_{\text{calculado}} = [5 * 5] - 21 = 4$$

- En esta fórmula, U' corresponde al valor más alto ( $U_1 = 21$ ) . El resultado en ambos casos es el mismo (4).

$n_2 = 5$

Tamaño muestral

$U$	$n_1$	1	2	3	4	5
0		0,167	0,047	0,018	0,008	0,004
1		0,333	0,095	0,036	0,016	0,008
2		0,500	0,190	0,071	0,032	0,016
3		0,667	0,286	0,125	0,056	0,028
$U_{\text{calculado}} = 4$			0,429	0,196	0,095	0,048
5			0,571	0,286	0,143	0,075
6				0,393	0,206	0,111
7				0,500	0,278	0,155
8				0,607	0,365	0,210
9					0,452	0,274
10					0,548	0,345
11						0,421
12						0,500
13						0,579

$p = 0,048$

## Decisión

A la probabilidad del valor U de Mann-Whitney, calculado anteriormente ( $U_{\text{calculado}}$ ), corresponde a un valor de la p de 0,048, el cual es más pequeño que el nivel de significancia (0,05); por lo tanto, se acepta  $H_a$  y se rechaza  $H_o$ .

## Interpretación

Entre las calificaciones del examen de conocimientos adquiridos mediante las dos modalidades del curso de enseñanza de la bioestadística, existe una diferencia significativa con un nivel de probabilidad de error alfa menor del 5% ( $p < 0,05$ ); es decir, aun cuando las muestras de ambos grupos son pequeñas, las calificaciones más altas mediante el curso clásico PRESENCIAL señalan más efectividad, con la probabilidad de equivocarse de 0,048 para aceptarlo por la influencia del AZAR (4,8%).

## Ejemplo aplicable cuando la muestra es mayor de 25 y donde $n_1$ y $n_2$ pueden ser iguales o de un tamaño diferente:

El Profesor del curso de Bioestadística aplicada para médicos, entusiasmado por las observaciones y conclusiones preliminares, decide repetir el experimento aumentando el tamaño muestral en ambos grupos. Diseña un nuevo estudio con 10 alumnos médicos que realizan el curso con el método a DISTANCIA y 25 médicos mediante el procedimiento clásico PRESENCIAL. Los nuevos datos del estudio se muestran en la tabla siguiente.

## Planteamiento de la hipótesis

- Hipótesis nula ( $H_0$ ). Las posibles diferencias entre las calificaciones dadas por los alumnos en ambos tipos de cursos se deben al azar.
- Hipótesis alternativa ( $H_a$ ). Las calificaciones de los alumnos inscritos en el curso realizado por el método clásico PRESENCIAL, propuesto por el experimentador, son diferentes y con valores más altos.
- Nivel de significación: Para todo valor de probabilidad igual o menor que 0,05, se acepta  $H_a$  y se rechaza  $H_0$ .
- Zona de rechazo: Para todo valor de probabilidad mayor que 0,05, se acepta  $H_0$  y se rechaza  $H_a$ .

## Población de alumnos médicos a los cuales se les aplicó dos tipos de cursos (método presencial, método a distancia)

### Curso PRESENCIAL N=25

PRESENCIAL	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Notas	55	70	90	110	45	60	60	75	80	95	100	110	95	60	70	80	40	65	50	75	90	90	100	80	100

### Curso a DISTANCIA (online) N=10

DISTANCIA N=10	1º	2º	3º	4º	5º	6º	7º	8º	9º	10º
Notas	60	80	25	30	40	60	90	100	60	55

¿Cuál es la prueba estadística más adecuada?

## Test de la U de Mann-Whitney (paso I)

- Aplicación de la prueba estadística. Primero ordenamos los rangos de todas las observaciones.
- Dirección de las ligas o empates y el tamaño de estas.

Rangos y empates	Rangos y empates	Rangos y empates
25 (1)	65 (15)	95 (28.5) $\frac{28+29}{2} = 28.5$
30 (2)	70 (16.5) $\frac{16+17}{2} = 16.5$	95 (28.5)
40 (3.5) $\frac{3+4}{2} = 3.5$	70 (16.5)	100 (31.5) $\frac{30+31+32+33}{4} = 31.5$
40 (3.5)	75 (18.5) $\frac{18+19}{2} = 18.5$	100 (31.5)
45 (5)	75 (18.5)	100 (31.5)
50 (6)	80 (21.5) $\frac{20+21+22+23}{4} = 21.5$	100 (31.5)
55 (7.5) $\frac{7+8}{2} = 7.5$	80 (21.5)	100 (31.5)
55 (7.5)	80 (21.5)	110 (34.5) $\frac{34+35}{2} = 34.5$
60 (11.5) $\frac{9+10+11+12+13+14}{6} = 11.5$	80 (21.5)	110 (34.5)
60 (11.5)	90 (25.5) $\frac{24+25+26+27}{4} = 25.5$	
60 (11.5)	90 (25.5)	
60 (11.5)	90 (25.5)	
60 (11.5)	90 (25.5)	

## Test de la U de Mann-Whitney (paso 2)

Calculamos las U de Mann-Whitney ( $U_1$  y  $U_2$ )

$$U_1 = n_1 n_2 + \frac{n_1 (n_1 + 1)}{2} - \Sigma R_1 = 10 \times 25 + \frac{10 (10 + 1)}{2} - 127 = 178$$

$$U_2 = n_1 n_2 + \frac{n_2 (n_2 + 1)}{2} - \Sigma R_2 = 10 \times 25 + \frac{25 (25 + 1)}{2} - 503 = 72$$

## Test de la U de Mann-Whitney (paso 3)

Teniendo en cuenta los pasos del test, hay que saber que cuando el tamaño muestral total es mayor de 25 (como ocurre en este caso), el estadístico del test de Mann-Whitney se distribuye normalmente, por tanto se calcula el estadístico Z, que se compara con el tabulado para conocer la probabilidad (p). Esto se calcula como sigue:

Z

$$Z = \frac{U - \bar{U}}{\sigma_U}$$

## Test de la U de Mann-Whitney (paso 3a)

Donde:

Z = valor estadístico de la curva normal.

U = cualquier valor de U calculado (ya sea  $U_1$  o  $U_2$ ).

$\bar{U}$  = valor promedio de U.

$\sigma_U$  = desviación estándar de U.

Calculamos el valor promedio de U:

$$\bar{U} = \frac{n_1 * n_2}{2} \quad \bar{U} = \frac{10 * 25}{2} = 125$$

## Test de la U de Mann-Whitney (paso 3b)

- La desviación estándar de U ( $\sigma_U$ ) se determina de la forma siguiente:

- Donde:

$\sigma_U$  = desviación estándar de U.

$n_1$  y  $n_2$  = tamaño de la muestra de los grupos 1 y 2.

N = tamaño total de la muestra (la suma de  $n_1$  y  $n_2$ ).

$L_i$  = sumatoria de las ligas o empates.

El cálculo de  $L_i$  se realiza de la siguiente manera:

$$L_i = \sum_i^k \frac{t_i^3 - t_i}{12}$$

$$\begin{aligned}
 L_i &= \frac{\sum L_i^3 - L_i}{12} \\
 &= \frac{(2^3 - 2) + (2^3 - 2) + (6^3 - 6) + (2^3 - 2) + (2^3 - 2) + (4^3 - 4) + (4^3 - 4)}{12} \\
 &\quad + \frac{(2^3 - 2) + (4^3 - 4) + (2^3 - 2)}{12} = 35.5
 \end{aligned}$$

$$Z = \frac{U - \bar{U}}{\sigma_U}$$

$$\bar{U} = \frac{n_1 * n_2}{2}$$

$$Li = \sum_i^k \frac{t_i^3 - t_i}{12}$$

$$\sigma_U = \sqrt{\frac{n_1 n_2}{n(n-1)} \left( \frac{n^3 - n}{12} - \sum_i^k \frac{t_i^3 - t_i}{12} \right)}$$

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2}{n(n-1)} \left( \frac{n^3 - n}{12} - \sum_i^k \frac{t_i^3 - t_i}{12} \right)}}$$

## Test de la U de Mann-Whitney (paso 3c)

- Una vez obtenida la sumatoria de  $L_i$ , se determinará la desviación estándar de  $U$  ( $\sigma_U$ ) mediante la expresión siguiente:

$$\sigma_U = \sqrt{\frac{n_1 n_2}{n(n-1)} \left( \frac{n^3 - n}{12} - \sum_i \frac{t_i^3 - t_i}{12} \right)}$$

$$\sigma_U = \sqrt{\left( \frac{10 \times 25}{35(35-1)} \right) \left( \frac{35^3 - 35}{12} - 35.5 \right)} = \sqrt{0.21 \times 3534.5} = 27.24$$

$$\sigma_U = 27,24$$

## Test de la U de Mann-Whitney (paso 3d)

- Una vez calculados todos los parámetros necesarios, se obtiene el valor Z conforme la fórmula anterior:

$$Z = \frac{U - \bar{U}}{\sigma_U}$$
$$Z = \frac{178 - 125}{27,24} = 1,95$$
$$Z = \frac{72 - 125}{27,24} = -1,95$$

U = cualquier valor de U calculado (ya sea  $U_1$  o  $U_2$ ).

## Test de la U de Mann-Whitney (paso 4)

- Para obtener la probabilidad (valor de la p) del valor  $Z_{\text{calculado}}$  de 1.95 (1,9 + 0,05), se debe consultar la tabla de tamaño de la muestra en función de los valores d y buscar la hilera 1.9, en cuya columna correspondiente al error alfa de 0,05 se localiza el número 0,9788, cuyo complementario ( $1 - 0,9744 = 0,0256$ ) corresponde a la probabilidad del valor de U con respecto al promedio ( $p = 0,0256$ ). Esto quiere decir que es menor que el nivel de significancia ( $p < 0,05$ ).

**TABLA D (continuación)** Valores de la función de distribución acumulativa normal estándar

z	.00	.01	.02	.03	.04	.05	.06	.07
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9907	0.9909	0.9911
2.4	0.9918	0.9920	0.9922	0.9924	0.9926	0.9928	0.9931	0.9932
2.5	0.9938	0.9940	0.9941	0.9943	0.9944	0.9946	0.9948	0.9949
2.6	0.9953	0.9954	0.9955	0.9956	0.9957	0.9958	0.9961	0.9962
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985

$$p = [ 1 - 0,9744 ] = 0,0256$$

## Decisión

A la cifra de Z de 1,95 le corresponde una probabilidad menor que 0,05 ( $p = 0,0256$ ), por lo cual se acepta  $H_a$  y se rechaza  $H_0$  (tabla de probabilidades asociadas en valores extremos como los de 2 en la distribución normal).

## Interpretación

El experimentador (profesor de bioestadística de los cursos), al aumentar el tamaño muestral en ambos grupos ( $n_1=10$ , y  $n_2=25$ ,  $N=35$ ), CONFIRMA la investigación preliminar con una muestra pequeña ( $n_1=5$ , y  $n_2=5$ ,  $N=10$ ), con lo cual da a entender que los resultados (calificaciones de los alumnos) logrados con el método PRESENCIAL son diferentes de los obtenidos con el método de enseñanza A DISTANCIA; además, este último revela calificaciones más bajas y permitiría concluir que es menos efectivo que el otro.

## Pregunta de investigación

¿Qué relación existe entre el grupo cultural/religioso y los niveles de anticuerpos frente al virus de la Rubeola en la población gestante inmunizada de Ceuta en el año 2010?

Variable PREDICTORA

Cualitativa dicotómica

Cristianas inmunizadas  
Musulmanas inmunizadas



Variable RESULTADO

Cuantitativa continua

Virus de la Rubeola

¿Cuál es el test estadístico más adecuado?

	SEXO	N	Mean Rank	Sum of Ranks
Anti-IgG-RUBEOLA	Cristianas	234	392,46	91836,0
	Musulmanas	487	345,88	168445,0
	Total	721		

Test estadísticos	IgG-Rubeola
Mann-Whitney U	49573,00
Wilcoxon W	168401,0
Zu	- 2,828
Significación a dos colas (bilateral)	0,005
Significación exacta a una cola	0,043

$$\text{Mean rank} = \frac{\text{Sum of ranks}}{N}$$

$$Z_u = \frac{U - R}{\frac{\sigma}{\sqrt{n}}}$$

El ranking promedio de los niveles de Anti-IgG-RUBEOLA en gestantes inmunizadas cristianas superior al de las musulmanas, es decir existen diferencias significativas entre ambos grupos culturales, y por tanto como  $p < 0,05$  se rechaza la  $H_0$ .

Conclusión: El grado de inmunización frente a la rubeola es diferente en cristianas y musulmanas, siendo mayor en las primeras.

## Pregunta de investigación

¿Qué relación existe entre el grupo cultural/religioso y los niveles de anticuerpos frente a *Toxoplasma* en la población gestante inmunizada de Ceuta en el año 2010?

Variable PREDICTORA  
Cualitativa dicotómica

Cristianas inmunizadas  
Musulmanas inmunizadas



Variable RESULTADO  
Cuantitativa continua

*Toxoplasma*

¿Cuál es el test estadístico más adecuado?

Test de normalidad	Test de Kolmogorov-Smirnov				
Año 2010	N	Media	DE	Estadístico Z	Significación
Niveles de anticuerpos IgG anti-toxoplasma	297	77,5	133,6	5,321	0,000

Test de normalidad	Test de Kolmogorov-Smirnov					
Año 2010	Tipo de cultura	N	Media	DE	Estadístico Z	Significación
Niveles de anticuerpos IgG anti-toxoplasma	Cristianas	32	67,9	58,6	1,260	0,084
	Musulmanas	265	78,62	139,9	5,110	0,000

$H_0: p > 0,05$  test paramétricos

$H_a: p < 0,05$  test NO paramétricos

Los test de normalidad ensayados dan un valor de p significativo ( $p < 0,05$ ) en ambos grupos culturales. Por tanto hay evidencias suficientes para rechazar la hipótesis nula de normalidad de la variable reducción del peso en ambas muestras. NO PODEMOS USAR TEST PARAMÉTRICOS.

	SEXO	N	Mean Rank	Sum of Ranks
Anti-IgG-toxoplasma	Cristianas	32	162,11	5187,50
	Musulmanas	265	147,42	39065,50
	Total	297		

Test estadísticos	IgG-toxoplasma
Mann-Whitney U	3820,50
Wilcoxon W	39065,50
Zu	- 0,914
Significación a dos colas (bilateral)	0,361

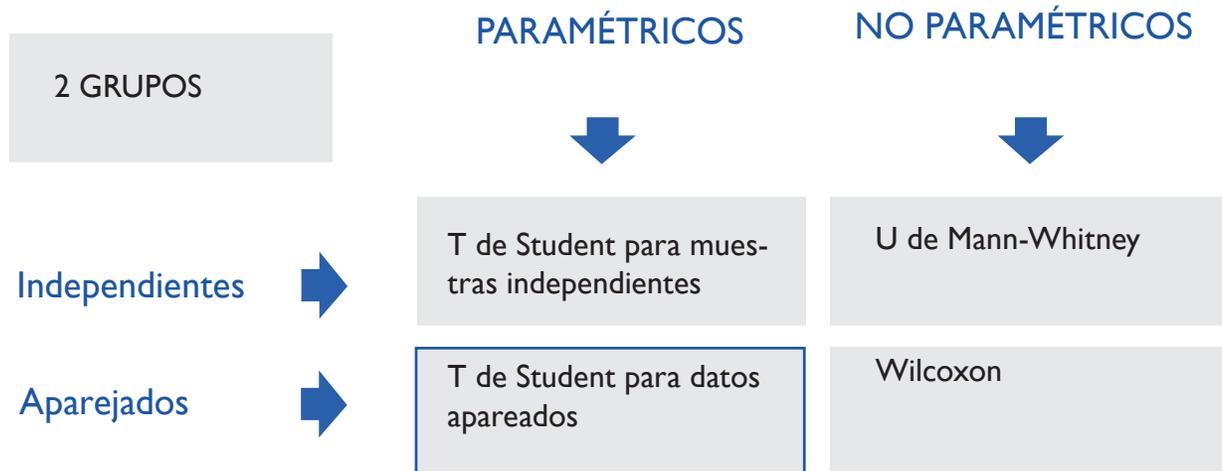
$$\text{Mean rank} = \frac{\text{Sum of ranks}}{N}$$

$$Z_u = \frac{U - R}{\frac{\sigma}{\sqrt{n}}}$$

El ranking promedio de los niveles de Anti-IgG-toxoplasma en gestantes inmunizadas cristianas similar al de las musulmanas, es decir NO existen diferencias significativas entre ambos grupos culturales, y por tanto como  $p > 0,05$  se acepta la  $H_0$  (igualdad).

Conclusión: El grado de inmunización (nivel de anticuerpos específicos) frente a Toxoplasma es igual cristianas y musulmanas.

## Cualitativa dicotómica versus cuantitativa



Variable predictora Cualitativa versus variable resultado Cuantitativa.

## Variables cuantitativas:

### Comparación de la media entre dos grupos

Muestras independientes	
Tipos de dieta	Peso (Kg)
Pescado	23
Pescado	19
Pescado	28
Pescado	23
Pescado	5
Pescado	17
Pescado	14
Pescado	13
.....	
Huevos	21
Huevos	22
Huevos	26
Huevos	25
Huevos	2
Huevos	28
Huevos	17
Huevos	4
Huevos	12
.....	

Muestras relacionadas		
Peso inicial	Peso final (Kg)	DIF
23	21	- 2
19	22	+ 3
28	26	- 2
23	25	+ 2
5	2	- 3
17	28	+ 11
14	17	+ 3
13	4	- 9
.....	.....	

Se compara cada individuo consigo mismo antes y después de un tratamiento con una determinada dieta.

## Test de la t de Student para datos apareados (muestras relacionadas)

Características del test:

- Se realizan mediciones repetidas en un mismo individuo.
  - Se aplican dos fármacos consecutivos a un mismo paciente.
  - Se desea evaluar el cambio de una variable cuantitativa tras aplicar un tratamiento (cambio entre antes y después).
    - Calcular una nueva variable que es la diferencia entre los valores antes y después de cada individuo.
    - Cuando el tamaño muestral es inferior al 30, se debe probar si la nueva variable sigue una distribución normal.
  
- En las muestras relacionadas la variabilidad ALEATORIA es menor que en las muestras independientes.

## Test de la t de Student para muestras relacionadas

Pregunta de investigación: ¿Existen variación del peso de la madre antes y después del embarazo?

Hipótesis nula  $H_0$   $X_{\text{antes}} = X_{\text{después}}$  ( DIF = POST – FREC = 0)

Hipótesis alternativa  $H_a$   $X_{\text{antes}} \neq X_{\text{después}}$  ( DIF = POST –FREC  $\neq$  0)

$$t_{\text{calculada}} = \frac{\bar{d} - 0}{\sqrt{\frac{S^2}{n}}}$$

	Variables		
Estadísticos	Peso antes del embarazo	Peso después del parto	Diferencia de peso
Media	54,26	55,64	1,38
Desviación estándar (S)	6,653	7,624	4,096
n	1626	1626	1626

Variables	Media	N	Desviación estándar (s)	Error estándar de la media (EEM)
Peso ANTES embarazo	54,26	1626	6,653	0,165
Peso DESPUÉS del parto	55,64	1626	7,624	0,189

Variable	N	Correlación (r)	Significación
Peso ANTES embarazo versus Peso DESPUÉS del parto	1626	0,844	0,000

Variables	Diferencias relacionadas			Intervalo de confianza de la diferencia		$t_{calculada}$	gl	Significación bilateral
	Media	s	EEM	Inferior	Superior			
Peso después del parto – peso antes del embarazo	1,378	4,096	0,102	1,178	1,577	13,562	1625	0,000

$$t_{calculada} = 13,56 > t_{tabulado} = 1,96; \quad p < 0,001$$

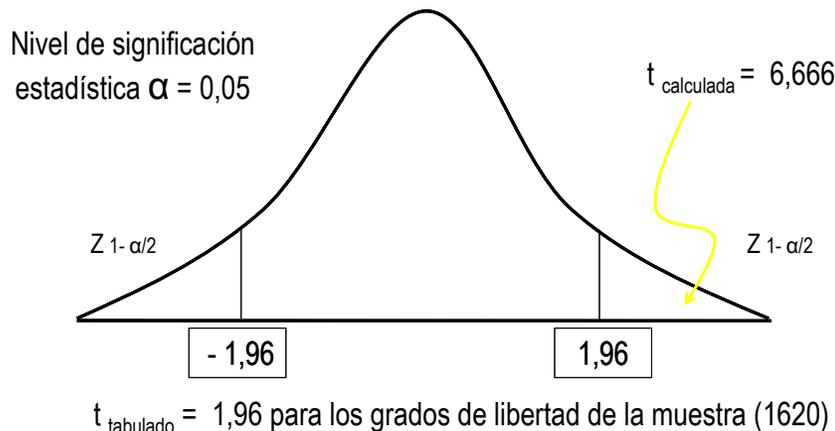
$$t_{calculada} > t_{tabulada} \quad \rightarrow \quad p < 0,05 \text{ (rechazamos la hipótesis nula)}$$

$H_a$ : Existe variación en el peso de las madres antes y después de un embarazo.

## Distribución de la t de Student ( $gl \geq 30$ )

El test de la T de Student para muestras de más de 30 individuos presenta una distribución muy similar a la distribución normal tipificada Z.

El test de la T de Student es una prueba BILATERAL, y por tanto NO especifica la dirección de la diferencia ( $A \neq B$ ).



## Test de la t de Student para datos apareados (muestras relacionadas)

### Ejemplo

Estudio sobre la forma física de un grupo de 30 individuos antes y después de un programa de entrenamiento deportivo (PED).

Pregunta de investigación:

¿Ha mejorado la forma física del grupo durante las sesiones del PED?

¿Reduce el PED la frecuencia cardíaca media?

Variable PREDICTORA

Cualitativa dicotómica

PED



Variable RESULTADO

Cuantitativa

Frecuencia Cardíaca Media

Objetivo: Conocer si el entrenamiento deportivo (PED) mejora la forma física de un grupo de individuos.

Material: Grupo de 30 individuos antes y después de un programa de entrenamiento deportivo (PED)

Método: Se evalúa la frecuencia cardiaca antes (FRECPRE) y después del PED (FRECPOST).

Se define una nueva variable  $DIF = FRECPRE - FRECPOST$ .

Hipótesis nula  $H_0$   $X_{\text{antes}} = X_{\text{después}}$   
(  $DIF = FRECPRE - FRECPOST =$

Hipótesis alternativa  $H_1$   $X_{\text{antes}} \neq X_{\text{después}}$   
(  $DIF = FRECPRE - FRECPOST \neq 0$  )

## Resultados

Variables	Media	N	Desviación estándar (s)	Error estándar de la media (EEM)
FRECPRE	79,53	30	19,15	3,50
FRECPOST	70,40	30	12,95	2,36

Variables	N	Correlación (r)	Significación
FRECPRE versus FRECPOST	30	0,828	0,000

Variables	Diferencias relacionadas			Intervalo de confianza de la diferencia		$t_{\text{calculada}}$	gl	Significación bilateral
	Media	s	EEM	Inferior	Superior			
FRECPRE - FRECPOST	9,13	11,12	2,031	4,98	13,29	4,49	29	0,000

$t_{\text{calculada}} < t_{\text{tabulada}} \rightarrow p > 0,05$  (aceptamos la  $H_0$ )

$t_{\text{calculada}} > t_{\text{tabulada}} \rightarrow p < 0,05$  (rechazamos la hipótesis nula)

## Test de la t de Student para datos apareados (muestras relacionadas)

### Ejemplo - Conclusiones

1. Puede concluirse que ha existido un descenso significativo ( $p < 0,001$ ) de la frecuencia cardíaca después del PED. Existen evidencias, con una confianza del 95%, para afirmar que el PED se ha acompañado de un descenso de la frecuencia cardíaca .
2. El descenso medio estimado fue de 9 latidos por minuto después del PED.

## Mito I: “La sonrisa de la esfinge”

La t de Student de datos apareados es un test paramétrico, es decir para usarlo se debe cumplir la normalidad de la variable cuantitativa.

¿Se cumple? ¿La nueva variable  $DIF = FRECPRE - FRECPPOST$  sigue una distribución normal?

He observado en tus resultados de DIF un valor de la desviación típica (11,1) muy superior a la media (9,1),

¿Cómo justificas esa supuesta normalidad? ¿Qué test has empleado para comprobar esa supuesta normalidad?

La frecuencia cardiaca es una variable cuantitativa DISCRETA.

¿Por qué empleas un test paramétrico para este tipo de variables? ¿No sería más adecuado utilizar el test de WILCOXON?

Objetivo: Comprobar si la reducción ponderal conseguida tras intervención dietética en los pacientes con obesidad mórbida disminuye los factores de riesgo cardiovascular.

Prueba de la t de Student para datos apareados

Prueba de Shapiro-Wilk (test de normalidad)

IMC > 40

	Inicio	Final	p	IC (95%)
Peso (Kg)	149,1 (22)	115 (18)	← < 0,001	26,9-39,8
IMC (Kg/m <sup>2</sup> )	49,2 (7,7)	38,8 (5)	← < 0,001	7,8-13,2
Cintura (cm)	144,4 (17,3)	119,6 (12,1)	← < 0,001	18,4-36,5
Cadera (cm)	143,3 (17,5)	127,7 (14,6)	← 0,02	9,1-24,1
Indice C/c	1,01 (0,12)	0,94 (0,08)	← 0,02	0,017-0,16

Los valores se expresan como media (desviación estándar); p nivel de significación; IC intervalo de confianza; IMC: índice de masa corporal; C/c cintura-cadera.

Objetivo: Comprobar si la reducción ponderal conseguida tras intervención dietética en los pacientes con obesidad mórbida disminuye los factores de riesgo cardiovascular.

Prueba de la t de Student para datos apareados

IMC > 40

	Inicio	Final	p	IC (95%)
Peso (Kg)	126 (19)	107 (18)	← < 0,001	13,5-20,3
IMC (Kg/m <sup>2</sup> )	48,7 (5,8)	42,5 (6,2)	← < 0,001	4,5-7,5
Cintura (cm)	129,8 (17,7)	118,7 (15,4)	← < 0,001	9,1-21,6
Cadera (cm)	142,5 (15,6)	132,8 (13)	← 0,003	3,6-15,9
Índice C/c	0,92 (0,11)	0,89 (0,09)	← 0,04	0,017-0,16

Los valores se expresan como media (desviación estándar); p nivel de significación; IC intervalo de confianza; IMC: índice de masa corporal; C/c cintura-cadera

La reducción ponderal se acompañó de una mejora del perfil lipídico (triglicéridos, cHDL y apoproteína B<sub>100</sub> principalmente en los varones.

Varones	Inicio	Final	p	IC (95%)
Colesterol total (mg/dl)	192 (49,7)	179,7 (34,5)	0,16	-5,1 a 29,6
Triglicéridos (mg/dl)	162,4 (73,6)	107,7 (33,7)	◀ 0,001	25,4 - 90,0
cLDL (mg/dl)	124,1 (38,5)	115 (30,4)	0,12	-12,9 a 5,7
cHDL (mg/dl)	35,4 (86,3)	43,5 (8,5)	◀ < 0,001	5,7 - 12,6
cVLDL (mg/dl)	32,5 (14,7)	21,9 (6,8)	◀ 0,002	4,8 - 18,5
Apo B <sub>100</sub> (mg/dl)	122,1 (31,3)	100,7 (19,6)	◀ 0,01	6,5 - 40,5
Apo A <sub>1</sub> (mg/dl)	117,8 (13,2)	126 (18,8)	0,24	- 17,5 a 4,9

Los valores se expresan como media (desviación estándar); p nivel de significación; IC intervalo de confianza;

La reducción ponderal se acompañó de una mejora del perfil lipídico (triglicéridos, cHDL y apoproteína B<sub>100</sub> principalmente en los varones.

Mujeres	Inicio	Final	p	IC (95%)
Colesterol total (mg/dl)	189 (29,2)	183,6 (32)	0,32	-5,7 a 16,5
Triglicéridos (mg/dl)	137 (60,8)	109,5 (43,7)	0,22	-5,8 a 23,5
cLDL (mg/dl)	117,8 (13,2)	115,6 (26,4)	0,07	-0,8 a 18,5
cHDL (mg/dl)	41,9 (15,6)	45,6 (13,1)	0,03	0,4 – 6,6
cVLDL (mg/dl)	32,3 (15,9)	22 (8,4)	0,2	-1,1 a 4,8
Apo B <sub>100</sub> (mg/dl)	110 (30,6)	106,1 (20,6)	0,7	-11,4 a 15,9
Apo A <sub>1</sub> (mg/dl)	124,1 (23,2)	129 (21,4)	0,052	-0,07 a 19,4

Los valores se expresan como media (desviación estándar); p nivel de significación; IC intervalo de confianza;

## Discusión - Conclusiones

- Los varones presentaron una mayor reducción ponderal que las mujeres.
- Con la intervención dietética empleada conseguimos una mejora en el perfil aterogénico, principalmente en los triglicéridos, VLDL y HDL.
- Los varones presentaron una mejoría más importante que las mujeres en el perfil lipídico.
- La reducción ponderal ha conseguido una reducción de algunos de los factores de riesgo cardiovascular (FRCV) analizados.
- El sexo masculino presenta una mejor respuesta global con reducción de los FRCV y mejora de la distribución grasa.

## Mito I: “La sonrisa de la esfinge”

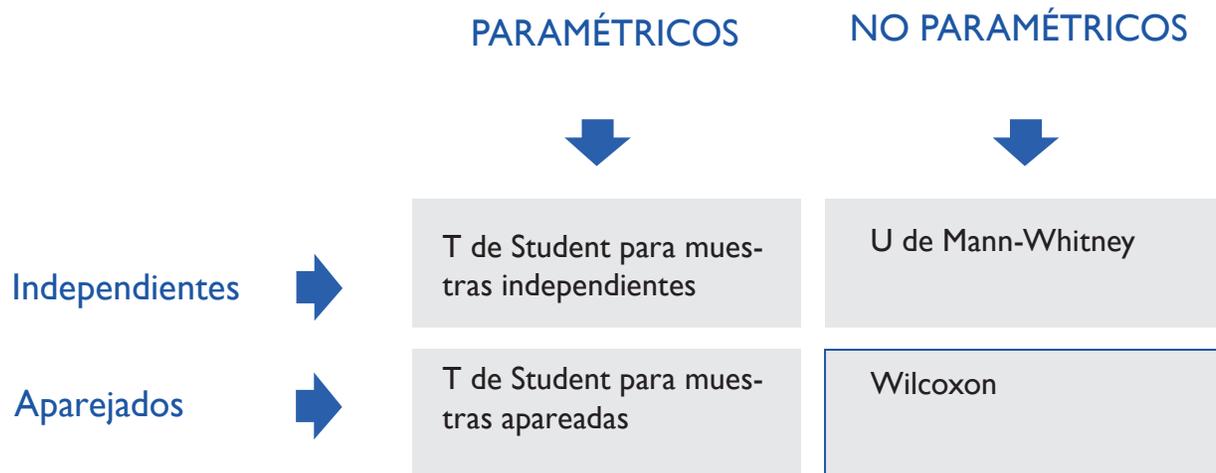
La t de Student de datos apareados es un test paramétrico, es decir para usarlo se debe cumplir la normalidad de las variables cuantitativas.

¿Se cumple en todas las variables? ¿Las nuevas variables DIF = FRECPRE – FRECPST siguen una distribución normal? ¿Cómo justificas esa supuesta normalidad? ¿Que test has empleado?

He observado que en tus resultados NO aparece las nuevas variables DIF ni su desviación típica.

¿Porqué no se ha incluido en los resultados?

## Cualitativa versus cuantitativa



Variable predictora Cualitativa versus variable resultado Cuantitativa.

## Test de Wilcoxon

Prueba NO PARAMÉTRICA o de libre distribución

- No se basa en la media, desviación estándar, etc.

Condiciones de aplicación:

- Diseño emparejado o de medidas repetidas
- Datos a comparar son  
ORDINALES o  
Cuantitativos con BAJO tamaño muestral ( $n < 30$ ) y  
NO siguen una distribución NORMAL

## Ejemplo (muestras relacionadas)

### Test de Wilcoxon para datos apareados

Estudio sobre la mejoría clínica del dolor en pacientes con artrosis tratados con un infiltración de corticoides.

Pregunta de investigación:

¿Ha mejorado el grado de bienestar de los pacientes con artrosis tras el tratamiento con infiltraciones?

¿Existen evidencias de una mejoría en el nivel de dolor en los pacientes con artrosis después del tratamiento con infiltraciones intrarticulares?

Nivel de dolor ANTES  
del tratamiento



Infiltración

Nivel de dolor DESPUÉS  
del tratamiento

Objetivo: Conocer si la infiltración con corticoides mejora el bienestar clínico de en un grupo de los pacientes con artrosis de rodilla...

Material: Grupo de pacientes con artrosis de rodilla del Hospital X seleccionados para el ensayo.

Método: Se evalúa el nivel de dolor en los pacientes mediante un test estandarizado antes (NDA) y después del tratamiento infiltrativo (NDD). Se define una nueva variable  $DIF = NDA - NDD$ .

Hipótesis nula  $H_0$      $X_{\text{antes}} = X_{\text{después}}$     ( $DIF = NDA - NDD = 0$ )

$H_0$  (hipótesis nula) = La infiltración con corticoides NO mejora el bienestar clínico de los pacientes con artrosis de rodilla.

Hipótesis alternativa  $H_1$      $X_{\text{antes}} \neq X_{\text{después}}$     ( $DIF = NDA - NDD \neq 0$ )

$H_a$  (hipótesis alternativa) = La infiltración con corticoides SI mejora el bienestar clínico de los pacientes con artrosis de rodilla.

## Test de $\chi^2$ de tendencia lineal

Ejercicio: Realizamos un estudio hipotético de cohortes recogiendo datos sobre la actividad deportiva que practicaban en su juventud 10 pareja de gemelos de Ceuta, donde después de un seguimiento de 20 años, uno de los gemelos ha muerto y el otro sobrevive.

Variable predictora:

Actividad deportiva

Variable cuantitativa ordinal

S= sedentarios

O= deporte ocasional

R= deporte regular

C = deporte competitivo

E = deporte de elite

Variable resultado:

Pareja	Gemelo vivo	Gemelo muerto
1 <sup>a</sup>	S	S
2 <sup>a</sup>	E	S
3 <sup>a</sup>	R	E
4 <sup>a</sup>	R	E
5 <sup>a</sup>	C	S
6 <sup>a</sup>	R	E
7 <sup>a</sup>	E	R
8 <sup>a</sup>	E	R
9 <sup>a</sup>	C	E
10 <sup>a</sup>	O	S

## Test de Wilcoxon para datos apareados

Ejercicio: Realizamos un estudio hipotético de cohortes recogiendo datos sobre la actividad deportiva que practicaban en su juventud 10 pareja de gemelos de Ceuta, donde después de un seguimiento de 20 años, uno de los gemelos ha muerto y el otro sobrevive.

Variable predictora:

Actividad deportiva

Variable cuantitativa ordinal

0 = sedentarios

1= deporte esporádico

2= deporte regular

3= deporte competitivo

4= deporte de élite

Variable resultado:

Pareja	Gemelo vivo	Gemelo muerto
1 <sup>a</sup>	0	0
2 <sup>a</sup>	1	0
3 <sup>a</sup>	2	1
4 <sup>a</sup>	2	1
5 <sup>a</sup>	4	0
6 <sup>a</sup>	2	1
7 <sup>a</sup>	1	2
8 <sup>a</sup>	1	2
9 <sup>a</sup>	3	1
10 <sup>a</sup>	2	0

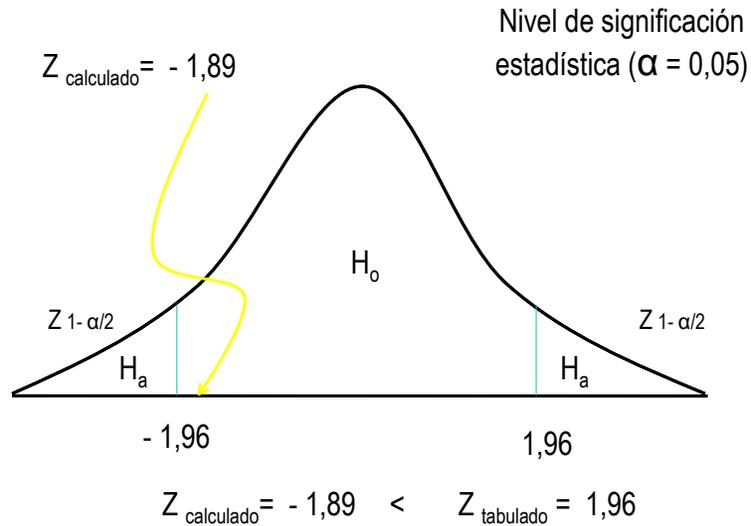
## Prueba de los rangos con signo de WILCOXON para datos apareados

Estadístico de contraste	DESENLACE - TIPO DE DEPORTE
Z	- 1,897
Significación asintomática (bilateral)	0,058

## Prueba de los signos para datos apareados

Estadístico de contraste	DESENLACE - TIPO DE DEPORTE
Significación asintomática (bilateral)	0,180

El test de Wilcoxon es una prueba BILATERAL, y por tanto NO especifica la dirección de la diferencia ( $A \neq B$ ).



Ho = La práctica deportiva en la juventud NO reduce la mortalidad

Ha = La práctica deportiva en la juventud REDUCE la mortalidad

$$z = \frac{W - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \sim N(0,1)$$

$$z = \frac{W - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24} - \sum_{i=1}^g \frac{t_i^3 - t_i}{2}}} \sim N(0,1)$$

1. Cálculo del estadístico z:  $Z_{\text{calculado}} = -1,89$
2. Comparación del  $Z_{\text{calculado}}$  con el  $Z_{\text{tabulado}}$  de las tablas de la normal.  $Z_{\text{calculado}} = -1,89 < Z_{\text{tabulado}} = -1,96$  para  $p < 0,05$  a dos colas. La significación estadística  $p = 0,058$
3. No se dispone de evidencias para rechazar la Ho de igualdad de medias entre el gemelo que sobrevivió y el que murió. La probabilidad de hallar unas diferencias al menos como las encontradas es de sólo el 5,8%. Las diferencias encontradas pueden ser debidas al AZAR.
4. No se puede rechazar Ho, ya que el valor de p es superior a 0,05. Estos datos NO proporcionan evidencias que indiquen que la práctica deportiva era superior en los gemelos supervivientes respecto a sus hermanos gemelos fallecidos. El estudio NO apoya que el deporte en la juventud reduzca la mortalidad en el futuro.
5. ERROR de DISEÑO: falta de potencia estadística por su escaso tamaño muestral. ERROR BETA.

	Desenlace (E)	Variable resultado
Exposición (F) Variable predictora	Muerte	Sobreviviente
Sedentarismo (+)	A	B
Sedentarismo (-)	C	D

A: Fallecidos con sedentarismo

B: Sobrevivientes sedentarios

C: Fallecidos deportistas

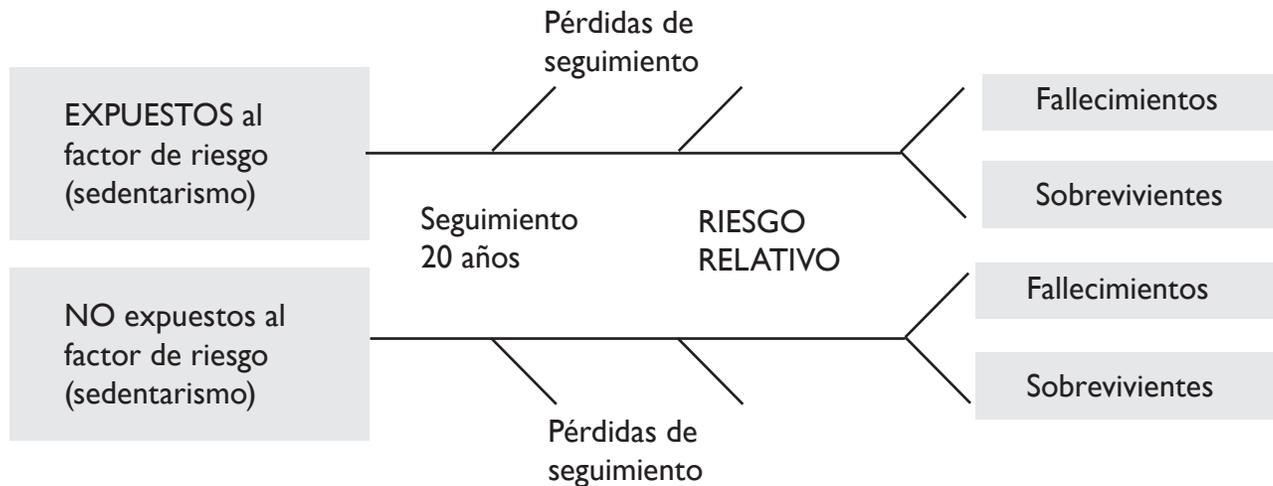
D: Sobrevivientes deportistas

$$RR = \frac{[ A / (A + B) ]}{[ C / (C + D) ]}$$

Es la probabilidad de morir (E) en el grupo de sedentarios (expuestos al factor de riesgo) en relación con la probabilidad de muerte en deportistas (grupo no expuesto).

RR = 1.41 IC95%=1.23-1.61 de los sedentarios frente a los para los “ocasionalmente activos”

RR = 1.52 IC95%=1.06-2.17 de los sedentarios frente a los “entrenados”



Kumala et al. Relationship of leisure-time physical activity and mortality: the Finnish twin cohort. JAMA 1998; 11; 279: 440-4.

	Desenlace (E)	Variable resultado
Exposición (F) Variable predictora	Muerte	Sobreviviente
Deporte (+)	A	B
Deporte (-)	C	D

A: Fallecidos con sedentarismo

B: Sobrevivientes sedentarios

C: Fallecidos deportistas

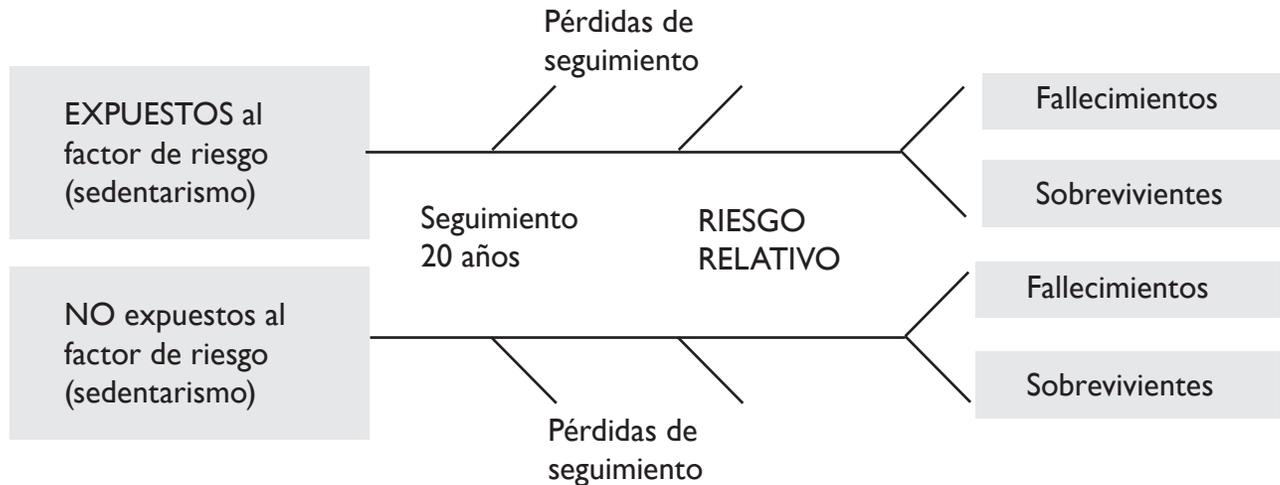
D: Sobrevivientes deportistas

$$RR = \frac{[ A / (A + B) ]}{[ C / (C + D) ]}$$

Es la probabilidad de morir (E) en el grupo de deportistas (expuestos al factor de protección) en relación con la probabilidad de muerte en sedentarios (grupo no expuestos).

RR = 0.71 IC95%=0.62-0.81 para los “ocasionalmente activos”

RR = 0.66 IC95%=0.46-0.94 para los “entrenados”



Kumala et al. Relationship of leisure-time physical activity and mortality: the Finnish twin cohort. JAMA 1998; 11; 279: 440-4.

- Diseño: estudio de COHORTES
  - 7925 varones gemelos
  - 7977 mujeres gemelas
- Resultados:
  - RR = 0,71: Se encontró una reducción del 29% en la mortalidad para los “ocasionalmente activos” comparados con los “sedentarios”.
  - RR = 0,66: Se encontró una reducción del 34% en la mortalidad para los “entrenados” comparados con los “sedentarios”.

$$RRR = [ 1 - RR ] * 100$$

$$RRR_1 = 29\%$$

$$RRR_2 = 34\%$$

- Conclusiones: Este estudio es importante porque deja claro que la mejor supervivencia de los más activos NO se debe a una predisposición genética (controlada por el emparejamiento entre gemelos), sino al efecto real del ejercicio físico sobre la supervivencia.

Kumala et al. Relationship of leisure-time physical activity and mortality: the Finnish twin cohort. JAMA 1998; 11; 279: 440-4.

## Ejemplo (muestras relacionadas)

Estudio sobre los conocimientos adquiridos en Metodología de Investigación en Ciencias de la Salud (MICS) en los alumnos del curso de Sevilla.

Pregunta de investigación:

¿Han mejorado los conocimientos adquiridos en Metodología de Investigación en los alumnos del curso de Sevilla?

¿Aumenta el curso el nivel de conocimientos sobre la materia impartida?

¿Existen evidencias de una mejoría en los conocimientos teóricos de Metodología de investigación en los alumnos después del curso?

Conocimientos de MICS  
previos al curso



Curso

Conocimientos de MICS  
después del curso

Objetivo: Conocer si el curso de Metodología de Investigación (CMI) realizado en Sevilla mejora los conocimientos sobre dicha materia en un grupo de alumnos (licenciados en Ciencias de la Salud).

Material: Grupo de 8 individuos FEA y residentes del Hospital Universitario de Valme (Sevilla ) antes y después de realizar el curso de Metodología de Investigación en Ciencias de la Salud (CMICS)

Método: Se evalúa los conocimientos mediante un test (examen) antes (RCAC) y después del CMICS (RCDC). Se define una nueva variable  $DIF = RCDC - RCAC$ .

Hipótesis nula  $H_0$   $X_{\text{antes}} = X_{\text{después}}$  ( $DIF = RCDC - RCAC = 0$ )

$H_0$  (hipótesis nula) = La acción formativa (CMICS) no mejora los conocimientos sobre metodología de investigación en los alumnos.

Hipótesis alternativa  $H_1$   $X_{\text{antes}} \neq X_{\text{después}}$  ( $DIF = RCDC - RCAC \neq 0$ )

$H_a$  (hipótesis alternativa) = La acción formativa (CMICS) mejora los conocimientos sobre metodología de investigación en los alumnos.

## ¿Existen evidencias de una mejoría en los conocimientos teóricos de metodología de investigación en los alumnos después del curso?

Nº del Alumno	Nº respuestas correctas ANTES del curso (RCAC)	Nº respuestas correctas DESPUÉS del curso (RCDC)	Diferencia de respuestas acertadas DESPUÉS – ANTES (RCDC- RCAC)
1ª	10	14	4
2ª	7	14	7
3ª	7	14	7
4ª	5	14	9
5ª	4	13	9
6ª	3	13	10
7ª	2	13	11
8ª	2	11	9

¿Cuál es el test estadístico más acertado?

## Prueba de los rangos con signo de WILCOXON para datos apareados

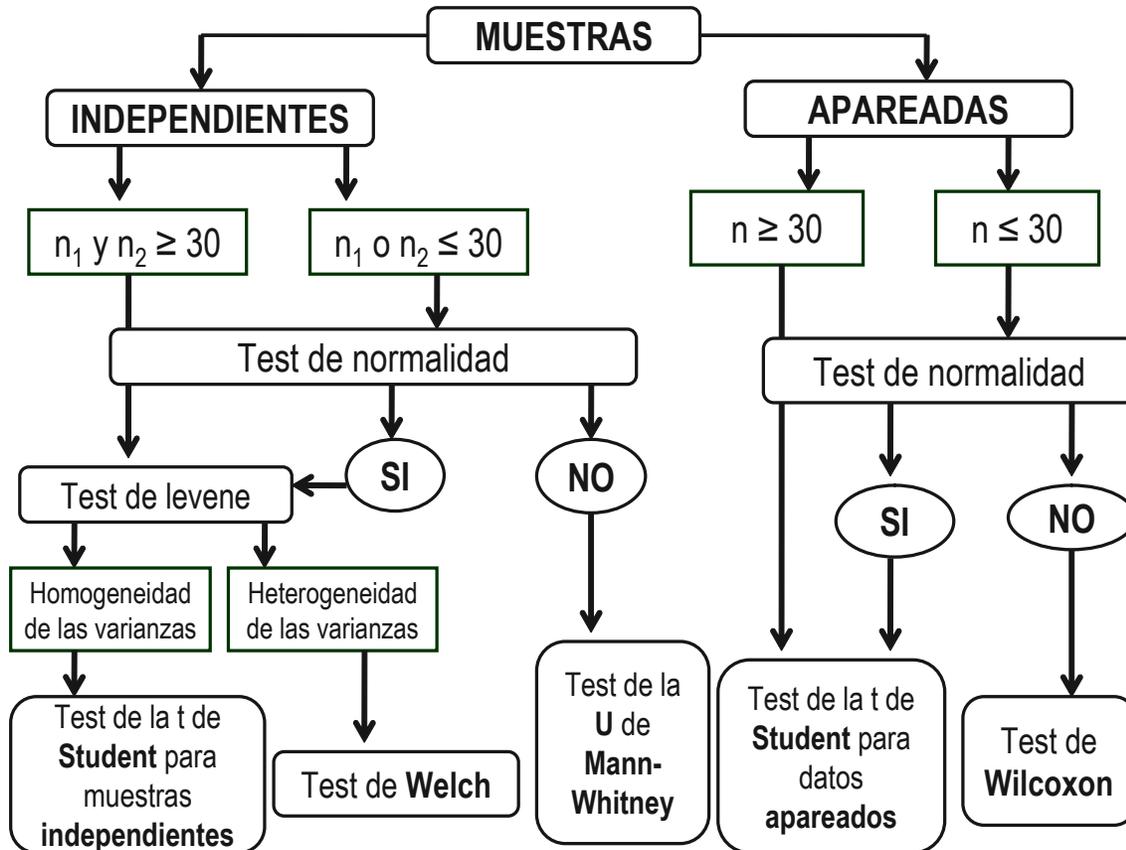
Estadístico de contraste	DESENLACE - TIPO DE DEPORTE
Z	-2,536
Significación asintomática (bilateral)	0,011

## Prueba de los signos para datos apareados

Estadístico de contraste	DESENLACE - TIPO DE DEPORTE
Significación asintomática (bilateral)	0,008

**RESULTADOS:** Se realizó la comparación del  $Z_{\text{calculado}}$  con el  $Z_{\text{tabulado}}$  de las tablas de la normal mediante el programa estadístico SPSS. El valor del estadístico  $Z_{\text{calculado}} = 2,536$  fue superior al  $Z_{\text{tabulado}} = 1,96$  para  $p < 0,05$  a dos colas (test de contraste bilateral). La significación estadística encontrada fue de  $p = 0,011$ . Estos datos permiten rechazar  $H_0$ , ya que el valor de  $p$  es inferior a  $0,05$ .

**CONCLUSIONES:** Puede concluirse que ha existido un aumento significativo ( $p < 0,05$ ) del número de respuestas acertadas después del curso. Existen evidencias, con una confianza del 95% (error alfa del 5%), para afirmar que la acción formativa ha producido una mejora en los conocimientos. Por tanto, estos datos proporcionan evidencias que indican una mejoría objetiva y significativa en los conocimientos teórico-prácticos de Metodología de investigación en los alumnos después del curso (hipótesis alternativa).



## Pregunta de investigación

¿Qué relación existe entre los niveles de colesterol-LDL en sangre y la probabilidad de padecer un IAM?

Variable PREDICTORA  
CUANTITATIVA continua

Niveles colesterol-LDL



Variable RESULTADO  
CUALITATIVA dicotómica

IAM

¿Cuál es el test estadístico más acertado?

## Pregunta de investigación

¿Qué relación existe entre los niveles de alcoholemia y la probabilidad de tener un accidente de tráfico?

Variable PREDICTORA  
CUANTITATIVA continua

Niveles de Alcoholemia



Variable RESULTADO  
CUALITATIVA dicotómica

Accidente de tráfico

¿Cuál es el test estadístico más acertado?

## Pregunta de investigación

¿Qué relación existe entre la ingesta crónica de alcohol y la probabilidad de desarrollar cáncer de esófago?

Variable PREDICTORA  
CUANTITATIVA continua

Ingesta crónica de alcohol



Variable RESULTADO  
CUALITATIVA dicotómica

Cáncer de esófago

¿Cuál es el test estadístico más acertado?

## Pregunta de investigación

¿Qué relación existe entre la edad gestacional y la probabilidad de muerte del recién nacido?

Variable PREDICTORA  
CUANTITATIVA continua

Edad gestacional



Variable RESULTADO  
CUALITATIVA dicotómica

Muerte del RN

¿Cuál es el test estadístico más acertado?

## Depósitos de Fe en médula ósea



Predice

VCM (hemograma)



Produce



Produce

Ferritina (sérica)



Predice

## Pregunta de investigación

¿Qué relación existe entre el VCM del hemograma y la probabilidad de tener déficit de los depósitos de hierro en médula ósea?

Variable PREDICTORA  
CUANTITATIVA continua

VCM  
(hemograma)



Variable RESULTADO  
CUALITATIVA dicotómica

Depósitos de Fe en médula ósea

¿Cuál es el test estadístico más acertado?

## Pregunta de investigación

¿Qué relación existe entre los niveles de ferritina sérica y la probabilidad de tener déficit de los depósitos de hierro en médula ósea?

Variable PREDICTORA  
CUANTITATIVA continua

Ferritina (sérica)



Variable RESULTADO  
CUALITATIVA dicotómica

Depósitos de Fe en médula ósea

¿Cuál es el test estadístico más acertado?

## Análisis estadístico

Tipos de test analíticos estadísticos de contraste de hipótesis según el tipo de variables implicadas en el estudio

Variable dependiente

Variable resultado

		Dicotómica	Continua
Variable Predictora	Dicotómica	Test de $\chi^2$	t de Student
Variable Independiente	Continua	Regresión logística simple	Regresión lineal

## Análisis estadístico

ANÁLISIS ESTADÍSTICO		Variable resultado	
		Dicotómica	Continua
Variable Predictora	Dicotómica	Test de $\chi^2$	t de Student
Variable Independiente	Continua	Regresión logística	Regresión lineal

REGRESIÓN LOGÍSTICA SIMPLE

## Análisis de Regresión Logística Simple

- Variable RESULTADO: Dicotómica
- Variable PREDICTORA:
  - Cuantitativa
  - Categórica

$$\log \frac{P}{1 - P} = \beta_0 + \beta_1 X$$

VARIABLES PREDICTORAS	VARIABLE RESULTADO
Alcoholemia	Accidente de tráfico
Ingesta de alcohol	Cáncer de esófago
Colesterolemia	IAM
Edad gestacional	Muerte del RN

## Pregunta de investigación

¿Qué relación existe entre los niveles de colesterol-LDL en sangre y la probabilidad de padecer un IAM?

Variable PREDICTORA  
CUANTITATIVA continua

Niveles colesterol-LDL



Variable RESULTADO  
CUALITATIVA dicotómica

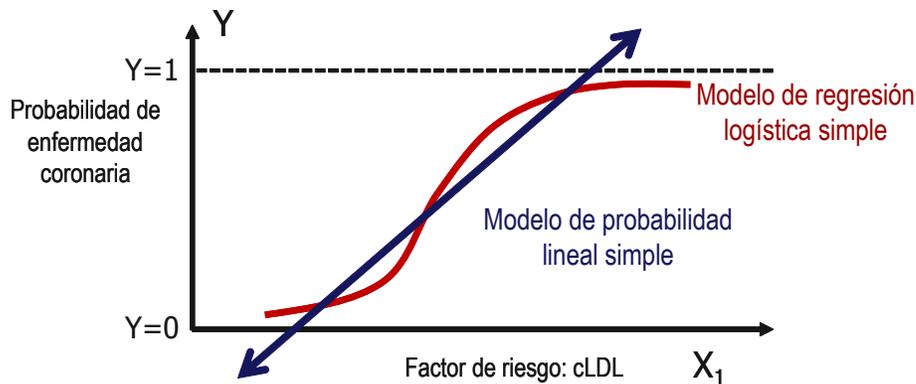
IAM

¿Cuál es el test estadístico más acertado?

## Modelo de Análisis de Regresión Logística Simple

$$P = \frac{e^{\beta_1 X_1}}{1 + e^{\beta_1 X_1}}$$

$$\log \frac{P}{1-P} = \beta_0 + \beta_1 X_1$$



## Análisis de regresión logística simple

$$\log \frac{P}{1 - P} = \beta_0 + \beta_1 X$$

$$\log \frac{P_1}{1 - P_1} = \beta_0 + \beta_1 x_1$$

$$\log \frac{P_2}{1 - P_2} = \beta_0 + \beta_1 x_2$$

$$\log \frac{P_1}{1 - P_1} - \log \frac{P_2}{1 - P_2} = \log \frac{\frac{P_1}{1 - P_1}}{\frac{P_2}{1 - P_2}} = \beta_1 (x_1 - x_2)$$

$$\frac{\frac{P_1}{1 - P_1}}{\frac{P_2}{1 - P_2}} = O.R.$$

$$\log(O.R.) = \beta_1$$

$$O.R. = e^{\beta_1}$$

## Pregunta de investigación

¿Qué relación existe entre la edad gestacional y la probabilidad de muerte del recién nacido?

Estudio caso-control realizado por Brand et al con 410 RN de edad gestacional entre 24-29 semanas.

Variable PREDICTORA  
CUANTITATIVA continua

Edad gestacional



Variable RESULTADO  
CUALITATIVA dicotómica

Probabilidad de morir

Análisis de Regresión Logística Simple

## Análisis de regresión logística simple

¿Qué relación existe entre la edad gestacional y la probabilidad de muerte del recién nacido?

Variable PREDICTORA  
Edad gestacional

Variable RESULTADO  
Riesgo de muerte del RN

	Edad gestacional						
Semanas	24	25	26	27	28	29	Total
Vivo	3	16	38	49	45	30	181
Muerto	26	56	69	50	21	7	229

$$\log \frac{P}{1 - P} = \beta_0 + \beta_1 X$$

$$\log \frac{P}{1 - P} = 18,36 - 0,6819 \text{ Edad}$$

Estimaciones de los parámetros del modelo logístico propuesto para los datos anteriores

Variable	Estimación	Error estándar
Constante	18,36	2,375
Edad gestacional	- 0,6819	0,08908

P = probabilidad que muera un RN

## Regresión logística. Coeficiente de regresión logística $\beta_1$

¿Existe relación entre la probabilidad de muerte neonatal (Variable Resultado) y la edad gestacional (Variable Predictora) en los RN de 24-29 semanas?

Z	Valor	$\alpha$
$Z_{95\%}$	1,96	95%
$Z_{99\%}$	2,56	99%
$Z_{99,5\%}$	3,29	99,5%

Estimaciones de los parámetros del modelo logístico propuesto para los datos anteriores

Variable	Estimación	Error estándar
Constante	18,36	2,375
Edad gestacional	- 0,6819	0,08908

Coeficiente de regresión logística estandarizado

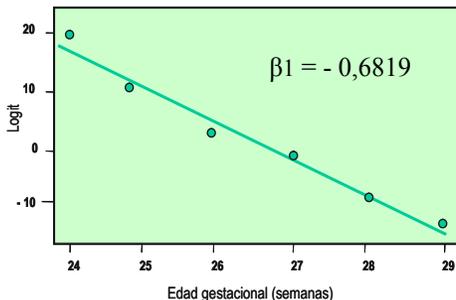
$$\frac{\beta_1}{EE\beta_1} = \frac{0,6819}{0,08908} = 7,65 > Z_{95\%}$$

Podemos decir con error inferior al 5% que existe asociación entre la mortalidad de los RN prematuros y su edad gestacional en el rango estudiado.

## Análisis de regresión logística simple

Estimaciones de los parámetros del modelo logístico propuesto para los datos anteriores

Variable	Estimación	Error estándar
Constante	18,36	2,375
Edad gestacional	- 0,6819	0,08908



Representación gráfica de los logits en las diferentes edades gestacionales estudiadas

$$\log \frac{P}{1 - P} = 18.36 - 0.6819 \text{ Edad}$$

$$O.R. = e^{\beta_1}$$

$$1/OR = 1,98$$

Un RN presenta casi 2 veces más riesgo de morir que otro RN con una semana más

$$OR = 0,506$$

Es el riesgo de morir de un RN respecto a otro que tiene una semana menos

¿Cuál es el riesgo de morir de un RN con 25 semanas en relación a otro de 29 semanas?

$$OR = e^{\beta_1 (X_1 - X_2)}$$

$$OR = e^{-0,6819 (25-29)} = 15,30$$

## Pregunta de investigación

¿Qué relación existe entre el VCM del hemograma y la probabilidad de tener déficit de los depósitos de hierro en médula ósea?

Variable PREDICTORA  
CUANTITATIVA continua

VCM  
(hemograma)



Variable RESULTADO  
CUALITATIVA dicotómica

Depósitos de Fe en médula ósea

¿Cuál es el test estadístico más acertado?

## Análisis de Regresión Logística Simple

¿Qué relación existe entre el VCM y los depósitos de hierro en médula ósea?

Variable PREDICTORA

Variable RESULTADO

VCM



Depósitos de Fe medular

Estimaciones de los parámetros del modelo logístico propuesto para los datos anteriores

Variable	Estimación	Error estándar
Constante	- 6,1112	
VCM	0,08485	0,02689

$$\log \frac{P}{1 - P} = \beta_0 + \beta_1 X$$

$$\log \frac{P}{1 - P} = -6,1112 + 0,08485 \text{ VCM}$$

P = probabilidad de que existan depósitos de hierro medular

## Análisis de Regresión Logística Simple

¿Qué relación existe entre la Ferritina sérica y los depósitos de hierro medular?

Variable PREDICTORA

Ferritina sérica

Variable RESULTADO

Depósitos de Fe medular



Estimaciones de los parámetros del modelo logístico propuesto para los datos anteriores

Variable	Estimación	Error estándar
Constante	- 2,1290	
FERRITINA	0,1713	0,04085

$$\log \frac{P}{1 - P} = \beta_0 + \beta_1 X$$

$$\log \frac{P}{1 - P} = - 2,1290 + 0,1713 \text{ FERRITINA}$$

P = probabilidad de que existan depósitos de hierro medular

## Tipos de test analíticos estadísticos bivariantes de contraste de hipótesis más frecuentes

Variable PREDICTORA o Independiente (factor de estudio)	Variable RESULTADO o Dependiente (Criterio de evaluación)			
	Cualitativa nominal dos categorías	Cualitativa nominal > 2 categorías	Cualitativa ordinal o Cuantitativa NO paramétrica	Cuantitativa paramétrica
Cualitativa dos grupos independientes	Chi cuadrado Z comparación de proporciones Test exacto de Fisher	Chi cuadrado	U de Mann-Whitney	T de Student (prueba de Welch)
Cualitativa dos grupos apareados	Test de McNemar	Q de Cochran	Prueba de los rangos de Wilcoxon	T de Student para datos emparejados
Cualitativa > 2 grupos independientes	Chi cuadrado	Chi cuadrado	Prueba de Kruskal-Wallis	Análisis de la Varianza (ANOVA)
Cualitativa > 2 grupos apareados	Q de Cochran	Q de Cochran	Prueba de Friedman	Análisis de la Varianza de dos vías
Cuantitativa	Regresión logística simple	Regresión logística multivariante	Correlación de Spearman Tau de Kendall	Regresión lineal Correlación de Pearson

## Capítulo 9

Publicación Bioestadística: Análisis bivariante.  
Test de contraste de hipótesis.  
Cualitativa policotómica versus cuantitativa.

Dr. Jacobo Díaz Portillo

Coordinador de Formación Continuada

Hospital Universitario de Ceuta

# Materiales y métodos

## Estructura

1. Tipo de diseño (estudio).
2. Sujetos incluidos (criterios de selección de la muestra)
3. Fuentes de información.
4. Descripción de la intervención y definición del punto final de la investigación (estudios experimentales).
5. Instrumentos y procedimientos de medida (cuestionarios).
6. Definición de las variables y covariables incluidas en el estudio (factor de estudio y criterio de evaluación).
7. Limitaciones y consideraciones éticas de la investigación.
8. Pruebas estadísticas a utilizar (programa estadístico).

## Fases del análisis estadístico

1. ANÁLISIS DESCRIPTIVO



2. ANÁLISIS BIVARIANTE



3. ANÁLISIS MULTIVARIANTE

## Pruebas estadísticas de contraste de hipótesis más aplicables a cada situación

Variable Independiente o PREDICTORA	Variable Dependiente o RESULTADO	Pruebas de contraste de hipótesis	Observaciones y/o limitaciones del test
Cualitativa	Cualitativa	Ji cuadrado Prueba exacta de Fisher Test de McNemar Q de Cochran REGRESIÓN LOGÍSTICA	Si el tamaño muestral es grande Si el tamaño muestral es pequeño Dos grupos de muestras repetidas o apareadas Más de dos grupos de muestras repetidas o apareadas Varias variables predictoras cualitativas
Cualitativa	Cuantitativa	t de Student Análisis de la Varianza Test de Mann-Whitney Test de Wilcoxon Test de Kruskal-Wallis Test de Friedman	Compara medias entre dos grupos independientes Compara medias entre más de dos grupos independientes Dos grupos independientes No paramétricos Dos grupos medidas repetidas No paramétrico Más de dos grupos No paramétrico, medidas repetidas Más de dos grupos No paramétrico medidas repetidas
Cuantitativa	Cualitativa	Regresión logística simple Regresión logística multivariante	Una sola variable predictora cuantitativa Varias variable predictoras cuantitativa o cualitativas
Cuantitativa	Cuantitativa	Regresión lineal simple Correlación de Pearson Correlación de Sperman REGRESIÓN MÚLTIPLE	Paramétrico Paramétrico No paramétrico Varias variables predictoras cuantitativas
Cualitativa	Supervivencia	Kaplan-Meier Long-Rank Regresión de COX	Curvas de supervivencia Compara curvas de supervivencia Análisis Multivariante

# Tipos de test analíticos estadísticos bivariantes de contraste de hipótesis más frecuentes

Variable PREDICTORA o Independiente (factor de estudio)	Variable RESULTADO o Dependiente (Criterio de evaluación)			
	Cualitativa nominal dos categorías	Cualitativa nominal > 2 categorías	Cualitativa ordinal o Cuantitativa NO paramétrica	Cuantitativa paramétrica
Cualitativa dos grupos independientes	Chi cuadrado Z comparación de proporciones Test exacto de Fisher	Chi cuadrado	U de Mann-Whitney	T de Student (prueba de Welch)
Cualitativa dos grupos apareados	Test de McNemar	Q de Cochran	Prueba de los rangos de Wilcoxon	T de Student para datos emparejados
Cualitativa > 2 grupos independientes	Chi cuadrado	Chi cuadrado	Prueba de Kruskal-Wallis	Análisis de la Varianza (ANOVA)
Cualitativa > 2 grupos apareados	Q de Cochran	Q de Cochran	Prueba de Friedman	Análisis de la Varianza de dos vías
Cuantitativa	Regresión logística	Regresión logística	Correlación de Spearman Tau de Kendall	Regresión lineal Correlación de Pearson

## Cualitativa versus cuantitativa

¿Cuáles son las pruebas estadísticas que tenemos que emplear?

Variable PREDICTORA  
CUALITATIVA dicotómica



Variable RESULTADO  
CUANTITATIVA

## Cualitativa policotómica versus cuantitativa



Variable predictora Cualitativa versus variable resultado Cuantitativa.

## Análisis de la varianza para un factor (ANOVA):

### Condiciones de aplicación

- Tipos de variables.
  - Variable resultado: Cuantitativa numérica.
  - Variable predictora: Cualitativa *policotómicas*.
  
- *Normalidad*: Variable resultado siga una distribución normal o de gauss.
  
- Homogeneidad de las varianzas (*homocedasticidad*): Las varianzas de cada grupo deben ser iguales.
  - Prueba F de Snedecor.
  - Test de Bartlett.
  - Test de Levene.

## Homogeneidad de varianzas (homocedasticidad)

- Prueba F de Snedecor
- Test de Bartlett
- Test de Levene



$$F = \frac{\text{Varianza mayor}}{\text{Varianza menor}}$$

$H_0$ : Las varianzas de ambos grupos son iguales → ANOVA

$H_a$ : Existen diferencias entre las varianzas → KRUSKAL-WALLIS

$F_{\text{calculada}} < F_{\text{tabulada}} \quad p > 0,05$  (aceptamos la  $H_0$ )

$F_{\text{calculada}} > F_{\text{tabulada}} \quad p < 0,05$  (rechazamos la hipótesis nula)

HIPERLIPIDEMIA

EDAD

SEXO

NIVEL CULTURAL

TABAQUISMO

OBESIDAD IMC

DIABETES

Excreción urinaria  
de sodio en 24 horas

# Test de la t de Student

## Análisis de la varianza para un factor

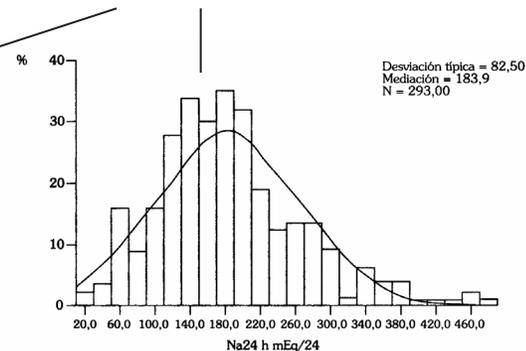
### Variables PREDICTORAS

**TABLA 2**  
**Excreción urinaria de sodio en los distintos subgrupos de pacientes**

Variable	Valores (mmEq/24 h)	P
Edad		0,002
< 40 años	208,62 ± 80,27	
40-65 años	185,81 ± 86,01	
> 65 años	158,12 ± 65,84	
Sexo		0,000
Hombres	205,19 ± 86,52	
Mujeres	169,05 ± 76,40	
IMC		0,103
Normal	150,80 ± 79,00	
Sobrepeso	184,18 ± 80,85	
Obesidad	193,02 ± 91,43	
Nivel cultural		0,011
Leer y escribir	164,90 ± 71,73	
Educación primaria	201,22 ± 87,40	
Educación secundaria	193,85 ± 90,59	
Estudios superiores	181,59 ± 78,43	
Tabaquismo		0,000
Sí	207,92 ± 80,90	
No	171,76 ± 80,82	
Diabetes		0,634
Sí	190,42 ± 102,99	
No	183,05 ± 79,80	
Hiperlipidemia		0,100
LDL > 150	172,99 ± 81,31	
LDL < 150	189,07 ± 79,67	

IMC: índice de masa corporal; LDL: lipoproteínas de baja densidad.

### Variable RESULTADO



¿Sigue una Distribución Normal?

Distribución ANORMAL

## Pregunta de investigación

¿Tienen la misma eficacia las dietas ensayadas para reducir el nivel de colesterol-LDL?

Variable PREDICTORA  
CUALITATIVA policotómica

- Tipos de dieta
- Dieta control
- Dieta moderada
- Dieta estricta



Variable RESULTADO  
CUANTITATIVA continua

LDL- colesterol

¿Cuál es la prueba estadística más adecuada?

Grupo Control (n=10)	Dieta moderada (n=10)	Dieta estricta (n=10)
+ 13	4	0
+ 10	8	12
+ 2	9	15
2	10	21
- 8	20	-10
+ 5	15	-18
3	6	-16
+ 8	7	-10
+ 2	16	-20
-2	6	-15
Media = + 2,5	Media = - 10,1	Media = - 13,7

## Análisis de la varianza para un factor

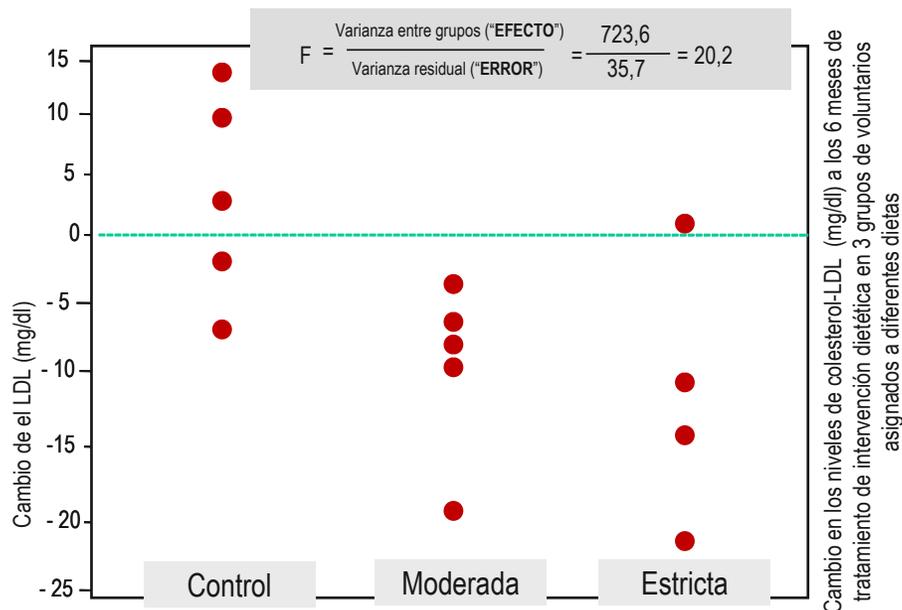
Hipótesis **nula**:  $\mu_1 = \mu_2 = \mu_3 = \dots = \mu_n$

Hipótesis **alternativa**: al menos una de las medias  $\mu_x$  es distinta

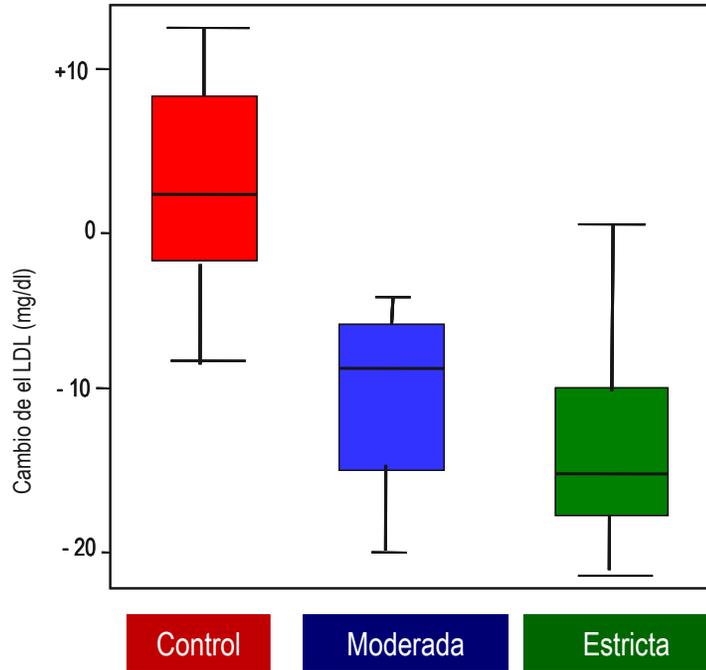
$$\frac{\text{"EFECTO" debido a la dieta}}{\text{Dispersión debida al AZAR ("error")}} = \frac{\text{Varianza entre grupos ("EFECTO")}}{\text{Varianza residual ("ERROR")}}$$

$$F = \frac{\text{Varianza entre grupos ("EFECTO")}}{\text{Varianza residual ("ERROR")}} = \frac{723,6}{35,7} = 20,2$$

¿Tienen la misma eficacia las dietas ensayadas para reducir el nivel de colesterol-LDL?



¿Cuál es la intervención dietética que más reduce los niveles de colesterol-LDL?



Cambio en los niveles de colesterol-LDL (mg/dl) a los 6 meses de tratamiento de intervención dietética en 3 grupos de voluntarios asignados a diferentes dietas

## Análisis de la varianza para un factor

Peso del RN	Suma de cuadrados	gl	Media cuadrática	F	Significación
Inter-grupos	1447,2	2	723,6	20,235	0,000
Intra-grupos	965,5	27	35,7		
Total	2412,7	29			

$$F_{\text{calculada}} = 20,235 > F_{\text{tabulado}} = 3,23; \quad p < 0,001$$

Hipótesis alternativa  $H_a$ : El reducción del LDL difiere según el tipo de dieta.

¿Cuál es la intervención dietética que más reduce los niveles de colesterol-LDL?

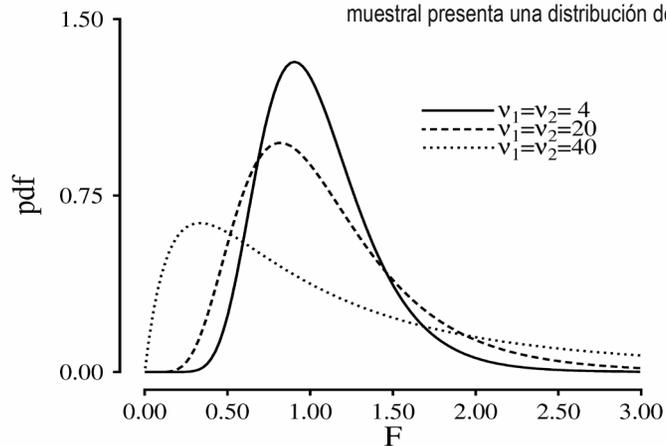
## Distribución de la F de Snedecor

**valores "F":**  $F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$

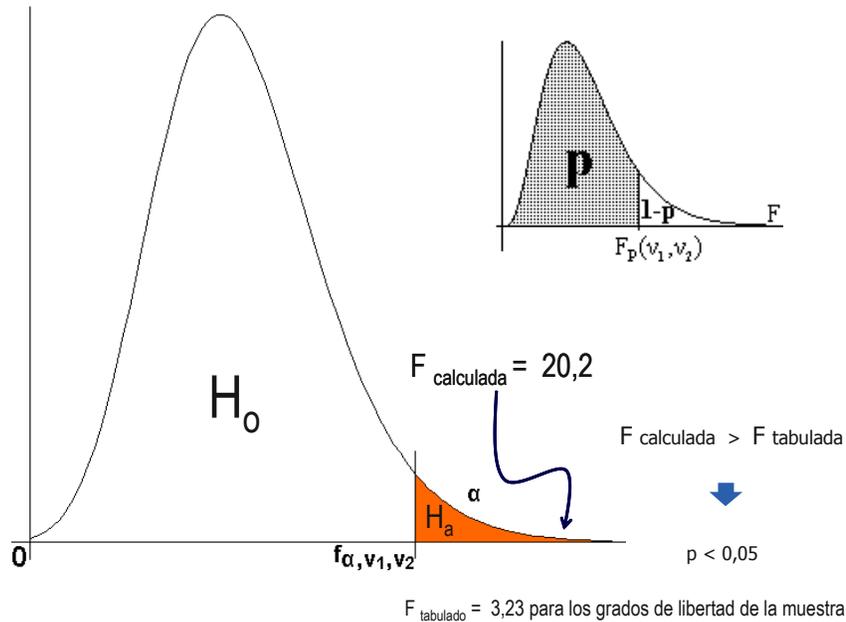
Grados de libertad  $u_1$  y  $u_2$

$$v_1 = n_1 - 1 \quad v_2 = n_2 - 1$$

El ANOVA para un factor para poblaciones con alto tamaño muestral presenta una distribución de la F de Snedecor.



# Distribución de la F de Snedecor



# Análisis de la varianza para un factor

## Análisis de la Varianza para un factor

t calculada < t tabulada  
p > 0,05

$$F = \frac{\text{Varianza entre grupos ("EFECTO")} = 318,6}{\text{Varianza residual ("ERROR")} = 61,0} = 5,2$$

t calculada > t tabulada  
p < 0,05

Hipótesis nula  $H_0$

El cambio en el nivel de cLDL es independiente de la intervención dietética

Hipótesis alternativa  $H_a$

Las medias del cLDL de los distintos grupos son diferentes entre sí, es decir NO pueden considerarse iguales

¿Tienen la misma eficacia las dietas ensayadas para reducir el nivel de colesterol-LDL?  
NO

¿Cuál es la intervención dietética que más reduce los niveles de colesterol-LDL?  
¿?

## ¿Cuál es la intervención dietética que más reduce los niveles de colesterol-LDL?

### CONTRASTES A POSTERIORI

Procedimiento de SCHEFFE	Se comparan los grupos de 2 en 2
Procedimiento de BONFERRONI	Se comparan los grupos de 2 en 2
Método de DUNNET	Se comparan los grupos de 2 en 2
Método de TUKEY	Se comparan los grupos de 2 en 2
Método de DUCAN	Se comparan los grupos de 2 en 2
Método de STUDENT-NEWMAN-KEULS	Se obtienen intervalos de medias

## Contrastes a posteriori en SPSS

ANOVA de un factor: Comparaciones múltiples post hoc

Asumiendo varianzas iguales

<input type="checkbox"/> DMS	<input type="checkbox"/> S-N-K	<input type="checkbox"/> Waller-Duncan
<input checked="" type="checkbox"/> Bonferroni	<input checked="" type="checkbox"/> Tukey	Tasa de errores tipo I/tipo II: 100
<input type="checkbox"/> Sidak	<input type="checkbox"/> Tukey-b	<input checked="" type="checkbox"/> Dunnett
<input checked="" type="checkbox"/> Scheffe	<input checked="" type="checkbox"/> Duncan	Categoría de control: Ultima
<input type="checkbox"/> R-E-G-W F	<input type="checkbox"/> GT2 de Hochberg	Contraste
<input type="checkbox"/> R-E-G-W Q	<input checked="" type="checkbox"/> Gabriel	<input checked="" type="radio"/> Bilateral <input type="radio"/> < Control <input type="radio"/> > Control

No asumiendo varianzas iguales

<input type="checkbox"/> T2 de Tamhane	<input type="checkbox"/> T3 de Dunnett	<input type="checkbox"/> Games-Howell	<input type="checkbox"/> C de Dunnett
--	--	---------------------------------------	---------------------------------------

Nivel de significación: ,05

Continuar Cancelar Ayuda

## Pruebas Post Hoc

Hipótesis nula  $H_0$ : La dieta ESTRICTA reduce el LDL de forma similar al resto de las dietas comparadas.

Hipótesis alternativa  $H_a$ : La dieta ESTRICTA reduce el LDL de forma significativa con respecto al resto de las dietas comparadas.

DIETAS	Control	Moderada	Estricta
Control		$H_0 =$ $H_a \neq$	$H_0 =$ $H_a \neq$
Moderada			$H_0 =$ $H_a \neq$
Estricta			

## Procedimiento de Bonferroni

Dieta (I)	Dieta (J)	Diferencias de medias (I – J)	Error estándar	Significación
Control	Moderada	12,60	2,674	0,000
	Estricta	16,20	2,674	0,000
Moderada	Control	- 12,60	2,674	0,000
	Estricta	3,60	2,674	0,568
Estricta	Control	- 16,20	2,674	0,000
	Moderada	- 3,60	2,674	0,568

Hipótesis nula  $H_0$ : La dieta ESTRICTA reduce el LDL de forma similar al resto de las dietas comparadas.

## Procedimiento de Scheffe

Dieta (I)	Dieta (J)	Diferencias de medias (I – J)	Error estándar	Significación
Control	Moderada	12,60	2,674	0,000
	Estricta	16,20	2,674	0,000
Moderada	Control	- 12,60	2,674	0,000
	Estricta	3,60	2,674	0,416
Estricta	Control	- 16,20	2,674	0,000
	Moderada	- 3,60	2,674	0,416

Hipótesis nula  $H_0$ : La dieta ESTRICTA reduce el LDL de forma similar al resto de las dietas comparadas.

## Procedimiento de HSD de Tukey

Dieta (I)	Dieta (J)	Diferencias de medias (I – J)	Error estándar	Significación
Control	Moderada	12,60	2,674	0,000
	Estricta	16,20	2,674	0,000
Moderada	Control	- 12,60	2,674	0,000
	Estricta	3,60	2,674	0,383
Estricta	Control	- 16,20	2,674	0,000
	Moderada	- 3,60	2,674	0,383

Hipótesis nula  $H_0$ : La dieta ESTRICTA reduce el LDL de forma similar al resto de las dietas comparadas.

## Procedimiento de Gabriel

Dieta (I)	Dieta (J)	Diferencias de medias (I – J)	Error estándar	Significación
Control	Moderada	12,60	2,674	0,000
	Estricta	16,20	2,674	0,000
Moderada	Control	- 12,60	2,674	0,000
	Estricta	3,60	2,674	0,459
Estricta	Control	- 16,20	2,674	0,000
	Moderada	- 3,60	2,674	0,459

Hipótesis nula  $H_0$ : La dieta ESTRICTA reduce el LDL de forma similar al resto de las dietas comparadas.

## Procedimiento de Sidak

Dieta (I)	Dieta (J)	Diferencias de medias (I – J)	Error estándar	Significación
Control	Moderada	12,60	2,674	0,000
	Estricta	16,20	2,674	0,000
Moderada	Control	- 12,60	2,674	0,000
	Estricta	3,60	2,674	0,467
Estricta	Control	- 16,20	2,674	0,000
	Moderada	- 3,60	2,674	0,467

Hipótesis nula  $H_0$ : La dieta ESTRICTA reduce el LDL de forma similar al resto de las dietas comparadas.

## Procedimiento de la t de Dunnett (bilateral)

Dieta (I)	Dieta (J)	Diferencias de medias (I – J)	Error estándar	Significación
Control	Estricta	16,20	2,674	0,000
Moderada	Estricta	3,60	2,674	0,315

La prueba de Dunnett trata como un grupo como control y lo compara con todos los demás grupos.

Hipótesis nula  $H_0$ : La dieta ESTRICTA reduce el LDL de forma similar al resto de las dietas comparadas.

## Método de Duncan

Tipo de dieta			Subconjunto para alfa $\alpha = 0,05$	
Test de Duncan	DIETA	N	1	2
	Estricta	10	-13,700	
	Moderada	10	-10,100	
	Control	10		2,500
	Significación			0,189

Comparación de la dieta estricta con la moderada ( $p > 0,1$ ).

Comparación de la dieta control con las demás ( $1 - p = 1,000$ ;  $p = 0,000$ ).

Hipótesis nula  $H_0$ : La dieta CONTROL reduce el LDL de forma similar al resto de las dietas comparadas. No existe diferencia entre la moderada y la estricta).

## Mito I: “La sonrisa de la esfinge”

ANOVA es un test paramétrico, es decir para usarlo se debe cumplir la normalidad de la variable cuantitativa. ¿Se cumple?

He observado que en los tres grupos de estudio sólo tienen 10 sujetos. Ante tan bajo volumen muestral...

¿Cómo justificas esa supuesta normalidad?

¿Qué test has empleado para comprobar esa normalidad de la reducción del LDL?

¿Estás seguro que se cumplen los criterios para poder usar ANOVA?

## Realidad I: “Respuesta de Edipo”

		Estadísticos descriptivos del LDL en cada dieta							
		Tipo de dieta	Media	DE	Mediana	$\gamma_1$	EE $\gamma_1$	$\gamma_2$	EE $\gamma_2$
Reducción del LDL	Control		2,500	6,53	2,000	0,127	0,687	-0,732	1,334
	Moderada		-10,100	5,19	-8,500	- 0,880	0,687	-0,307	1,334
	Estricta		- 13,700	6,12	-15,000	1,172	0,687	1,909	1,334

Distribución normal

$$\rightarrow [\text{Asimetría}] < 2 \text{EE}_{\text{asimetría}} \quad \gamma_1 < 2 \text{EE} \gamma_1$$

$$\rightarrow [\text{Curtosis}] < 2 \text{EE}_{\text{curtois}} \quad \gamma_2 < 2 \text{EE} \gamma_2$$

## Realidad I: “Respuesta de Edipo”

		Test de Kolmogorov-Smirnov		
	Tipo de dieta	Estadístico Z	g.l	Significación
Reducción del LDL	Control	0,488	10	0,971
	Moderada	0,657	10	0,782
	Estricta	0,582	10	0,887

Ho:  $p > 0,05$  NORMALIDAD



Test paramétricos

Ha:  $p < 0,05$  ANORMALIDAD



Test NO paramétricos

Los test de normalidad ensayados dan un valor de p NO significativo ( $p > 0,05$ ) en los 3 grupos (tipos de dieta). Por tanto no hay evidencias suficientes para rechazar la hipótesis nula de normalidad de la variable reducción del peso en ambas muestras: podemos usar test PARAMÉTRICOS.

## Test ANOVA: Observaciones

- Los resultados del contraste estadístico F en la prueba ANOVA son sustancialmente válidos aunque los datos no sean normales, en ese sentido se dice que es una técnica "ROBUSTA" frente a desviaciones de la normalidad.
- El efecto de desigualdad de las varianzas en los grupos sobre el contraste F y los contrastes de medias dependen de que el número de observaciones en cada grupo sea igual o muy distinto. Si todos los grupos tienen el mismo número de observaciones, el contraste F es igualmente exacto aunque las varianzas sean distintas. Es decir, podemos despreocuparnos de las varianzas a efectos de contrastes de medias, siempre que haya aproximadamente el mismo número de observaciones por grupo, en caso contrario, diferencias entre las varianzas pueden ser graves.

## Pregunta de investigación

¿El peso al nacer de los RN difiere según la localidad de origen?

Variable PREDICTORA  
CUALITATIVA policotómica

- Ciudad de origen
- 1. Ceuta
- 2. Algeciras
- 3. Tarifa
- 4. La Línea



Variable RESULTADO  
CUANTITATIVA continua

Peso del RN

¿Cuál es la prueba estadística más adecuada?

## Análisis de la varianza para un factor

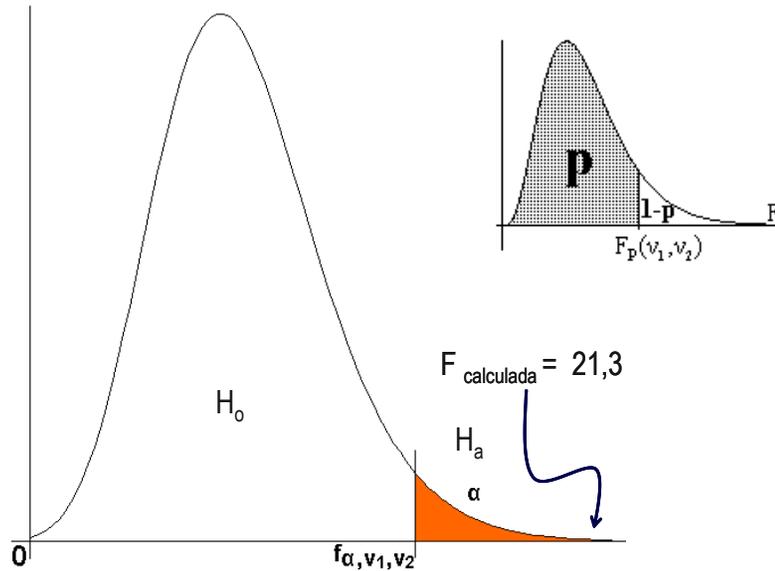
Hipótesis nula  $H_0$ : El peso al nacer de los RN NO difiere según la localidad de origen.

Hipótesis alternativa  $H_a$ : El peso al nacer de los RN difiere según la localidad de origen.

Estadísticos	Peso al nacer del RN			
	Ceuta	Algeciras	Tarifa	La Línea
Localidad de origen	Ceuta	Algeciras	Tarifa	La Línea
Media	3366,2	3217,2	3188,4	3169,0
Desviación estándar (S)	393,0	444,1	471,5	361,4
n	1146	214	80	186

$$F = \frac{\text{Varianza entre grupos ("EFECTO")}}{\text{Varianza residual ("ERROR")}} = 21,3$$

# Distribución de la F de Snedecor



$F_{\text{tabulado}} = 3,23$  para los grados de libertad de la muestra

## Análisis de la varianza para un factor

Peso del RN	Suma de cuadrados	gl	Media cuadrática	F	Significación
Inter-grupos	10290751	3	3430250,3	21,351	0,000
Intra-grupos	260596358	1622	160663,599		
Total	270887109	1625			

$$F_{\text{calculada}} = 21,3 > F_{\text{tabulado}} = 3,23; \quad p < 0,001$$

Hipótesis alternativa  $H_a$ : El peso al nacer de los RN difiere según la localidad de origen.

¿Cuál es la ciudad donde los niños nacen con mayor peso?

## Pregunta de investigación

¿Cuál es la ciudad donde los niños nacen con mayor peso?

Variable PREDICTORA  
CUALITATIVA policotómica

- Ciudad de origen
1. Ceuta
  2. Algeciras
  3. Tarifa
  4. La Línea



Variable RESULTADO  
CUANTITATIVA continua

Peso del RN

¿Cuál es la prueba estadística más adecuada?

## Pruebas Post Hoc paramétrico

### CONTRASTES A POSTERIORI

Procedimiento de SCHEFFE	Se comparan los grupos de 2 en 2
Procedimiento de BONFERRONI	Se comparan los grupos de 2 en 2
Método de DUNNET	Se comparan los grupos de 2 en 2
Método de TUKEY	Se comparan los grupos de 2 en 2
Método de DUCAN	Se comparan los grupos de 2 en 2
Método de STUDENT-NEWMAN-KEULS	Se obtienen intervalos de medias

## Pruebas Post Hoc

Hipótesis nula  $H_0$ : Los RN que nacen en Ceuta tiene un peso medio similar al resto de los RN de las otras ciudades.

Hipótesis alternativa  $H_a$ : Los RN que nacen en Ceuta tiene mayor peso medio que el resto de las ciudades.

Ciudades	Ceuta	Algeciras	Tarifa	La Línea
Ceuta		$H_0 =$ $H_a \neq$	$H_0 =$ $H_a \neq$	$H_0 =$ $H_a \neq$
Algeciras			$H_0 =$ $H_a \neq$	$H_0 =$ $H_a \neq$
Tarifa				$H_0 =$ $H_a \neq$
La Línea				

## Procedimiento de Scheffe

Ciudad (I)	Ciudad (J)	Diferencias de medias (I – J)	Error estándar	Significación
Ceuta	Algeciras	148,92	29,85	0,000
	Tarifa	177,79	46,35	0,002
	La Línea	197,13	31,69	0,000
Algeciras	Ceuta	- 148,92	29,85	0,000
	Tarifa	28,87	52,53	0,960
	La Línea	48,21	40,18	0,696
Tarifa	Ceuta	- 177,79	46,35	0,002
	Algeciras	- 28,87	52,53	0,960
	La línea	19,34	53,59	0,988
La Línea	Ceuta	- 197,13	31,69	0,000
	Algeciras	- 48,21	40,18	0,696
	Tarifa	- 19,92	53,59	0,988

Hipótesis alternativa  $H_a$ : Los RN que nacen en Ceuta tiene mayor peso medio que el resto de las ciudades.

## Procedimiento de Bonferroni

Ciudad (I)	Ciudad (J)	Diferencias de medias (I – J)	Error estándar	Significación
Ceuta	Algeciras	148,92	29,85	0,000
	Tarifa	177,79	46,35	0,001
	La Línea	197,13	31,69	0,000
Algeciras	Ceuta	- 148,92	29,85	0,000
	Tarifa	28,87	52,53	1,000
	La Línea	48,21	40,18	1,000
Tarifa	Ceuta	- 177,79	46,35	0,001
	Algeciras	- 28,87	52,53	1,000
	La línea	19,34	53,59	1,000
La Línea	Ceuta	- 197,13	31,69	0,000
		- 48,21	40,18	1,000
	Tarifa	- 19,92	53,59	1,000

Hipótesis alternativa  $H_a$ : Los RN que nacen en Ceuta tiene mayor peso medio que el resto de las ciudades.

## Procedimiento de Duncan

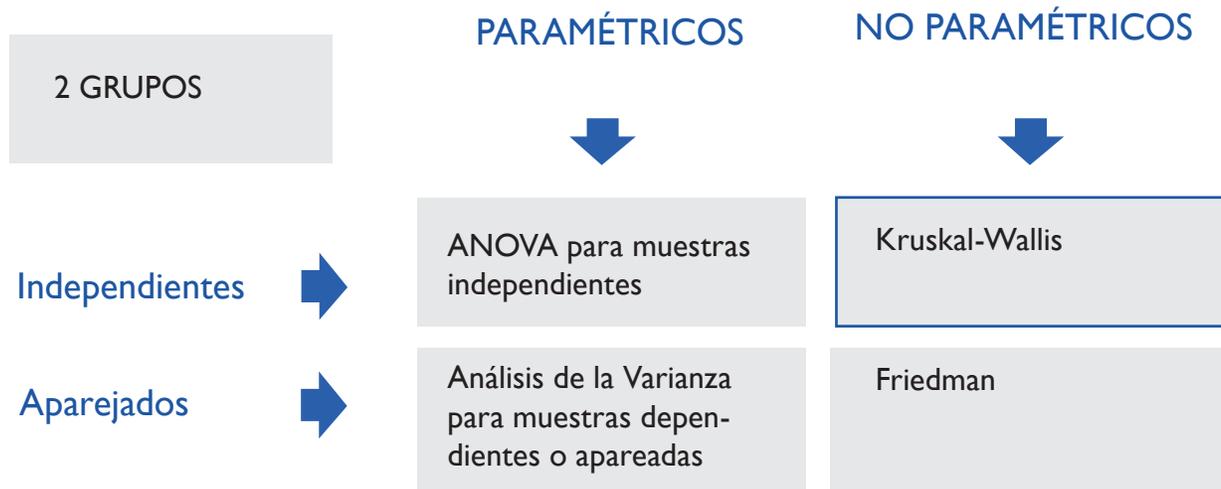
Procedencia		Subconjunto para alfa $\alpha = 0,05$		
Test de Duncan	Ciudad	N	1	2 (Ceuta)
	La Línea	186	3169,03	
	Tarifa	80	3188,38	
	Algeciras	214	3217,24	
	Ceuta	1146	3366,16	0,960
	Significación			0,298

Comparación de las 3 ciudades  
( $p > 0,1$ ).

Comparación de Ceuta con las demás  
ciudades ( $1 - p = 1,000$ ;  $p = 0,000$ ).

Hipótesis alternativa  $H_a$ : Los RN que nacen en Ceuta tiene mayor peso medio que el resto de las ciudades.

## Cualitativa policotómica versus cuantitativa



Variable predictora Cualitativa versus variable resultado Cuantitativa.

## La prueba de Kruskal-Wallis (comparar más de dos grupos)

- Para la comparación de una misma variable cuantitativa en más de tres grupos. Es la prueba NO paramétrica más potente para comparar >2 variables continuas independientes. Es el “similar” al test de ANOVA para un factor.
- La prueba de Kruskal-Wallis, es una alternativa a la prueba F del ANOVA (análisis de varianza) para diseños de clasificación simple donde NO se cumple la normalidad de la variable cuantitativa.
- En este caso se comparan varios grupos de la variable predictora (cualitativa POLICOTÓMICA) pero usando la MEDIANA de cada uno de ellos, en lugar de las medias, donde, n es el total de datos.

$$H = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(n+1)$$

$H_0$ : La mediana de las k poblaciones consideradas son iguales y

$H_a$ : Al menos una de las poblaciones tiene mediana distinta a las otras

## Pregunta de investigación

¿Tienen la misma eficacia las dietas ensayadas en grupos de 5 individuos para reducir el nivel de colesterol-LDL?

Variable PREDICTORA  
CUALITATIVA policotómica

- Tipos de dieta
- Dieta control
- Dieta moderada
- Dieta estricta



Variable RESULTADO  
CUANTITATIVA continua

LDL- colesterol

¿Cuál es la prueba estadística más adecuada?

## ¿Tienen la misma eficacia las dietas ensayadas para reducir el nivel de colesterol-LDL?

Variaciones en los niveles de colesterol-LDL (mg/dl) a los 6 meses de seguimiento en 3 grupos de voluntarios asignados a diferentes medidas de intervención dietética.

Grupo Control	Dieta moderada	Dieta estricta
+ 13	- 4	0
+ 10	- 8	- 12
+ 2	- 9	- 15
- 2	- 10	- 21
- 8	- 20	
N = 5	N = 5	N = 4
Media = + 3	Media = - 10,2	Media = - 12
N = 14		Media total = - 6

## Resultados: Prueba de Kruskal-Wallis

Rangos		Cambio del LDLc
Tipo de DIETAS	N	Rango promedio
CONTROL	5	12,30
MODERADA	5	6,70
ESTRICTA	5	5,00
Total		15

Estadístico de contraste	
N	15
Chi-cuadrado	7,308
gl (K – 1)	2
Significación	0,026

# Test de Kruskal-Wallis

## Test de Kruskal-Wallis

$t$  calculada <  $t$  tabulada  
 $p > 0,05$

$X^2 H = 7,3$   $p = 0,026$

$t$  calculada >  $t$  tabulada  
 $p < 0,05$

Hipótesis nula  $H_0$

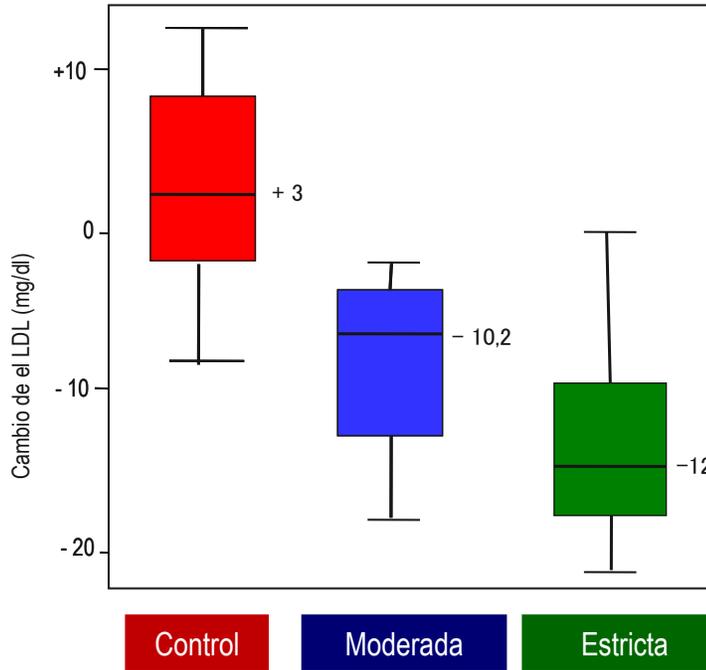
El cambio en el nivel de cLDL es independiente de la intervención dietética

Hipótesis alternativa  $H_a$

Las medias del cLDL de los distintos grupos son diferentes entre sí, es decir NO pueden considerarse iguales

¿Tienen la misma eficacia las dietas ensayadas para reducir el nivel de colesterol-LDL?  
NO

¿Cuál es la intervención dietética que más reduce los niveles de colesterol-LDL?  
¿?



Cambio en los niveles de colesterol-LDL (mg/dl) a los 6 meses de tratamiento de intervención dietética en 3 grupos de 5 voluntarios asignados a diferentes dietas

## Pruebas Post Hoc no paramétricas

CONTRASTES A POSTERIORI

Procedimiento de BONFERRONI

Procedimiento de ŠIDÁK

Método de HOLM

Método de HOLLAND-COPENHAVER

Método de HOCHBERG

Método de HOMMEL

Método de ROM

## ***ORIGINAL***

# **Precisión diagnóstica de la calprotectina fecal para predecir una colonoscopia patológica**

María del Valle García Sánchez<sup>a</sup>, Raúl González<sup>b</sup>, Eva Iglesias Flores<sup>a</sup>, Federico Gómez Camacho<sup>a</sup>, Luis Casais Juanena<sup>a</sup>, Antonio Cerezo Ruiz<sup>a</sup>, Manuel Montero Pérez-Barquero<sup>c</sup>, Jordi Muntané<sup>b</sup> y Juan Francisco de Dios Vega<sup>a</sup>

<sup>a</sup>Unidad Clínica Aparato Digestivo. Hospital Universitario Reina Sofía. Córdoba.

<sup>b</sup>Unidad de Investigación. Hospital Universitario Reina Sofía. Córdoba.

<sup>c</sup>Medicina Interna. Hospital Universitario Reina Sofía. Córdoba. España.

**Med Clin (Barc) 2006;127:41-6**

**FUNDAMENTO Y OBJETIVO:** El dolor abdominal y la diarrea son síntomas comunes en la población general que no permiten discriminar enfermedad intestinal orgánica. La colonoscopia es la prueba de referencia, pero es cara, invasiva, no puede repetirse frecuentemente y supone una gran carga asistencial. La calprotectina fecal (CPF) es un marcador fácil y rápido de medir, no invasivo y barato, que podría detectar la enfermedad orgánica en la zona del colon. El objetivo de nuestro estudio es valorar la utilidad de este marcador para predecir una colonoscopia patológica y correlacionar sus valores con el grado de actividad en la enfermedad inflamatoria intestinal (EII).

**PACIENTES Y MÉTODO:** Se determinó la calprotectina en muestras fecales (mediante enzimoanálisis) en 190 personas sometidas a una colonoscopia. Los diagnósticos endoscópicos fueron: 117 normales, 28 pólipos colónicos, 20 cáncer colorrectal (CCR) y 25 EII. El análisis de los datos se realizó con el programa SPSS 11.0.

**RESULTADOS:** La media (DE) de CPF estaba aumentada en CCR (726,6 [533] mg/kg) y EII (2.171,1 [2.133,6] mg/kg). No hubo diferencias en pacientes con pólipos (158,3 [15,8] mg/kg). En pacientes con EII, los valores se correlacionaron con el grado de actividad de la enfermedad. El valor de corte fue 217 mg/kg, y se obtuvo una sensibilidad de la CPF del 85% y un valor predictivo negativo del 93%. En el grupo control, la única variable que se relacionó con un aumento de CPF fue la medicación previa con ácido acetilsalicílico (AAS) y/o antiinflamatorios no esteroideos (AINE).

**CONCLUSIONES:** La CPF es una prueba no invasiva, barata y sensible para detectar enfermedad orgánica en la zona del colon, que puede ser útil para seleccionar a pacientes que deben someterse a una colonoscopia. Se correlaciona con el grado de actividad en la EII. La toma de AAS y/o AINE podría aumentar la tasa de falsos positivos.

*Palabras clave:* Calprotectina fecal. Colonoscopia. Inflamación intestinal.

## Pregunta de investigación

¿Qué relación existe entre los niveles de calprotectina fecal y las diferentes enfermedades orgánicas intestinales?

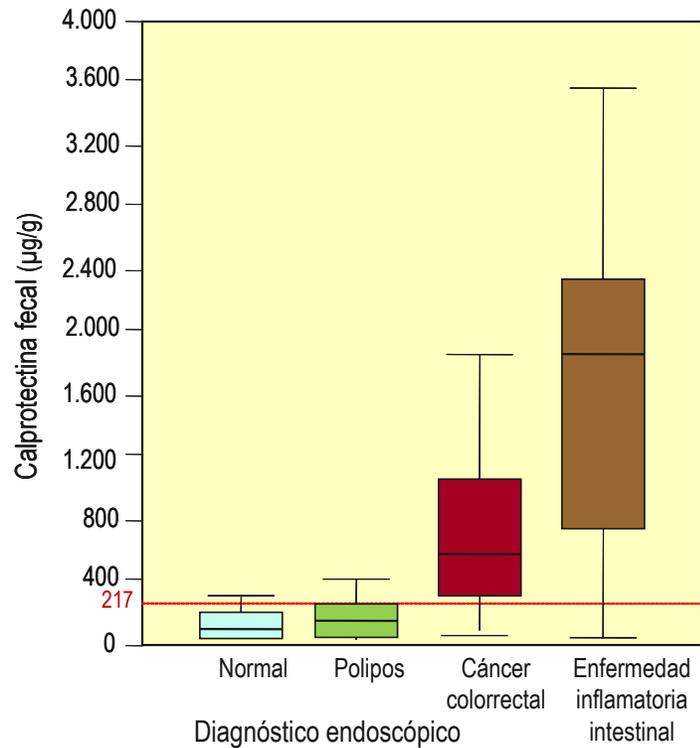
Variable PREDICTORA  
CUALITATIVA policotómica

- Controles
- Pólipos
- Cáncer colorrectal
- Enfermedad inflamatoria intestinal



Variable RESULTADO  
CUANTITATIVA continua

Calprotectina fecal



¿Cuál es la prueba estadística más adecuada?

## ¿Qué relación existe entre los niveles de calprotectina fecal y las diferentes enfermedades orgánicas intestinales?

Concentraciones de calprotectina fecal ( $\mu\text{g/g}$ )			
Grupos	N	Media	DE
Controles	117	113,7	113,1
Pólipos	28	158,3	155,8
Cáncer colorrectal	20	726,6	533,2
Enfermedad inflamatoria intestinal	25	2171,1	2133,6

Test de KRUSKAL-WALLIS



$p < 0,001$

## Análisis estadístico (fragmento del texto)

El análisis de los datos se llevó a cabo mediante el programa estadístico SPSS 11.0. El valor de la CPF se presenta como media y con un intervalo de confianza (IC) del 95%. Se utilizó el test de Kruskal-Wallis para evaluar las diferencias entre los valores de CPF entre los diferentes grupos y el coeficiente de Spearman para valorar la correlación entre este marcador y el índice de actividad de la enfermedad en el grupo de pacientes con enfermedad inflamatoria intestinal (EII).

## Resultados (fragmento del texto)

“En la figura 1 se muestra los valores de CPF en los diferentes grupos de pacientes. Se encontró un incremento del valor medio (DE) en pacientes con enfermedad inflamatoria intestinal (EII) EII (2.171 [2.134] mg/kg) y CCR (727 [533] mg/kg) con diferencias estadísticamente significativas con respecto al grupo 1 (114 [113] µg/g,  $p < 0,001$ )”.

## Discusión (fragmento del texto)

“El presente trabajo muestra que la concentración de CPF está significativamente elevada en pacientes con procesos orgánicos en colon, fundamentalmente en el cáncer colorrectal y enfermedad inflamatoria intestinal (EII). Además, los valores más altos se alcanzan en pacientes con EII, y sus valores se correlacionan con el grado de actividad de la enfermedad”.

## Mito I: “La sonrisa de la esfinge”

Me parece acertada la elección del test de KRUSKAL-WALLIS como contraste de hipótesis, sin embargo...

¿Qué test has empleado para comprobar esa supuesta anormalidad de la calprotectina fecal?

Si la calprotectina no sigue la distribución normal...

¿Por qué utilizas la media y la desviación estándar como estadísticos descriptivos?

## Mito I: “La sonrisa de la esfinge”

El test de KRUSKAL-WALLIS muestra un nivel crítico que es menor de 0,05, y sólo podemos rechazar la hipótesis de igualdad de promedios, concluyendo que los grupos comparados (control, pólipos, cáncer colorrectal y enfermedad inflamatoria intestinal), difieren en los niveles de calprotectina fecal. Sin embargo el test no te permite decir que en la enfermedad inflamatoria intestinal los valores son mayores.

¿Por qué dices que “los valores más altos se alcanzan en pacientes con enfermedad inflamatoria intestinal”, cuando eso no te lo permite el test estadístico empleado?

## Mito I: “La sonrisa de la esfinge”

Aunque existen procedimientos POCT HOC para efectuar comparaciones múltiples tras obtener un estadístico H significativo, para analizar con SPSS que grupo tiene un mayor valor de CPF puedes utilizar la U de Mann-Whitney para dos muestras independientes, pero acompañada de la corrección de Bonferroni para controlar la probabilidad de cometer un error alfa.

Puesto que con 4 grupos de la variable predictora; control (CT), polipos (PL), cáncer colorrectal (CCR) y enfermedad inflamatoria intestinal (EII), necesitas hacer 6 comparaciones 2 a 2 (CT-PL, CT-CCR, CT-EII, PL-CCR, PL-EII, CCR-EII), la aplicación de la corrección de Bonferroni te llevará a basar vuestras decisiones en un nivel de significación de  $0,05/6 = 0,0083$ . Es decir, consideraremos que dos grupos comparados difieren significativamente cuando el nivel crítico obtenido ( $p$ ) sea inferior a 0,0083.

## Test de Kruskal-Wallis

VENTAJAS	INCONVENIENTES
Permite datos ordinales y cuantitativos discretos	Menor potencia estadística
No es necesaria la normalidad (test no paramétrico)	Menor sensibilidad para detectar diferencias entre grupos
El tamaño muestral puede ser pequeño	No permite construir intervalos de confianza

## Prueba de la Mediana

- Su objetivo es comparar las medianas de dos muestras y determinar si pertenecen a la misma población o no.
- Se calcula la **mediana** de todos los datos conjuntamente. Después, se divide cada muestra en dos subgrupos: uno para aquellos datos que se sitúen por encima de la mediana y otro para los que se sitúen por debajo.
- La prueba de **chi-cuadrado** determinará si las frecuencias observadas en cada grupo difieren de las esperadas con respecto a una distribución de frecuencias que combine ambas muestras.
- Esta prueba está especialmente indicada cuando los datos sean extremos o estén sesgados.

## Pregunta de investigación

¿Tienen la misma eficacia las dietas ensayadas para reducir el nivel de colesterol-LDL?

Variable PREDICTORA  
CUALITATIVA policotómica

- Tipos de dieta
- Dieta control
- Dieta moderada
- Dieta estricta



Variable RESULTADO  
CUANTITATIVA continua

LDL- colesterol

¿Cuál es la prueba estadística más adecuada?

Grupo Control (n=10)	Dieta moderada (n=10)	Dieta estricta (n=10)
+ 13	4	0
+ 10	8	12
+ 2	9	15
2	10	21
- 8	20	-10
+ 5	15	-18
3	6	-16
+ 8	7	-10
+ 2	16	-20
-2	6	-15
Media = + 2,5	Media = - 10,1	Media = - 13,7

## Prueba de la Mediana

	Tipo de dietas		
	Control	Moderada	Estricta
> Mediana	9	4	1
≤ Mediana	1	6	9

Hay 3 casillas con frecuencia esperada menor de 5.  
La frecuencia de casilla esperada mínima es de 4,7

Estadístico de contraste	LDL
N	30
Mediana	- 8,000
Chi-cuadrado	13,125
gl (K - 1)	2
Significación	0,001

## Test de Jonckheere-Terpstra

- Cuando no se cumplen las condiciones para el ANOVA, hay tres tests disponibles: H de Kruskal-Wallis, mediana y Jonckheere-Terpstra.
- El test H de Kruskal-Wallis, una extensión del test U de Mann-Whitney es el análogo no-paramétrico del ANOVA.
- El test de la mediana, que es una prueba más general pero no tan potente, detecta diferencias en la situación y forma de las distribuciones. Ambas pruebas asumen que no hay un ordenamiento a priori de las k poblaciones de las que se han extraído las muestras.
- Cuando hay una ordenación natural a priori (ascendente o descendente) de las k poblaciones, es preferible usar el test de Jonckheere-Terpstra.

74 pacientes ingresados en una UCI y queremos comparar las estancias generadas en dicha unidad entre tres grupos de pacientes clasificados en función de la edad (jóvenes, maduros y ancianos)

Variable PREDICTORA I  
CUALITATIVA ordinal

Grupos etarios  
- Jóvenes  
- Maduros  
- Ancianos



Variable RESULTADO  
CUANTITATIVA

Estancia  
- Días

¿Cuál es el test estadístico más acertado?

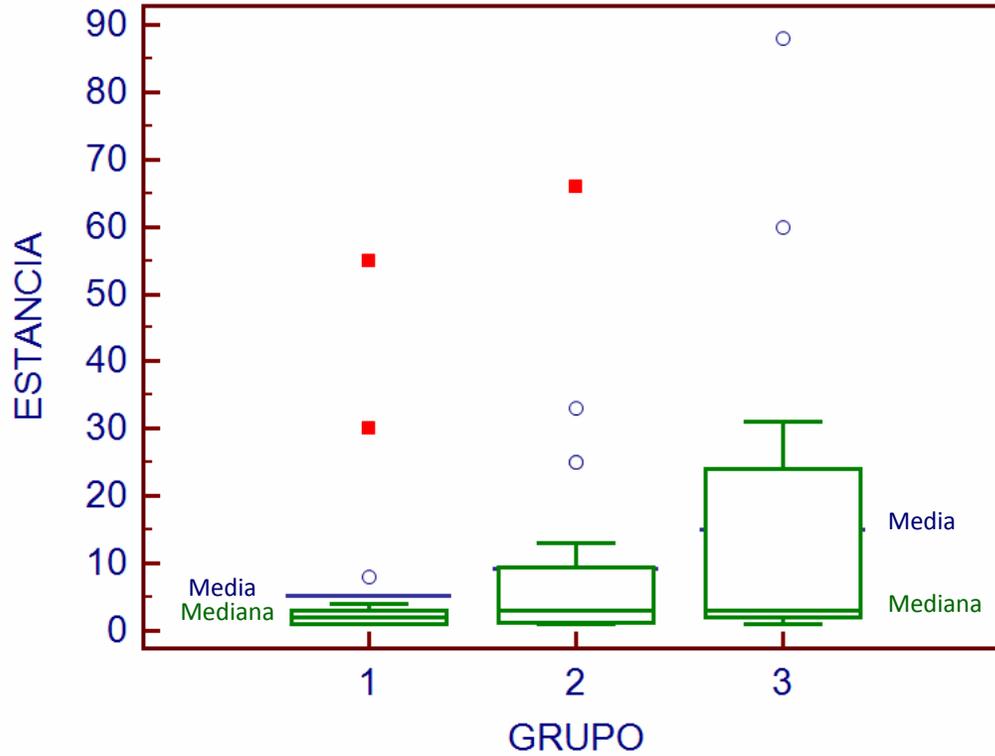
## Estadísticos descriptivos

Variable resultado	ESTANCIA	N	Media	DE	EEM	Mediana
Variable predictora Grupos etarios	JÓVENES	28	5,18	11,17	2,11	2,00
	MADUROS	23	9,22	15,29	3,19	3,00
	ANCIANOS	23	21,51	21,52	4,49	3,00

Variable resultado	ESTANCIA	N	$\gamma_1$	EE $\gamma_1$	$\gamma_2$	EE $\gamma_2$
Variable predictora Grupos etarios	JÓVENES	28	3,944	0,441	16,048	0,858
	MADUROS	23	2,760	0,481	8,430	0,935
	ANCIANOS	23	2,290	0,481	5,682	0,935

Distribución normal  $\rightarrow \gamma_1 < 2 EE \gamma_1 \quad \gamma_2 < 2 EE \gamma_2$

La variable resultado ESTANCIA, como casi todas las que se refieren al tiempo, es una variable conocida por su distribución alejada de lo normal. Si el tamaño de los grupos que se estudian es pequeño –menor de 30–, no sería adecuada la utilización de métodos paramétricos (ANOVA) y deberíamos recurrir a los clásicos no paramétricos (KRUSKAL-WALLIS).



## Test de normalidad

Variable resultado ESTANCIA		Test de Kolmogorov-Smirnov		
		N	Estadístico Z	Significación
Variable predictora Grupos etarios	Jóvenes	28	2,301	0,000
	Maduros	23	1,668	0,008
	Ancianos	23	1,236	0,097
	Total	74	2,829	0,000

$H_0$ :  $p > 0,05$  NORMALIDAD Test paramétricos

$H_a$ :  $p < 0,05$  Test NO paramétricos

El test de normalidad ensayado da un valor de  $p$  significativo ( $p < 0,05$ ). Por tanto hay evidencias suficientes para rechazar la hipótesis nula de normalidad de la variable ESTANCIA en UCI en los grupos. No se puede asumir la normalidad.

# Prueba de la homogeneidad de las varianzas

## Test de Levene

Variable resultado ESTANCIA en UCI (días)			
Estadístico de Levene	gl1	gl2	Significación
3,707	2	71	0,022

El test de Levene (1960) se usa para contrastar si los 3 grupos comparados tienen la misma varianza, es decir, la homogeneidad de varianzas. Para poder utilizar el análisis de la varianza es necesario que las varianzas son iguales para todos los grupos. De ahí la importancia de verificar con el test de Levene esa hipótesis.

$H_0$ : Las varianzas de ambos grupos son iguales ➔ ANOVA

$H_a$ : Existen diferencias entre las varianzas ➔ KRUSKAL-WALLIS

$p > 0,05$  (aceptamos la  $H_0$ ) ➔ ANOVA

$p < 0,05$  (rechazamos la hipótesis nula) ➔ KRUSKAL-WALLIS

## Reflexiones: existen varios incumplimientos para el ANOVA

1. Ausencia de normalidad.
2. Ausencia de homocedasticidad: el test de Levene para valorar la homogeneidad de las variancias muestra que existen diferencias entre ellas ( $p = 0,027$ ).
3. Y existencia de valores anómalos (outliers) que impiden la utilización de métodos paramétricos (ANOVA clásico).

Tenemos que recurrir a los clásicos no paramétricos de Kruskal-Wallis o, mejor, a métodos robustos como la generalización robusta del test Welch, que utiliza medias  $\alpha$ -recortadas muestrales y es capaz de detectar diferencias significativas entre los grupos.

## Resultados: Prueba H de Kruskal-Wallis

Cálculo de Rangos		
GRUPO ETARIO	N	Rango promedio
JÓVENES	28	31,00
MADUROS	23	38,37
ANCIANOS	23	44,54
Total		74

Estadístico de contraste	ESTANCIA
Chi-cuadrado	5,234
gl (K – 1)	2
Significación	0,073

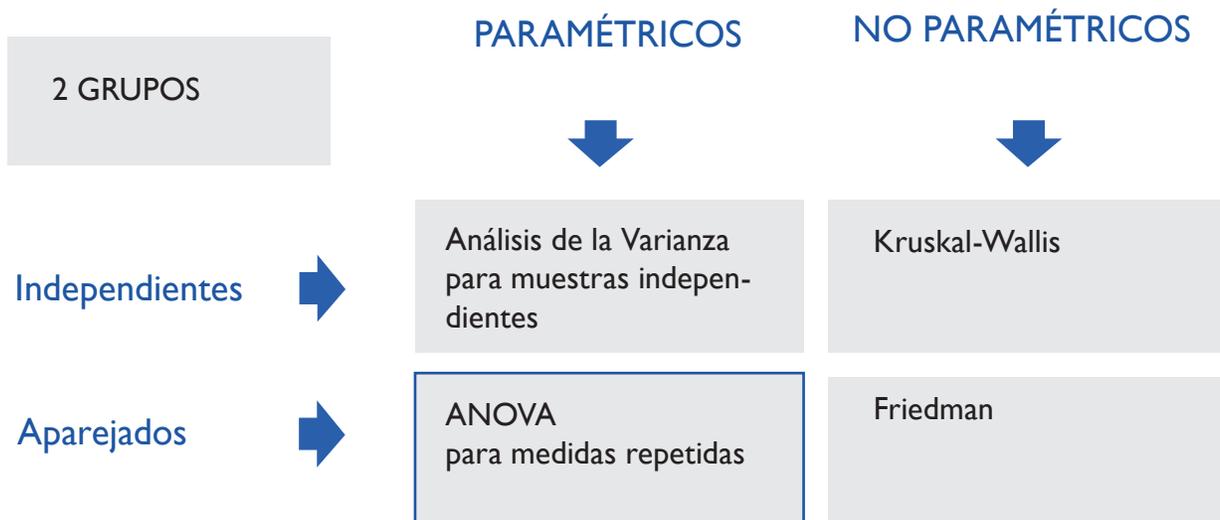
Hipótesis nula  $H_0$ : La estancia en UCI de los pacientes estudiados NO difiere según el grupo etario.

## Test de Jonckheere-Terpstra (J-T)

Estadístico de contraste	ESTANCIA
Número de niveles	3
N	74
Estadístico de J-T observado	1136,500
Media del estadístico J-T	908,500
Desviación típica del estadístico J-T	99,118
Estadístico de J-T tipificado	2,305
Significación (bilateral)	0,021

Hipótesis alternativa  $H_a$ : La estancia en UCI de los pacientes estudiados difiere según el grupo etario.

## Cualitativa policotómica versus cuantitativa



Variable predictora Cualitativa versus variable resultado Cuantitativa.

## Análisis de la Varianza (ANOVA) para muestras dependientes o apareadas (medidas repetidas)

- Análisis de la variancia (ANOVA) se utiliza para la obtención del coeficiente de correlación intraclase (CCI).
- En los estudios que evalúan la concordancia, cada sujeto es examinado por más de un observador, siendo adecuado realizar un análisis de la variancia (ANOVA) para medidas repetidas (generalización de la prueba de la t de Student para datos apareados).
- Este tipo de diseño permite aislar la variabilidad entre pacientes y concentrarnos en la variabilidad debida al factor de interés; en el análisis de la variancia simple, la variabilidad entre pacientes está incluida en la variabilidad residual.

Experimento diseñado para estudiar el efecto del paso del tiempo sobre la calidad del recuerdo y realización práctica de estudios con aplicación de la estadística de los 30 alumnos que han realizado el curso de BIOESTADÍSTICA APLICADA A LAS CIENCIAS DE LA SALUD. Se evalúa el grado de conocimientos con el mismo examen de 20 preguntas tipo test en 4 momentos diferentes; inmediatamente después de finalizar el curso, al año, a los 2 años, y a los 3 años del curso. Se valora el número de respuestas correctas.

TIEMPO	N	Media	DE	$\gamma_1$	EE $\gamma_1$	$\gamma_2$	EE $\gamma_1$
INICIO	30	15,53	2,776	0,014	0,427	- 0,946	0,833
PRIMER AÑO	30	12,00	2,626	- 0,159	0,427	- 0,515	0,833
SEGUNDO AÑO	30	9,37	2,282	0,334	0,427	0,412	0,833
TERCER AÑO	30	8,03	2,539	0,366	0,427	- 0,647	0,833

TIEMPO	Test de Kolmogorov-Smirnov		
	N	Estadístico Z	Significación
INICIO	30	0,618	0,839
PRIMER AÑO	30	0,548	0,925
SEGUNDO AÑO	30	0,679	0,746
TERCER AÑO	30	0,648	0,795

## Comparación de $>$ de 2 grupos apareados

Test de Kolmogorov-Smirnov

↓  $p > 0,05$ ; Normalidad

Prueba de la esfericidad de Mauchly

↓  $p > 0,05$  ↓  $p < 0,05$   
ESFERICIDAD NO ESFERICIDAD ↓

Contrastes UNIVARIADOS

Contrastes MULTIVARIADOS

↓  $p < 0,05$

$p < 0,05$  ↓

POST HOC: Procedimiento de BONFERRONI: Comparación por pares

## Prueba de la homogeneidad de las varianzas

### Prueba de la esfericidad de Mauchly

Efecto intra-sujetos	Prueba de esfericidad de Mauchly			
	W de Mauchly	Chi cuadrado	g.l	Significación
TIEMPO	0,803	6,096	5	0,297

Para el ANOVA de medidas repetidas es conveniente suponer que las varianzas de las diferencias entre cada dos niveles del factor TIEMPO son iguales. Con los 4 niveles, tenemos 6 pares de combinaciones dos a dos entre niveles; inicio-1º año, inicio-2º año, inicio-3º año, 1º año-2º año, 1º año-3º año, 2º año-3º año. Calculando las diferencias entre las puntuaciones de esos 6 pares, tendremos 6 nuevas variables. Este supuesto equivale a afirmar que la matriz de varianzas-covarianzas es circular o esférico. Para el procedimiento de medidas repetidas esto se realiza con la prueba de esfericidad de MAUCHLY. Puesto que la significación estadística asociada al estadístico W de Mauchly es de 0,297, (mayor de 0,05), NO podemos rechazar la hipótesis nula de esfericidad. La matriz de las varianzas-covarianzas es ESFÉRICA.

## Prueba de la esfericidad de Mauchly

Efecto intra-sujetos	Epsilon	Hunh-Feldt	Límite-interior
	Greenhous e-Geisser		
TIEMPO	0,865	0,958	0,333

Cuando el estadístico  $W$  de Mauchly nos lleve al rechazo de la esfericidad (NO ESFERICIDAD), podemos basar nuestra decisión en los estadísticos  $F$  MULTIVARIADOS que no les afecta la falta de esfericidad, o bien podemos utilizar el estadístico  $F$  UNIVARIADO aplicando el índice corrector EPSILON.

El índice corrector EPSILON expresa el grado en que la matriz de varianzas-covarianzas se aleja de la esfericidad. En condiciones de esfericidad perfecta,  $\epsilon$  vale 1. En este caso  $\epsilon$  es alto porque hay ESFERICIDAD.

Existen 3 estimaciones del índice  $\epsilon$ , Greenhous e-Geisser es la más conservadora. El límite inferior expresa el valor que adoptaría  $\epsilon$  en caso de incumplimiento extremo de la esfericidad. El valor de EPSILON se utiliza para corregir el valor del  $F$  UNIVARIADO. En este caso NO es necesario utilizar EPSILON porque se cumple la ESFERICIDAD ( $p > 0,05$ ).

TIEMPO	Contrastes MULTIVARIADOS				
	Valor	F	GI de la hipótesis	GI del error	Significación
Traza de Pillai	0,902	83,25	3	27	0,000
Lambda de Wilks	0,098	83,25	3	27	0,000
Traza de Hotelling	9,250	83,25	3	27	0,000
Raíz mayor de Roy	9,250	83,25	3	27	0,000

La tabla contiene 4 estadísticos multivariados: Traza de Pillai, Lambda de Wilks, Traza de Hotelling, Raíz mayor de Roy. Se interpretan de la misma forma que todos los estadísticos ya estudiados; puesto que la significación de cada uno de ellos ( $p < 0,001$ ) es menor que 0,05, podemos rechazar la hipótesis nula de igualdad de las medias y concluir que la calidad del recuerdo de los conocimientos de bioestadística adquiridos no es la misma en los 4 momentos temporales definidos en el factor tiempo.

$H_a$ : La calidad del recuerdo de los conocimientos de bioestadística NO es la misma en los 4 momentos considerados.

TIEMPO	Contrastes UNIVARIADOS: Prueba de efectos INTRA-SUJETOS				
	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación
Esfericidad asumida	984,067	3	328,022	103,236	0,000
Greenhouse-Geisser	984,067	2,596	379,006	103,236	0,000
Huynh-Feldt	984,067	2,874	342,361	103,236	0,000
Límite-inferior	984,067	1,000	984,067	103,236	0,000

Si se cumple la ESFERICIDAD (homogeneidad de las varianzas) es preferible utilizar la aproximación UNIVARIADA (esfericidad asumida), ya que en condiciones de esfericidad, el estadístico UNIVARIADO F ( $F = 103,236$ ) es más potente que los estadísticos MULTIVARIADOS de la tabla anterior ( $F = 83,246$ ), sobre todo con muestras pequeñas, aunque, si ambos estadísticos conducen a la misma conclusión es irrelevante utilizar uno u otro.

Observando los resultados de esta tabla vemos que las cuatro versiones del estadístico F UNIVARIADO (la no corregida y las 3 corregidas con el valor de EPSILON) conducen a la misma conclusión, que coincide con la ya alcanzada en la tabla de contrastes MULTIVARIADOS, por tanto, de nuevo rechazamos la  $H_0$  (igualdad de medias) y aceptamos la  $H_a$ .

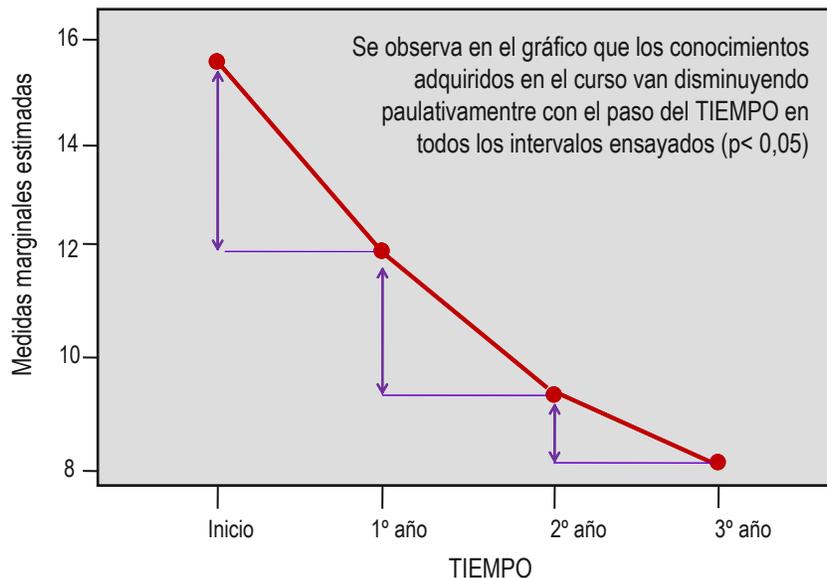
$H_a$ : la calidad del recuerdo de los conocimientos NO es la misma en los 4 momentos considerados.

TIEMPO	MEDIAS MARGINALES ESTIMADAS			
	Media	Error típico	Límite inferior del IC95%	Límite inferior del IC95%
INICIO	15,553	0,507	14,494	16,570
1º año	12,000	0,479	11,019	12,981
2º año	9,367	0,417	8,515	10,219
3º año	8,033	0,464	7,085	8,981

Se observa en la tabla que la media de los conocimientos adquiridos en el curso van disminuyendo paulatinamente con el paso del TIEMPO.

¿Esta disminución es estadísticamente significativa en cada uno de los estratos temporales?

## Gráfico de perfil representando el efecto del factor tiempo



¿Qué test has utilizado para llegar a esa conclusión?

TIEMPO	Pruebas de CONTRASTES INTRA-SUJETOS				
	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación
LINEAL	947,527	1	947,527	206,956	0,000
CUADRÁTICO	36,300	1	36,300	13,815	0,001
CÚBICO	0,240	1	0,240	0,103	0,750

La tabla ofrece contrastes polinómicos, de tipo ortogonal, mostrando tantos contrastes como niveles tiene el factor tiempo menos uno, como el factor tiempo tiene 4 niveles, aparecen 3 contrastes (lineal, cuadrático y cúbico).

Si miramos la significación (p) asociados a cada estadístico F podemos rechazar las hipótesis nulas referidas a los componentes lineal y cuadrático, pero no la referida al componente cúbico. Podemos concluir que las medias de los conocimientos adquiridos en el curso en cada momento temporal se ajustan significativamente tanto a una línea recta (componente lineal) como a una curva (componente cuadrático). Pero como existe más de un componente significativo se elige el de mayor orden (modelo lineal) (mayor valor de F, y mejor significación). El gráfico del perfil siguiente nos ayuda a comprender con más claridad lo que está ocurriendo.

## Pruebas Post Hoc Anova apareado

TIEMPOS	INICIO	1° AÑO	2° AÑO	3° AÑO
INICIO		$H_0 =$ $H_a \neq$	$H_0 =$ $H_a \neq$	$H_0 =$ $H_a \neq$
1° AÑO			$H_0 =$ $H_a \neq$	$H_0 =$ $H_a \neq$
2° AÑO				$H_0 =$ $H_a \neq$
3° AÑO				

## Procedimiento de Bonferroni: Comparación por pares

Tiempo (I)	Tiempo (J)	Diferencias de medias (I – J)	Error estándar	Significación
Inicio	1º año	3,533	0,417	0,000
	2º año	6,167	0,407	0,000
	3º año	7,500	0,544	0,000
1º año	Inicio	- 3,533	0,417	0,000
	2º año	2,633	0,405	0,000
	3º año	3,967	0,531	0,000
2º año	Inicio	- 6,167	0,401	0,000
	1º año	- 2,633	0,405	0,000
	3º año	1,333	0,440	0,031
3º año	Inicio	- 7,500	0,544	0,000
	1º año	- 3,9967	0,531	0,000
	2º año	- 1,333	0,440	0,031

## Relación entre las glucemias medias, la hemoglobina glicosilada y el riesgo de aparición de complicaciones de la diabetes

GLUCEMIAS MEDIAS	HbA1c (%)	RIESGO DE COMPLICACIONES		
60	4	<b>RIESGO BAJO</b>		
90	5			
120	6			
150	7	RIESGO MODERADO (control aceptable hasta 7,5%)		
180	8			
210	9	RIESGO AUMENTADO	<b>MAL CONTROL METABÓLICO (&gt; 9,5%)</b>	
240	10	RIESGO ALTO		
270	11	<b>RIESGO CRÍTICO</b>		
300	12			
330	13			
360	14			

Objetivo: Conocer si el curso de educación diabetológica mejora el control metabólico de un grupo de pacientes diabéticos mal controlados ( $HbA_{1c} > 8,0$ ).

Material: Grupo de 30 individuos antes y después de un programa de educación para el control diabético (ECD).

Método: Se evalúa la hemoglobina glicosilada antes ( $HbA_{1c} \text{ PRE}$ ) y después del ECD ( $HbA_{1c} \text{ POST}$ ).

Se define una nueva variable  $HbA_{1c} \text{ DIF} = HbA_{1c} \text{ PRE} - HbA_{1c} \text{ POST}$ .

Hipótesis nula  $H_0: X_{\text{antes}} = X_{\text{después}}$

$$(DIF = HbA_{1c} \text{ PRE} - HbA_{1c} \text{ POST} = 0)$$

Hipótesis alternativa  $H_1: X_{\text{antes}} \neq X_{\text{después}}$

$$(DIF = HbA_{1c} \text{ PRE} - HbA_{1c} \text{ POST} \neq 0)$$

¿El curso de educación diabetológica (ECD) mejora el control diabético de un grupo de pacientes mal controlados ( $HbA_{1c} > 8,0$ )?

Diabetes = Educación

$HbA_{1c}$  ANTES



$HbA_{1c}$  DESPUÉS

¿Cuál es la prueba estadística más adecuada?

## Resultados I° objetivo

HbA1c	N	Media	DE	$\gamma$ 1	EE $\gamma$ 1	$\gamma$ 2	EE $\gamma$ 2
INICIAL (PRE)	30	10,150	1,77	0,745	0,427	- 0,516	0,833
PRIMER AÑO (POST)	30	7,130	0,932	1,408	0,427	2,619	0,833

HbA1c	Test de Kolmogorov-Smirnov		
	N	Estadístico Z	Significación
INICIAL (PRE)	30	0,827	0,501
PRIMER AÑO	30	1,034	0,235

Experimento diseñado con dos objetivos:

1. Se evalúa el control metabólico con los niveles de HbA1c antes y después (al año) de la impartición de un curso intensivo de educación diabetológica a un grupo de 30 pacientes mal controlados ( $HbA1c > 8$ ) para comprobar si el curso ha sido útil en la mejora del control glucémico a CORTO plazo.
2. Estudiar el efecto del paso del tiempo (LARGO plazo) sobre el control metabólico tras la realización de un curso de educación diabetológica en un grupo de 30 pacientes con diabetes mal controlada ( $HbA1c > 8$ ), realizando controles sucesivamente cada año (2º, 3º y 4º año). Se quiere valorar si es necesario un nuevo curso y con que frecuencia debería realizarse.

## Resultados I° objetivo

Variables	Media	N	Desviación estándar (s)	Error estándar de la media (EEM)
HbA <sub>1c</sub> PRE	10,150	30	1,772	0,323
HbA <sub>1c</sub> POST	7,130	30	0,932	0,170

Variable	N	Correlación (r)	Significación
HbA <sub>1c</sub> PRE versus HbA <sub>1c</sub> POST	30	0,772	0,000

Variables	Diferencias relacionadas			Intervalo de confianza de la diferencia		t <sub>calculada</sub>	gl	Significación bilateral
	Media	s	EEM	Inferior	Superior			
HbA <sub>1c</sub> PRE – HbA <sub>1c</sub> POST	3,020	1,208	0,2206	2,569	3,471	13,692	29	0,000

$t_{\text{calculada}} < t_{\text{tabulada}} \rightarrow p > 0,05$  (aceptamos la Ho)

$t_{\text{calculada}} > t_{\text{tabulada}} \rightarrow p < 0,05$  (rechazamos la hipótesis nula)

## Resultados 2º objetivo

HbA1c	N	Media	DE	$\gamma_1$	EE $\gamma_1$	$\gamma_2$	EE $\gamma_2$
PRIMER AÑO	30	7,130	0,932	1,408	0,427	2,619	0,833
SEGUNDO AÑO	30	7,207	0,835	1,110	0,427	1,805	0,833
TERCER AÑO	30	7,707	0,712	1,056	0,427	0,998	0,833
CUARTO AÑO	30	9,050	0,910	0,652	0,427	- 0,084	0,833

HbA1c	Test de Kolmogorov-Smirnov		
	N	Estadístico Z	Significación
PRIMER AÑO	30	1,034	0,235
SEGUNDO AÑO	30	0,901	0,392
TERCER AÑO	30	0,990	0,281
CUARTO AÑO	30	0,737	0,649

2. Estudiar el efecto del paso del tiempo (LARGO plazo) sobre el control metabólico tras la realización de un curso de educación diabetológica en un grupo de 30 pacientes con diabetes mal controlada ( $HbA1c > 8$ ), realizando controles sucesivamente cada año (2º, 3º y 4º año). Se quiera valorar si es necesario un nuevo curso y con que frecuencia debería realizarse.

## Prueba de la homogeneidad de las varianzas

### Prueba de la esfericidad de Mauchly

Efecto intra-sujetos	Prueba de esfericidad de Mauchly			
	W de Mauchly	Chi cuadrado	g.l	Significación
HbA1c	0,368	27,717	5	0,000

Para el ANOVA de medidas repetidas es conveniente que las varianzas de las diferencias entre cada dos niveles del factor TIEMPO son iguales. Con los 4 niveles, tenemos 6 pares de combinaciones dos a dos entre niveles; 1º año-2º año, 1º año-3º año, 1º año-4º año, 2º año-3º año, 2º año-4º año, 3º año- 4º año. Calculando las diferencias entre las puntuaciones de esos 6 pares, tendremos 6 nuevas variables. Este supuesto equivale a afirmar que la matriz de varianzas es circular o esférico. Para el procedimiento de medidas repetidas esto se realiza con la prueba de esfericidad de MAUCHLY. Puesto que la significación estadística asociada al estadístico W de Mauchly es  $p < 0,001$ , (MENOR de 0,05), podemos rechazar la hipótesis nula de esfericidad.

NO HAY ESFERICIDAD.

## Prueba de la esfericidad de Mauchly

Efecto intra-sujetos	Epsilon	Hunh-Feldt	Límite-interior
	Greenhous e-Geisser		
TIEMPO	0,654	0,701	0,333

Cuando el estadístico  $W$  de Mauchly nos lleve al rechazo de la esfericidad (NO ESFERICIDAD), podemos basar nuestra decisión en los estadísticos  $F$  MULTIVARIADOS que no les afecta la falta de esfericidad, o bien podemos utilizar el estadístico  $F$  UNIVARIADO corregido aplicando el índice corrector EPSILON.

El índice corrector EPSILON expresa el grado en que la matriz de varianzas-covarianzas se aleja de la esfericidad. En condiciones de esfericidad perfecta,  $\epsilon$  vale 1. Cuanto menor es  $\epsilon$  menor es la esfericidad.

Existen 3 estimaciones del índice  $\epsilon$ , Greenhous e-Geisser es la más conservadora. El límite inferior expresa el valor que adoptaría  $\epsilon$  en caso de incumplimiento extremo de la esfericidad. El valor de EPSILON se utiliza para corregir el valor del  $F$  UNIVARIADO cuando no se cumple la esfericidad.

TIEMPO	Contrastes UNIVARIADOS: Prueba de efectos INTRA-SUJETOS				
	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación
Esfericidad asumida	71,079	3	23,693	238,282	0,000
Greenhouse-Geisser	71,079	1,961	36,238	238,282	0,000
Huynh-Feldt	71,079	2,102	33,810	238,282	0,000
Límite-inferior	71,079	1,000	71,079	238,282	0,000

Si se cumple la ESFERICIDAD es preferible utilizar la aproximación UNIVARIADA (versión esfericidad asumida), pues en condiciones de esfericidad el estadístico F UNIVARIADO es más potente que los estadísticos MULTIVARIADOS, sobre todo con muestras pequeñas.

Si NO se cumple la ESFERICIDAD es preferible utilizar los estadísticos F MULTIVARIADOS (como en este caso).

## Comparación de $>$ de 2 grupos apareados

Test de Kolmogorov-Smirnov

↓  $p > 0,05$ ; Normalidad

Prueba de la esfericidad de Mauchly

↓  $p > 0,05$  ↓  $p < 0,05$   
ESFERICIDAD NO ESFERICIDAD ↓

Contrastes UNIVARIADOS

Contrastes MULTIVARIADOS

↓  $p < 0,05$  ↓  $p < 0,05$  ↓

POST HOC: Procedimiento de BONFERRONI: Comparación por pares

TIEMPO	Contrastes MULTIVARIADOS				
	Valor	F	GI de la hipótesis	GI del error	Significación
Traza de Pillai	0,931	120,741	3	27	0,000
Lambda de Wilks	0,069	120,741	3	27	0,000
Traza de Hotelling	13,416	120,741	3	27	0,000
Raíz mayor de Roy	13,416	120,741	3	27	0,000

La tabla contiene 4 estadísticos multivariados: Traza de Pillai, Lambda de Wilks, Traza de Hotelling, Raíz mayor de Roy. Se interpretan de la misma forma que todos los estadísticos ya estudiados; puesto que la significación de cada uno de ellos ( $p = 0,000$ ) es menor que 0,05, podemos rechazar la hipótesis nula de igualdad de las medias y concluir que el control metabólico debido a la educación del diabético por el curso no es el mismo en los 4 momentos temporales definidos en el factor tiempo.

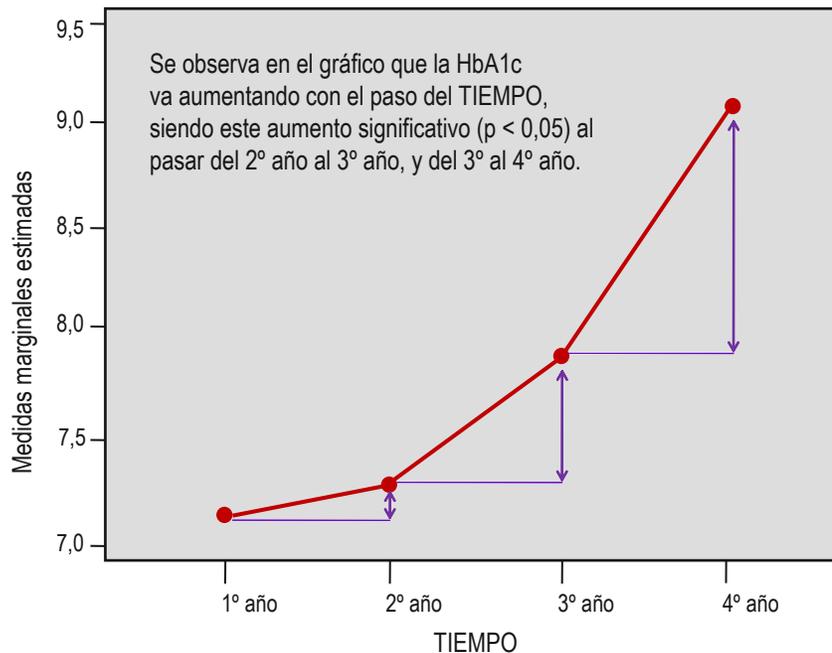
$H_a$ : el control metabólico debido a la educación del diabético por el curso no es la misma en los 4 momentos temporales definidos en el factor tiempo.

TIEMPO	MEDIAS MARGINALES ESTIMADAS			
	Media	Error típico	Límite inferior del IC95%	Límite inferior del IC95%
1º año	7,130	0,170	6,782	7,478
2º año	7,207	0,152	6,895	7,518
3º año	7,707	0,130	7,441	7,973
4º año	9,050	0,166	8,710	9,390

Se observa en la tabla que la media de la hemoglobina glicosilada va aumentando paulatinamente con el paso del TIEMPO.

¿Este aumento es estadísticamente significativo en cada uno de los estratos?

## Gráfico de perfil representando el efecto del factor tiempo

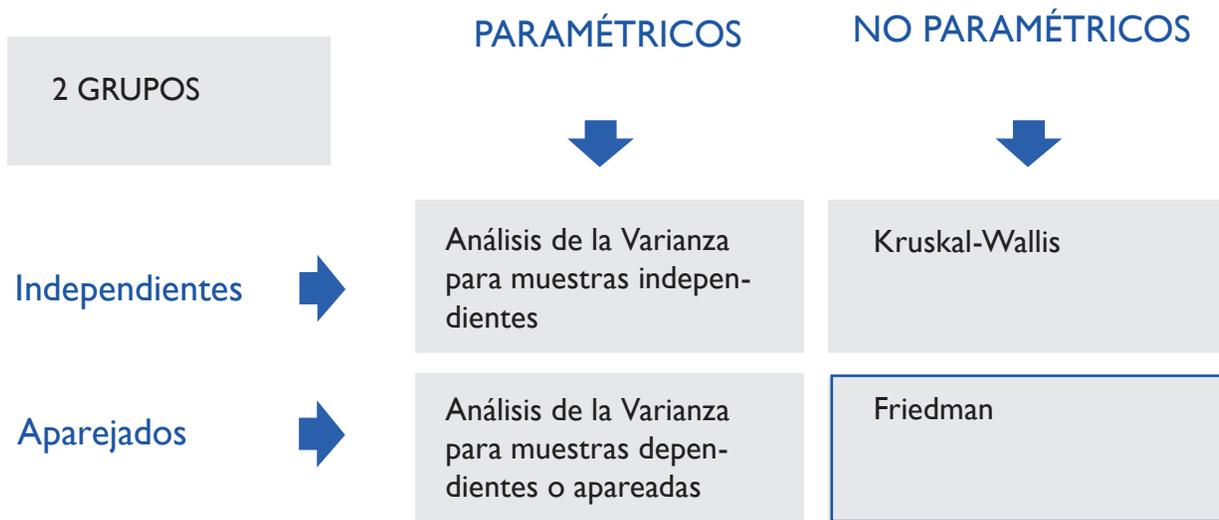


¿Qué test has utilizado para llegar a esa conclusión?

## Procedimiento de Bonferroni: Comparación por pares

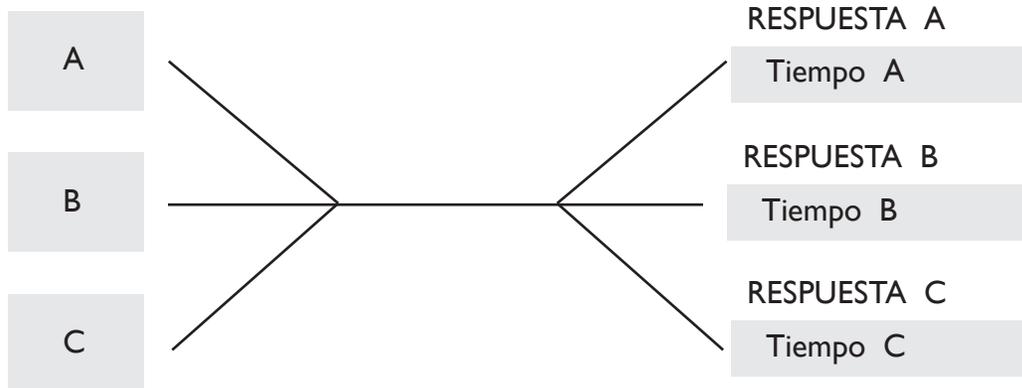
Tiempo (I)	Tiempo (J)	Diferencias de medias (I – J)	Error estándar	Significación
1º año	2º año	- 0,077	0,041	0,425
	3º año	0,577	0,070	0,000
	4º año	- 1,920	0,098	0,000
2º año	1º año	0,077	0,041	0,425
	3º año	- 0,500	0,068	0,000
	4º año	- 1,843	0,104	0,000
3º año	1º año	0,577	0,070	0,000
	2º año	0,500	0,068	0,000
	4º año	- 1,343	0,090	0,000
4º año	1º año	1,920	0,098	0,000
	2º año	1,843	0,104	0,000
	3º año	1,343	0,090	0,000

## Cualitativa policotómica versus cuantitativa



Variable predictora Cualitativa versus variable resultado Cuantitativa.

## ¿Existen diferencias en el tiempo de reacción frente a los 3 fármacos ensayados?



¿Cuál es el test estadístico más adecuado?

## Test de Friedman

- Es la versión NO PARAMÉTRICA del ANOVA para muestras **apareadas**.
- Utiliza la modificación estadística de Iman & Davenport (1980) que lo hace menos conservador.
- El valor de z de cada hipótesis se puede utilizar para calcular la probabilidad de error (valor p) mediante la distribución normal.
- ¿Se cumplen las condiciones impuestas por el test ANOVA?  
SI: Utilizar ANOVA  
NO: Utilizar el test de FRIEDMAN

$$z = (R_i - R_j) \sqrt{\frac{M(M+1)}{6N}}$$

Se ha medido el tiempo de la reacción farmacológica de 10 pacientes a 3 medicamentos diferentes.

Paciente	Fármaco A	Fármaco B	Fármaco C
1º	10	10	15
2º	10	15	20
3º	11	15	12
4º	8	12	10
5º	7	12	9
6º	15	10	15
7º	14	12	18
8º	10	14	17
9º	9	9	12
10º	10	14	16

$H_0$ : No existen diferencias entre los tiempos de reacción frente a los 3 fármacos.

$H_a$ : Existen diferencias entre los tiempos de reacción frente a los 3 fármacos.

## Test de Friedman

Estadística descriptiva					
Tratamiento	N	Mean	Standars Desviation	Minimum	Maximum
Fármaco A	10	10,40	2,458	7,00	15,00
Fármaco B	10	12,30	2,162	9,00	15,00
Fármaco C	10	14,40	3,565	8,00	20,00

Ranks	
Tratamiento	Promedio de rangos
Fármaco A	1,35
Fármaco B	2,00
Fármaco C	2,65

Friedman test Statistics	
N	10
Chi cuadrado	9,135
gl (K – 1)	2
Significación	0,010

$H_a$ : Existen diferencias entre los tiempos de reacción frente a los 3 fármacos.

## Coeficiente de concordancia $W$ de Kendall

- Mide el grado de concordancia entre un grupo de elementos y un grupo de características.
- Si la concordancia es máxima  $W=1$ , si la concordancia es mínima  $W=0$ .

$$W = \frac{S}{1/12 K^2 (N^3 - N) - K \sum L_i}$$

Donde:

$w$  = coeficiente de concordancia de Kendall.

$S$  = suma de los cuadrados de las diferencias observadas con respecto a un promedio.

$N$  = Tamaño de la muestra en función del número de tripletes, tetrapletes, quintupletes, etc.

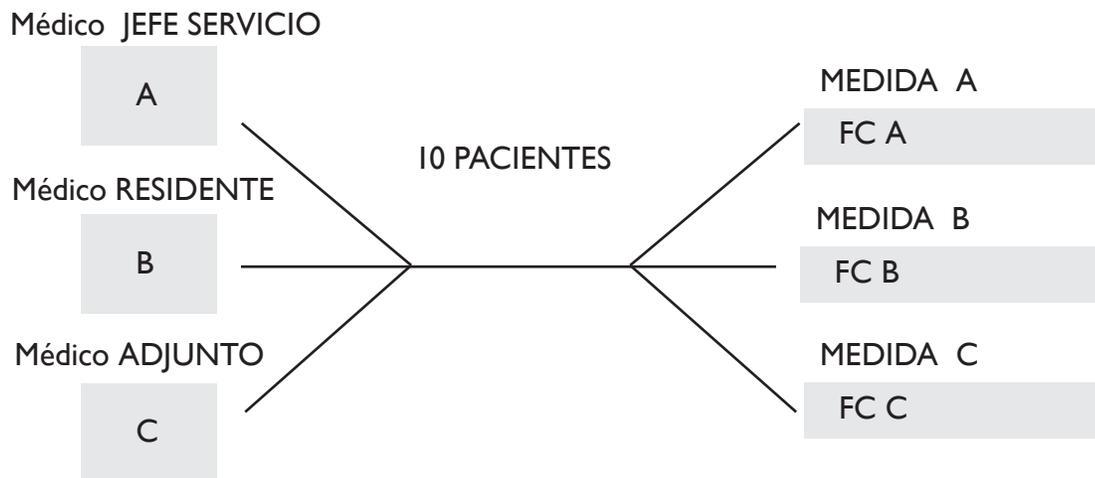
$K$  = número de variables incluidas.

$L_i$  = sumatoria de las ligas o empates entre los rangos.

## Coeficiente de concordancia $W$ de Kendall

- El coeficiente  $W$  de Kendall y el estadístico  $\chi^2$  de Friedman son aplicables al mismo tipo de situaciones.
- La hipótesis nula de que los tratamientos aplicados son iguales en su efectividad es exactamente lo mismo que mantener la hipótesis nula de ausencia de concordancia en los resultados de los tratamientos ensayados.

## ¿Existen diferencias en la medida de la frecuencia cardíaca en los tres médicos evaluados?



¿Cuál es el test estadístico más adecuado?

Comparación de las mediciones consecutivas de la FRECUENCIA CARDÍACA en pulsaciones por minuto (ppm) en 10 pacientes realizada por tres médicos diferentes: un médico A, jefe del servicio, un médico B en periodo de formación, y un médico C adjunto del mismo departamento.

Pacientes	Médico JEFE SERVICIO (A)	Medico RESIDENTE (B)	Médico ADJUNTO (C)
1	75	80	74
2	74	84	72
3	76	81	75
4	79	83	80
5	82	92	84
6	83	88	82
7	85	90	84
8	87	92	86
9	87	92	88
10	88	93	88
Media	81,6	87,5	81,3
DE	5,3	5,04	5,85

## Test de Friedman

Estadísticos descriptivos					
Médico	N	Media	Desviación estándar	Mínimo	Máximo
JEFE A	10	81,6	5,23	74	88
RESIDENTE B	10	87,5	5,04	80	93
ADJUNTO C	10	81,3	5,85	72	88

Rangos	
	Promedio de rangos
JEFE A	1,35
RESIDENTE B	2,00
ADJUNTO C	2,65

Estadístico de contraste prueba de Friedman	
N	10
Chi cuadrado	15,846
gl (K - 1)	2
Significación	0,000

$H_a$ : Existen diferencias entre la medida de la frecuencia cardiaca en los tres médicos.

## Coeficiente de concordancia $W$ de Kendall

Rangos	
Promedio de rangos	
JEFE A	1,35
RESIDENTE B	2,00
ADJUNTO C	2,65

Estadístico de contraste prueba de Kendall	
N	10
W de Kendall (*)	0,792
Chi cuadrado	15,846
gl (K - 1)	2
Significación	0,000

(\*) Coeficiente de concordancia de Kendall

$H_a$ : Existen diferencias entre la medida de la frecuencia cardiaca en los tres médicos.

Experimento diseñado para estudiar el efecto del paso del tiempo sobre la calidad del recuerdo y realización práctica de estudios con aplicación de la estadística de los 10 alumnos que han realizado el curso de BIOESTADÍSTICA APLICADA A LAS CIENCIAS DE LA SALUD. Se evalúa el grado de conocimientos con el mismo examen de 20 preguntas tipo test en 4 momentos diferentes; inmediatamente después de finalizar el curso, al año, a los 2 años, y a los 3 años del curso.

¿La calidad y cantidad del recuerdo de los conocimientos sobre bioestadística es la misma en los 4 momentos considerados?

Respuestas acertadas en función del tiempo transcurrido desde la realización del curso				
Alumnos	Tiempo 0	1º año	2º año	3º año
1º	16	8	8	12
2º	12	9	9	10
3º	12	10	10	8
4º	15	13	7	11
5º	18	12	12	12
6º	13	13	8	10
7º	18	16	10	13
8º	15	9	6	6
9º	20	9	11	8
10º	14	12	11	9

¿Cuál es el test estadístico más adecuado?

## Test de Friedman

Estadísticos descriptivos					
Tiempo	N	Media	Desviación estándar	Mínimo	Máximo
INICIO	10	15,30	2,710	12	20
1º AÑO	10	11,10	2,514	8	16
2º AÑO	10	9,20	1,932	6	12
3º AÑO	10	9,90	2,183	6	13

Rangos	
	Rango promedio
INICIO	3,95
1º AÑO	2,50
2º AÑO	1,70
3º AÑO	1,85

Estadístico de contraste prueba de Friedman	
N	10
Chi cuadrado	20,868
gl (K – 1)	3
Significación	0,000

$H_a$ : la calidad del recuerdo de los conocimientos no es la misma en los 4 momentos considerados.

## Coeficiente de concordancia $W$ de Kendall

Rangos	
	Rango promedio
INICIO	3,95
1º AÑO	2,50
2º AÑO	1,70
3º AÑO	1,85

Estadístico de contraste prueba de Kendall	
N	10
W de Kendall (*)	0,698
Chi cuadrado	20,868
gl (K - 1)	3
Significación	0,000

(\*) Coeficiente de concordancia de Kendall

$H_a$ : La calidad del recuerdo de los conocimientos no es la misma en los 4 momentos considerados

# Capítulo 10

## Publicación Bioestadística: Medidas de concordancia. Variables cualitativas.

Dr. Jacobo Díaz Portillo  
Coordinador de Formación Continuada  
Hospital Universitario de Ceuta

## Medidas de concordancia

- Fiabilidad de los procedimientos de medida empleados.
- Fleis: “ni el más elegante de los diseños sería capaz de paliar el daño causado por un sistema de medida poco fiable”.
- Error de medida: Variabilidad entre observadores.
- Estudios de fiabilidad.
  - Sesgo entre observadores.
  - Concordancia entre observadores.
- Indicadores del grado de variabilidad.
  - Coeficiente de correlación intraclase (variables cuantitativas).
  - Índice de Kappa (variables cualitativas tipo categóricos).

Valoración de la fiabilidad externa de las publicaciones de dos revistas científicas durante el último quinquenio.

En 5 (4 de JAMA y uno de MC) de los 12 artículos se han llevado a cabo tratamientos estadísticos no idóneos para medir la concordancia, mediante las siguientes pruebas de contraste de hipótesis:

- Comparación de medias con la t de Student
- Test exacto de Fisher,
- Test de chi-cuadrado  $\chi^2$
- Análisis de la varianza (ANOVA)
- Test de Wilcoxon
- Coeficiente de Spearman

## Análisis de concordancia: Variables Cualitativas

- Índice kappa de Cohen (2 observadores)
- Índice kappa de Fleiss ( $> 2$  observadores)
- Índice kappa ponderado
- Prueba de homogeneidad de kappas

## Índice kappa de Cohen (medida dicotómica)

OBSERVADOR A (ó 1ª medición)	OBSERVADOR B (ó 2ª medición)		TOTAL
	Anormal (+)	Normal (-)	
Anormal (+)	a	b	a + b
Normal (-)	c	d	c + d
Total	a + c	b + d	N

Statistical methods for rates and proportions. Joseph L. Fleis. Ed. John Wiley New York 1981.

Dos radiólogos del programa de tuberculosis se les remiten radiografías de tórax de 170 sujetos que están controlados en una unidad de neumología, y que se quiere estimar el grado de concordancia entre ellos. Los radiólogos A y B tienen que clasificar cada radiografía en una de dos categorías:

(+) “positiva” (sospechosa de lesión tuberculosa).

(-) “negativa” (no sospechosa de lesión tuberculosa).

RADIÓLOGO A	RADIÓLOGO B		TOTAL
	Anormal (+)	Normal (-)	
Anormal (+)	58	39	97
Normal (-)	12	61	73
Total	70	100	170

$k = 0,415$

$IC95\% = 0,2862 - 0,5430$

## ¿Cuál es el grado de concordancia encontrado?

Medida de acuerdo	Valor de Kappa	EEk	Z	Significación
Kappa de Cohen	+ 0,415	0,0655	5,688	0,000
Nº de casos válidos	200			

Junto al índice kappa aparece su error estándar EEk y el valor del estadístico Z.

Puesto que el valor de p es muy pequeño ( $<0,001$ ) podemos rechazar la hipótesis nula y concluir que existe un ACUERDO significativamente más alto que el esperado por el azar.

¿Cómo se clasifica el grado de concordancia encontrado?

## Índice kappa de Cohen

$k = 0$  cuando la concordancia observada  $p_o$  es la que se esperada a causa del azar  $p_e$

$k > 0$  cuando la concordancia observada  $p_o$  es mayor de la esperada sólo por el azar  $p_e$

$$k = \frac{\sum \left( \frac{X_{ii}}{n} \right) \left( \sum \left( \frac{X_{i1}}{n} \right) \left( \frac{X_{1i}}{n} \right) \right)}{1 - \sum \left( \frac{X_{i1}}{n} \right) \left( \frac{X_{1i}}{n} \right)} = \frac{n \sum X_{ii} - X_{1i} X_{i1}}{n^2 - \sum X_{i1} X_{1i}} \quad \hat{k} = \frac{p_o - p_e}{1 - p_e}$$

Valoración del Índice Kappa	
Valor de k	Fuerza de concordancia
< 0,20	Pobre
0,21 – 0,40	Débil
0,41 – 0,60	Moderada 
0,61 – 0,80	Buena
0,81 – 1,00	Muy buena

Con los datos de la tabla se obtiene aplicando esta fórmula un valor de  $k = 0,415$  (IC95%: 0,2862 – 0,5430), que según la clasificación anterior calificaríamos como una MODERADA concordancia.

## Pregunta de investigación

¿Tienen los dos radiólogos de mi servicio el mismo criterio-juicio clínico para informar las radiografías?

Radiólogo 1	Radiólogo 2		
	Anormal	Dudosa	Normal
Anormal	18	4	3
Dudosa	1	10	5
Normal	2	4	53

Se enfrenta ambos radiólogos a la tarea de categorizar una misma muestra de 100 radiografías mediante la escala: “anormal”, “dudosa”, “normal”.

¿Cuál es el test estadístico más acertado?

López de Ullibarri Galparsoro I, Pita Fernández S. Medidas de concordancia: el índice Kappa. Cad Aten Primaria 1999; 6: 169-171.

## Índice kappa de Fleiss

Medida de acuerdo	Valor de Kappa	EEk	T <sub>aprox</sub>	Significación
Kappa de Fleiss	+ 0,660	0,044	8,944	0,000
Nº de casos válidos	100			

## Coefficiente Gamma

Gamma	EEk	T <sub>aprox</sub>	p
0,899	0,044	8,744	0,000
Nº de casos válidos	100		

¿Cómo se clasifica el grado de concordancia encontrado?

## Índice kappa de Fleiss

$k = 0$  cuando la concordancia observada  $p_o$  es la que se esperada a causa del azar  $p_e$

$k > 0$  cuando la concordancia observada  $p_o$  es mayor de la esperada sólo por el azar  $p_e$

$$k = \frac{\sum \left( \frac{X_{ii}}{n} \right) \left( \sum \left( \frac{X_{i1}}{n} \right) \left( \frac{X_{i1}}{n} \right) \right)}{1 - \sum \left( \frac{X_{i1}}{n} \right) \left( \frac{X_{i1}}{n} \right)} = \frac{n \sum X_{ii} - X_i X_i}{n^2 - \sum X_i X_i} \quad \hat{k} = \frac{p_o - p_e}{1 - p_e}$$

Valoración del Índice Kappa	
Valor de k	Fuerza de concordancia
< 0,20	Pobre
0,21 – 0,40	Débil
0,41 – 0,60	Moderada
0,61 – 0,80	Buena 
0,81 – 1,00	Muy buena

Con los datos de la tabla se obtiene aplicando esta fórmula un valor de  $k = 0,660$  (IC95%: 0,527-0,793), que según la clasificación anterior calificaríamos como una buena concordancia.

## Pregunta de investigación

¿Tienen los dos psiquiatras de mi servicio el mismo criterio-juicio clínico al diagnosticar una muestra de 200 pacientes?

	Psiquiatra 2			
Psiquiatra 1	Fóbica	Histórica	Obsesiva	Depresiva
Fóbica	20	8	6	1
Histórica	7	36	14	4
Obsesiva	1	8	43	7
Depresiva	2	6	4	33

¿Cuál es el test estadístico más acertado?

Medida de acuerdo	Valor de Kappa	EEk	T <sub>aprox</sub>	Significación
Kappa de Fleiss	+ 0,538	0,046	12,921	0,000
Nº de casos válidos	200			

Junto al índice kappa aparece su valor estandarizado (T aproximada), resultado similar al resultante de dividir el valor de kappa por su error estándar (EEk) (11,69).

Puesto que el valor de p es muy pequeño ( $<0,001$ ) podemos rechazar la hipótesis nula y concluir que existe un ACUERDO significativamente más alto que el esperado por el azar.

¿Cómo se clasifica el grado de concordancia encontrado?

## Índice kappa de Fleiss

$k = 0$  cuando la concordancia observada  $p_o$  es la que se esperada a causa del azar  $p_e$

$k > 0$  cuando la concordancia observada  $p_o$  es mayor de la esperada sólo por el azar  $p_e$

$$k = \frac{\sum \left( \frac{X_{ij}}{n} \right) \left( \sum \left( \frac{X_{i1}}{n} \right) \left( \frac{X_{i2}}{n} \right) \right)}{1 - \sum \left( \frac{X_{i1}}{n} \right) \left( \frac{X_{i2}}{n} \right)} = \frac{n \sum X_{ij} - X_i X_j}{n^2 - \sum X_i X_j} \quad \hat{k} = \frac{p_o - p_e}{1 - p_e}$$

Valoración del Índice Kappa	
Valor de k	Fuerza de concordancia
< 0,20	Pobre
0,21 – 0,40	Débil
0,41 – 0,60	Moderada 
0,61 – 0,80	Buena
0,81 – 1,00	Muy buena

Con los datos de la tabla se obtiene aplicando esta fórmula un valor de  $k = 0,538$ , que según la clasificación anterior calificaríamos como una MODERADA concordancia.

## Pregunta de investigación

¿Podemos sustituir el diabur-test por el clinitest para la medida de la glucosuria?

Clinitest



Diabur-test 5000

¿Cuál es el test estadístico más acertado?

Banauch D, Koller PU, Bablok W. Evaluation of Diabur-Test 5000: a cooperative study carried out at 12 diabetes centers. *Diabetes Care* 1983; 6(3): 213-18.

## Pregunta de investigación

Se desea estimar la concordancia entre dos pruebas diagnósticas de diabetes, el Diabur-Test5000 y el Clinitest en 1677 muestras de orina. Calcula el valor de kappa sin ponderar y ponderado por pesos cuadráticos.

Clinitest	Diabur-test 5000					
	Negativo	Trazas	+	++	+++	++++
Negativo	452	5				
Trazas	133	270	28	1	2	
+	4	36	107	5	2	2
++		5	53	76	2	4
+++			12	28	81	35
++++			2	11	44	251

Índice kappa Ponderado

Kappa ponderado. Si las categorías de clasificación son más de dos y están definidas en una escala ordinal, como por ejemplo “sano”, “posiblemente enfermo” y “claramente enfermo”. A la hora de valorar el grado de discrepancia entre dos observadores, no es lo mismo que uno clasifique a un sujeto como “posiblemente enfermo” y el otro lo declare “sano” a que uno lo clasifique como “sano” y el otro como “claramente enfermo”. La “distancia” entre ambas discrepancias no es la misma.

Cuando ciertos tipos de desacuerdo son más acusados que otros, al investigador le puede interesar tenerlo en cuenta en la construcción del índice kappa. Para ello, se PONDERA las diferentes discrepancias, usando una matriz de pesos que pueden variar según el criterio del investigador en función de lo que esté analizando, aunque siempre cumpliendo ciertas restricciones:

1. La matriz debe ser simétrica;
2. La diagonal de acuerdos se le asigna el máximo peso, que es 1 y el resto de pesos deben ser inferiores, aunque siempre positivos o iguales a cero.

Medida de acuerdo	Valor de Kappa	EEk	Z	Significación
Kappa NO ponderado	+ 0,6706	0,013	57,098	0,0000
Nº de casos válidos	1677	IC95%: 0,6450 – 0,6961		

Medida de acuerdo	Valor
Acuerdo observado	+ 0,7376
Acuerdo esperado	+ 0,2035

$$k = \frac{P_o - P_e}{1 - P_e}$$

Medida de acuerdo	Valor de Kappa	EEk	Z	Significación
Kappa ponderado	+ 0,9491	0,0033	38,982	0,0000
Nº de casos válidos	1677	IC95%: 0,9427 – 0,9555		

Medida de acuerdo	Valor
Acuerdo observado	+ 0,9856
Acuerdo esperado	+ 0,7165

$$k = \frac{P_o - P_e}{1 - P_e}$$

## Variabilidad en datos clínicos

### Grado de acuerdo interobservador

Variables	Acuerdo preadiestramiento	Valor
(CCI / Kappa)	Acuerdo postadiestramiento	+ 0,9856
(CCI / Kappa)	+ 0,7165	+ 0,7165
Afectación simétrica	0.14	0.79
Clase funcional	-0.08	0.65
Numero criterios diagnósticos	0.02	0.93
Sexo	1	1
Anticuerpos antinucleares	0.62	0.84
Valor máximo VSG	0.70	0.99

A. Gómez de la Cámara et al. Med Clin (Barc) 1997; 108:377-381

## Limitaciones del estadístico Kappa

- El valor de kappa se ve afectado por la prevalencia del rasgo estudiado. Por tanto, es necesario ser cuidadoso a la hora de generalizar los resultados de comparación de observadores en situaciones con prevalencias diferentes; esto quiere decir que kappa es un estadístico descriptivo útil, pero es inadecuado con fines de predicción o inferencia.
- Kappa es dependiente del número de categorías. Cuantas más categorías se estén considerando, más difícil será clasificar correctamente los sujetos de observación, lo que implica valores de kappa más bajos. Por tanto, debe tenerse en cuenta el número de categorías a la hora de interpretar kappa.
- Para datos ordinales derivados de categorizar variables continuas, kappa depende fuertemente las definiciones arbitrarias que se hacen de las categorías.

Se desea evaluar los conocimientos prácticos adquiridos por los residentes de microbiología a su paso por el departamento de micobacterias. Para ello se mide la concordancia con el Especialista adjunto mediante la evaluación conjunta de un determinado número de tinciones de Ziehl Neelsen. Se desea estimar el grado de concordancia de cada residente con el adjunto, y si el grado de adiestramiento ha sido el mismo en los 4 residentes que han rotado por el departamento.

RESIDENTE



ADJUNTO

¿Cuál es el test estadístico más acertado?

## Prueba de homogeneidad de las kappas

Hipótesis **nula**:  $k_1 = k_2 = k_3 = \dots = k_n$

Hipótesis **alternativa**: al menos una de las kappas  $k_x$  es distinta

Comparación de kappas: Distribución de  $\chi^2$  de *Pearson*

Hipótesis nula  $H_0$ : Igualdad de las kappas comparadas. El grado de aprendizaje no varía entre los distintos residentes.

Hipótesis alternativa  $H_a$ : Desigualdad de las kappas comparadas. El grado de aprendizaje varía entre los distintos residentes. Existe al menos una kappa diferente al resto.

RESIDENTE R1	ADJUNTO		TOTAL
	Anormal (+)	Anormal (-)	
Anormal (+)	320	30	350
Dudosa (-)	120	29	149
Total	440	59	499

P = 72,7%

Kappa	EEk	IC95%	T <sub>aprox</sub>	p
0,132	0,043	0,048 – 0,215	3,448	0,001

Gamma (Q de Yule)	EEk	T <sub>aprox</sub>	p
0,441	0,113	3,027	0,002

RESIDENTE R2	ADJUNTO		TOTAL
	Anormal (+)	Anormal (-)	
Anormal (+)	890	210	1100
Dudosa (-)	290	700	990
Total	1180	59	2090

P = 75,4%

Kappa	EEk	IC95%	T <sub>aprox</sub>	p
0,518	0,019	0,481 – 0,555	23,764	0,000

Gamma (Q de Yule)	EEk	T <sub>aprox</sub>	p
0,822	0,017	27,549	0,000

RESIDENTE R3	ADJUNTO		TOTAL
	Anormal (+)	Anormal (-)	
Anormal (+)	350	120	470
Dudosa (-)	70	550	620
Total	420	670	1090

P = 83,3%

Kappa	EEk	IC95%	T <sub>aprox</sub>	p
0,640	0,024	0,593 – 0,686	21,226	0,000

Gamma (Q de Yule)	EEk	T <sub>aprox</sub>	p
0,916	0,013	25,914	0,000

RESIDENTE R4	ADJUNTO		TOTAL
	Anormal (+)	Anormal (-)	
Anormal (+)	280	80	360
Dudosa (-)	60	550	610
Total	340	630	970

P = 82,3%

Kappa	EEk	IC95%	T <sub>aprox</sub>	p
0,687	0,024	0,639 – 0,735	21,425	0,000

Gamma (Q de Yule)	EEk	T <sub>aprox</sub>	p
0,940	0,011	24,63	0,000

## Comparación de kappas

Número de Kappas a comparar	Nivel de confianza
4	95%

Kappa global	IC95%
0,687	0,5379 – 0,5855

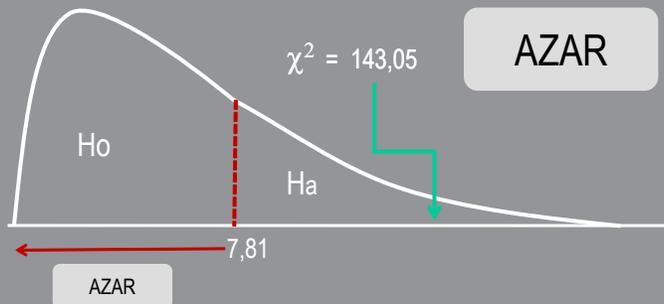
Prueba de la homogeneidad de kappas		
Chi cuadrado	gl	Valor p
143,0515	3	0,0000

Hipótesis alternativa  $H_a$ : Desigualdad de las kappas comparadas. El grado de aprendizaje varia entre los distintos residentes. Existe al menos una kappa diferente al resto.

## Comparación de kappas: Distribución de $\chi^2$ de Pearson

$$\text{Grados de libertad} = [K - 1] [F - 1] = (2-1) (4 - 1) = 3$$

$$\chi^2_{\text{Pearson calculado}} = 143,05 > \chi^2_{\text{Pearson tabulado}} = 7,81$$



Hipótesis alternativa  $H_a$ : Desigualdad de las kappas comparadas. El grado de aprendizaje varía entre los distintos residentes. Existe al menos una kappa diferente al resto.

## Mito I: “La sonrisa de la esfinge”

El valor de kappa se ve afectado por la prevalencia del rasgo estudiado. Por tanto, es necesario ser cuidadoso a la hora de generalizar los resultados de comparación de observadores en situaciones con prevalencias diferentes; esto quiere decir que kappa es un estadístico descriptivo útil, pero es inadecuado con fines de predicción o inferencia.

¿La prevalencia (porcentaje de tinciones positivas valoradas) es la misma en la evaluación de los 4 residentes comparados?

Si no son iguales NO podemos comparar los índices kappa.

Gardner PL. Measuring attitudes to science: Unidimensionality and internal consistency revisited. *Research in Science Education* 1995; 25: 283-9.

## Capítulo II

# Publicación Bioestadística: Medidas de concordancia. Variables cuantitativas.

Dr. Jacobo Díaz Portillo  
Coordinador de Formación Continuada  
Hospital Universitario de Ceuta

## Medidas de concordancia

- Fiabilidad de los procedimientos de medida empleados.
- Fleis: “ni el más elegante de los diseños sería capaz de paliar el daño causado por un sistema de medida poco fiable”.
- Error de medida: Variabilidad entre observadores.
- Estudios de fiabilidad.
  - Sesgo entre observadores.
  - Concordancia entre observadores.
- Indicadores del grado de variabilidad.
  - Coeficiente de correlación intraclase (variables cuantitativas).
  - Índice de Kappa (variables cualitativas tipo categóricos).

Valoración de la fiabilidad externa de las publicaciones de dos revistas científicas durante el último quinquenio.

En 5 (4 de JAMA y uno de MC) de los 12 artículos se han llevado a cabo tratamientos estadísticos no idóneos para medir la concordancia, mediante las siguientes pruebas de contraste de hipótesis:

- Comparación de medias con la t de Student
- Test exacto de Fisher,
- Test de chi-cuadrado  $\chi^2$
- Análisis de la varianza (ANOVA)
- Test de Wilcoxon
- Coeficiente de Spearman

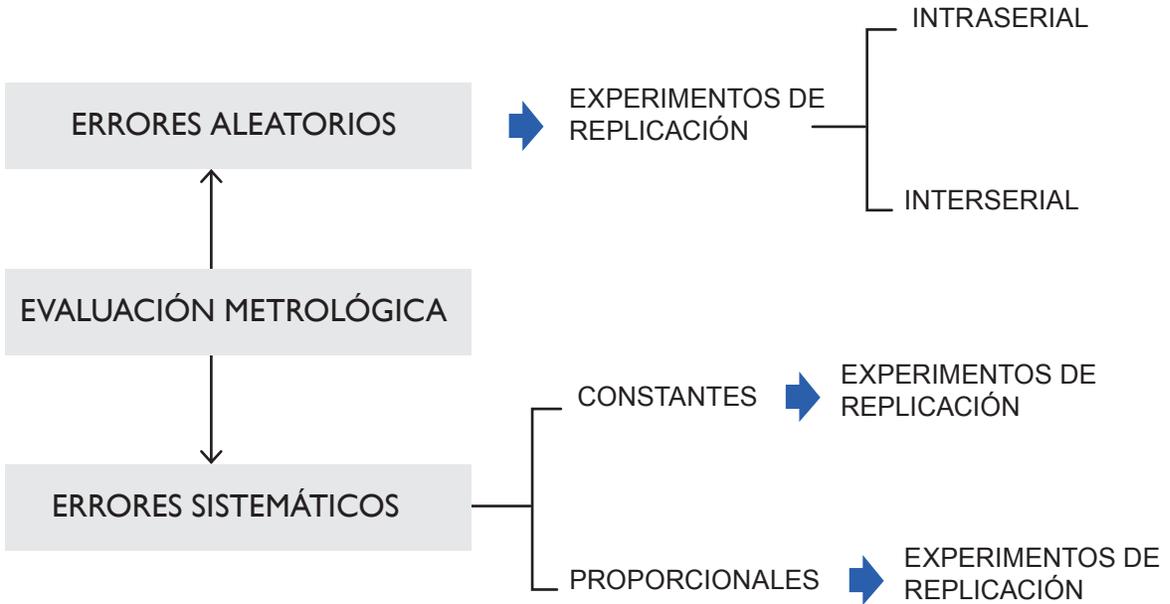
## Evaluación de métodos diagnósticos cuantitativos

- Variación metrológica.
  - Errores aleatorios.
  - Errores sistemáticos.
- Comparación de métodos.

## Evaluación metrológica

La evaluación metrológica de una prueba diagnóstica CUANTITATIVA persigue detectar los errores aleatorios y sistemáticos que origina, para cuantificarlos mediante cálculos estadísticos y determinar si son inferiores a los límites permisibles que no comprometen a la utilización clínica de los resultados.

Esta evaluación nos permite conocer el error que cometemos en nuestras condiciones de trabajo, y si el la prueba es adecuada para nuestros fines o es necesario cambiarlo por otros métodos.



## Variación metrológica

$$CV_a = \frac{SD}{\bar{X}} * 100$$

Error aleatorio

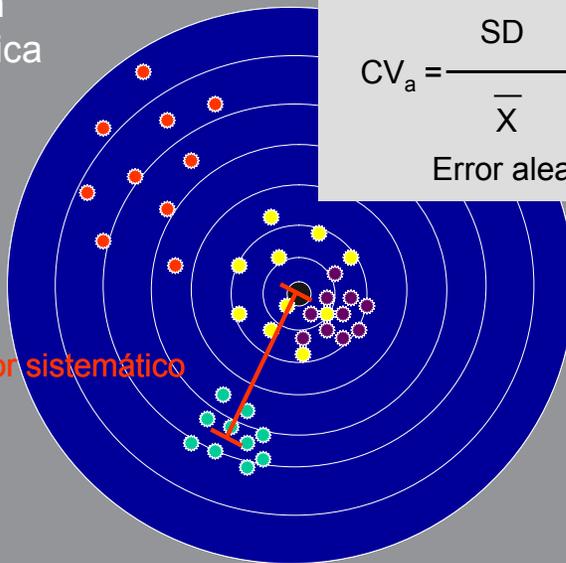
Error sistemático

● Inexactitud y imprecisión

● Inexactitud y precisión

● Exactitud y imprecisión

● Exactitud y precisión



## Especificaciones de la calidad para la imprecisión analítica

Prestación óptima

$$CV_a < \frac{1}{4} CV_{Bw}$$

Prestación deseable

$$CV_a < \frac{1}{2} CV_{Bw}$$

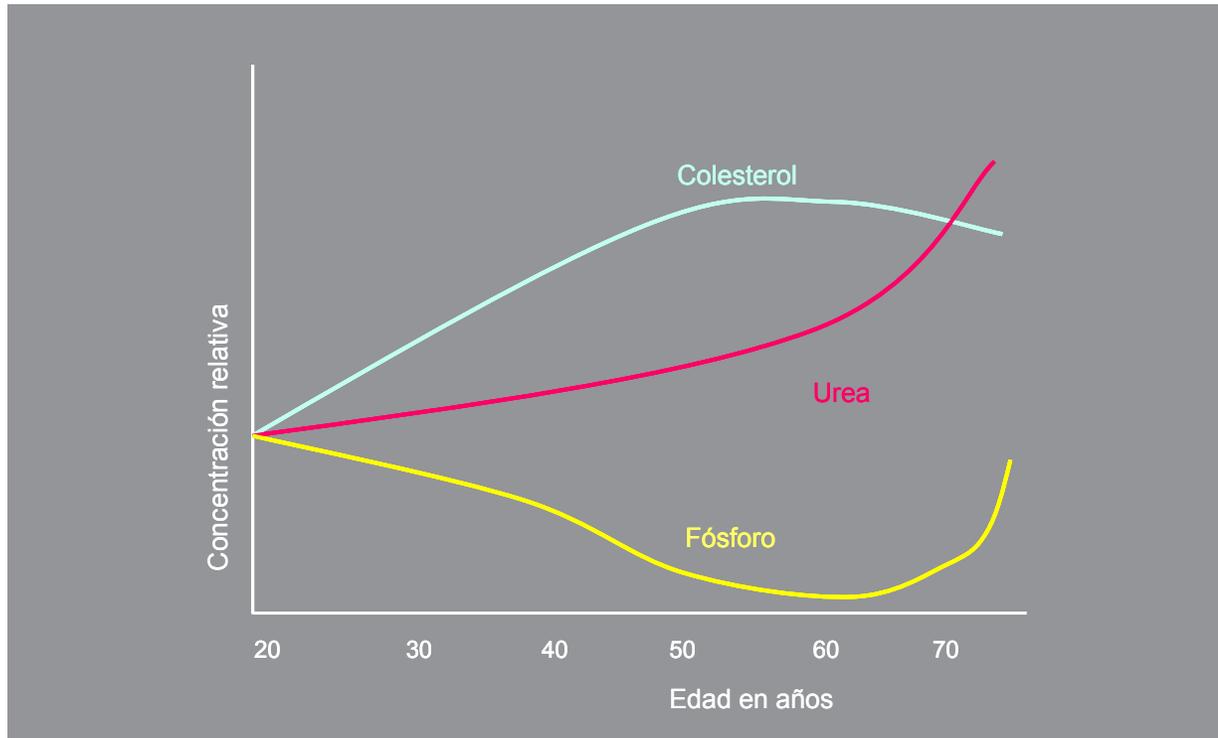
Prestación mínima

$$CV_a < \frac{3}{4} CV_{Bw}$$

## Fuentes de variación biológica

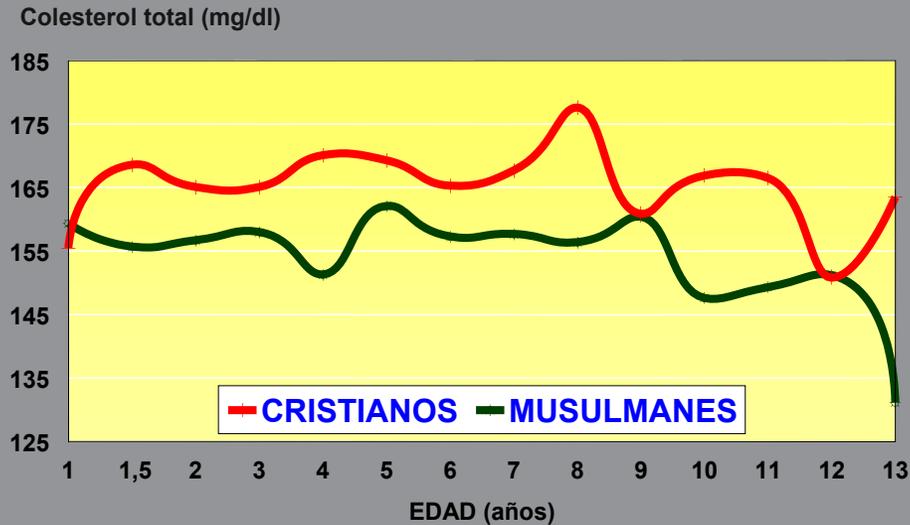
- Variación biológica NO aleatoria
  - A lo largo de la vida
  - Ritmo biológico diario
  - Ciclos mensuales
  - Ritmos estacionales
- Variación biológica aleatoria
  - Variación biológica intra-individual ( $CV_{BW}$ )
  - Variación biológica inter-individual ( $CV_{Bb}$ )

## Fuentes de variación biológica

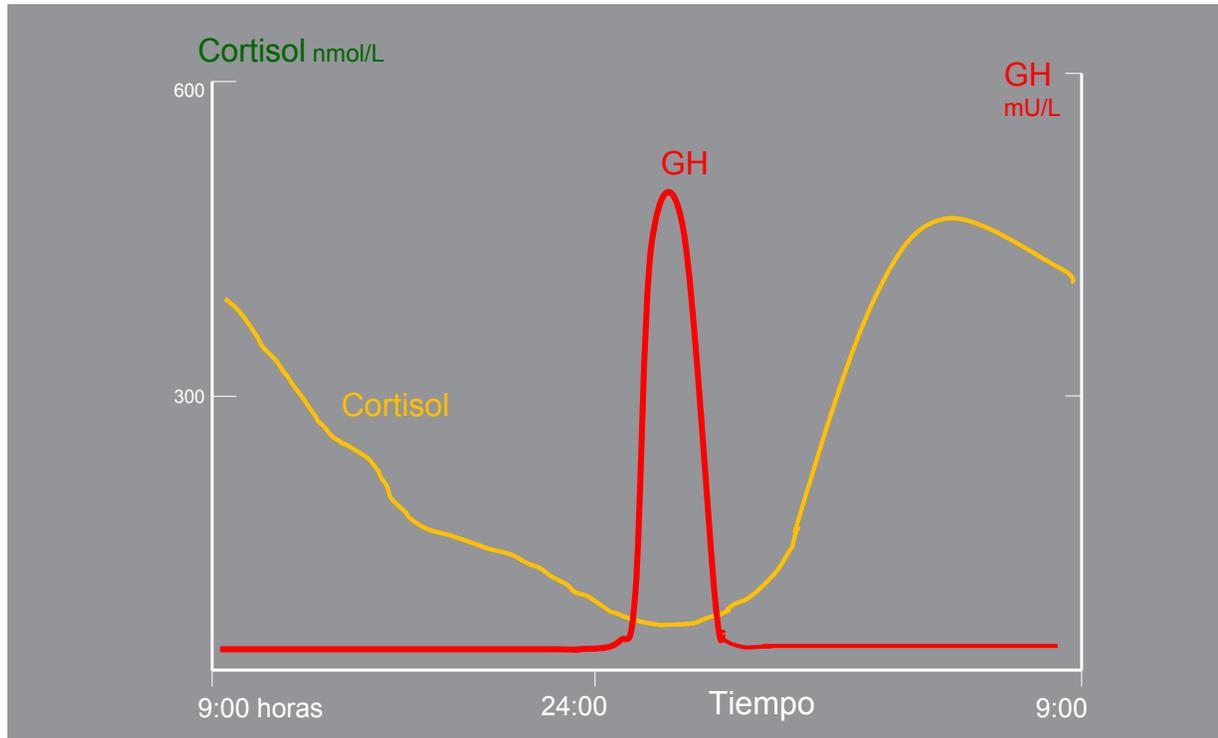


## Variación de los niveles de Colesterol total

### Distribución por grupos socioculturales



## Ritmo biológico circadiano



## Cálculo de la variación total ( $CV_t$ )

Cuando se obtiene una sola muestra y esta es analizada una sola vez....

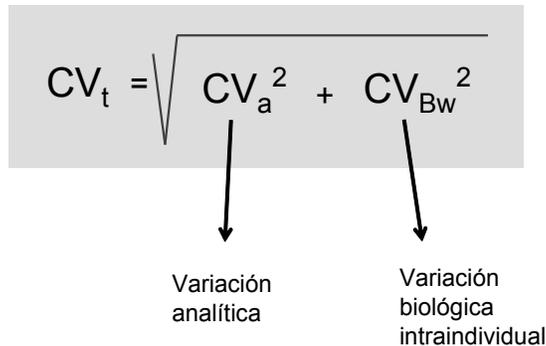
$$CV_t = \sqrt{CV_p^2 + CV_a^2 + CV_{Bw}^2}$$


Variación preanalítica

Variación analítica

Variación biológica intraindividual

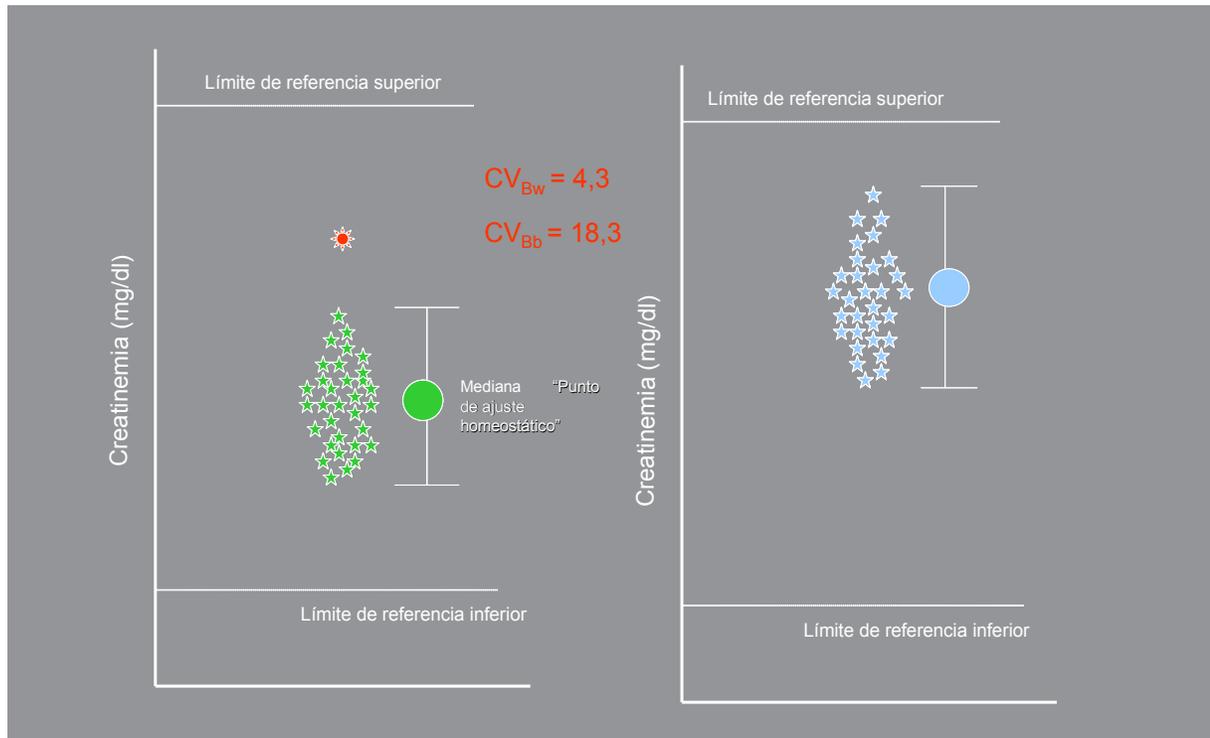
Si minimizamos la variabilidad preanalítica mediante protocolización, entrenamiento, buenas prácticas de laboratorio y seguimiento de los procedimientos operativos descritos y estandarizados....

$$CV_t = \sqrt{CV_a^2 + CV_{Bw}^2}$$


Variación analítica

Variación biológica intraindividual

## Variabilidad biológica aleatoria



$$\text{Índice de individualidad (II)} = \frac{CV_{Bw}}{CV_{Bb}}$$

$$II > 0,6$$

$$CV_{Bw} > CV_{Bb}$$

Escasa individualidad

Diagnóstico

Intervalos de referencia poblacionales

$$II < 0,6$$

$$CV_{Bw} < CV_{Bb}$$

Fuerte individualidad

Seguimiento

Diferencia crítica

$$DC = 2^{1/2} * Z [CV_{Bw}^2 + CV_a^2]^{1/2}$$

Cambio significativo con probabilidad del 95% (Z=1.96)

$$DC_{95} = 2,77 [CV_{Bw}^2 + CV_a^2]^{1/2}$$

Cambio significativo con probabilidad del 99% (Z=2.58)

$$DC_{99} = 3,65 [CV_{Bw}^2 + CV_a^2]^{1/2}$$

Varón de 63 años con hipertensión y una concentración de colesterol total de 254 mg/dl, la cual tras modificar sus hábitos de vida (tratamiento dietético) cambió a 224 mg/dl.

¿Cuánto ha cambiado la colesterolemia?

$$\text{Variación porcentual} = \frac{\text{CT final} - \text{CT inicial}}{\text{CT inicial}} = -11,8\%$$

¿Ha cambiado significativamente la colesterolemia?

# Interpretación del cambio aplicando criterios basados en la variabilidad biológica

Variación biológica: de la teoría a la práctica

Callum G. Fraser

Ninewells Hospital and medical School. Dundee. Scotland. 2001.

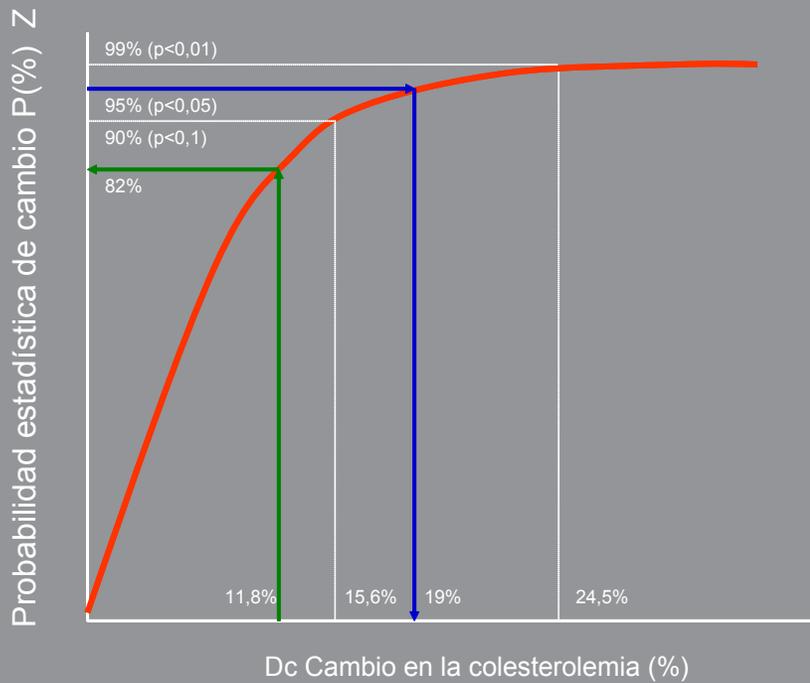
¿Ha cambiado significativamente la colesterolemia?

Cambio significativo con probabilidad del 95% ( $Z=1.96$ )( $p<0,05$ )

$$DC_{95} = 2,77 [CV_{Bw}^2 + CV_a^2]^{1/2} \quad DC = 19\%$$

Cambio significativo con probabilidad del 99% ( $Z=2.58$ )( $p<0,01$ )

$$DC_{99} = 3,65 [CV_{Bw}^2 + CV_a^2]^{1/2} \quad DC = 24,5\%$$



# Comparación de métodos

## Diagnósticos cuantitativos

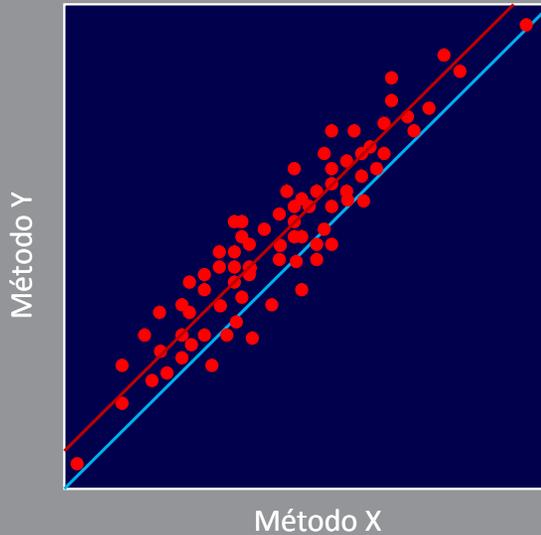
1. Métodos analíticos
2. Métodos de regresión
3. Métodos gráficos

# Comparación de métodos

## Diagnósticos cuantitativos

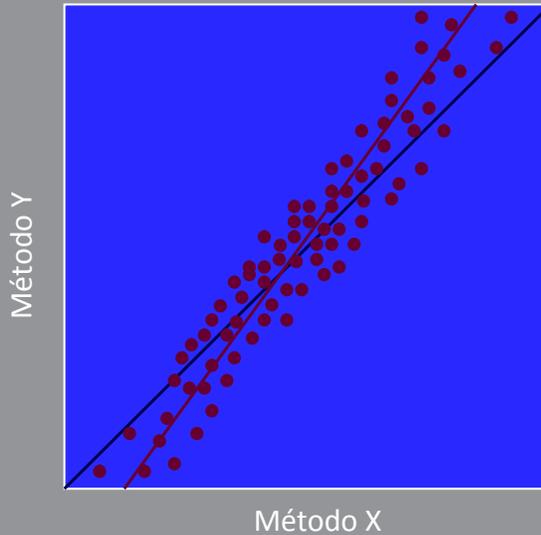
- Se utiliza para demostrar que el nuevo método (Y) presenta una buena concordancia con el método clásico de referencia (X).
- Demostrar que una misma medida tiene equivalencia con diferentes equipos.
- Detectar la presencia de SESGOS (errores sistemáticos):
  - Error (sesgo) constante.
  - Error (sesgo) proporcional.

## Error (sesgo) constante



La diferencia entre ambos métodos se mantiene constante a lo largo del rango de medidas

## Error (sesgo) proporcional



La diferencia entre ambos métodos se lineal a lo largo de todo el largo del rango de medición.

## Comparación de métodos diagnósticos cuantitativos

- Si los errores sistemáticos detectados y medidos están en un límite aceptable, la evaluación se debe completar con un experimento de comparación de métodos frente a los resultados obtenidos con un método de referencia.
- La utilización de la estadística en los experimentos de comparación de métodos es la parte más difícil del análisis estadístico en los estudios de evaluación de métodos.
- ¿Qué método estadístico es el que comete menos error para calcular el error que comete un método analítico?

# Análisis de concordancia: variables cuantitativas

## Métodos diagnósticos

- Coeficiente de correlación intraclass (ANOVA).
  - Modelo de efectos ALEATORIOS.
  - Modelo de efectos FIJOS.
- Coeficiente de correlación-concordancia de Lin.
- Concordancia de Krippendorff.

# Análisis de concordancia: variables cuantitativas

## Métodos de regresión

- Método de regresión lineal por mínimos cuadrados.
- Método de regresión ortogonal de Deming.
- Modelo de Passing-Bablok.
- Método de mínimos cuadrados ponderado.
- Método de regresión de Bland y Altman.
- ANOVA de medidas repetidas.
- Modelos de ecuaciones estructurales.

# Análisis de concordancia: variables cuantitativas

## Métodos gráficos

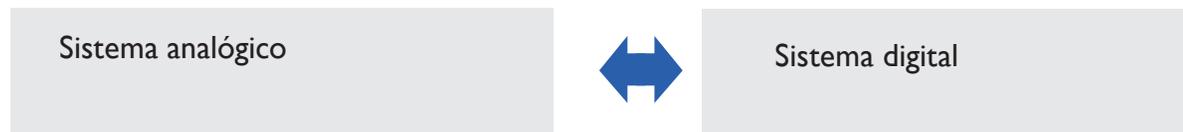
- Análisis de las diferencias individuales método de Bland y Altman (que calcula un intervalo de concordancia),
- Gráfico de distribución acumulativa empírica –“gráfico montaña” (mountain plot)–
- Gráfico de la concordancia- supervivencia (survival-agreement plot).

## ¿Cuál es el método estadístico más adecuado para calcular el error que comete un método diagnóstico cuantitativo?

La utilización de la estadística en los experimentos de comparación de métodos cuantitativos es la parte más difícil del análisis estadístico en los estudios de evaluación de métodos diagnósticos.

## Pregunta de investigación

¿Podemos sustituir en instrumento clásico analógico por el nuevo sistema digital para la medida de la tensión arterial?



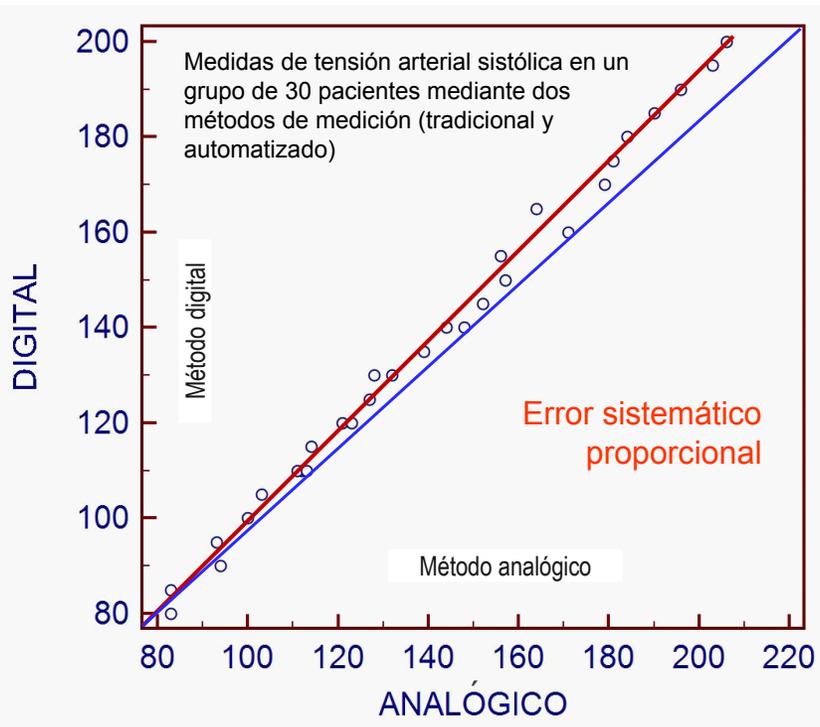
¿Cuál es el test estadístico más acertado?

# Regresión lineal por mínimos cuadrados:

## Condiciones de aplicación(\*)

1. Linealidad de datos.
2. Pequeños errores aleatorios.
3. Intervalo de valores normales suficientemente amplio de concentraciones.
4. Ausencia de valores aberrantes («outliers»).
5. Que la variable PREDICTORA «X» no tenga error, es decir que el método de referencia utilizado sea exacto, pero en la mayoría de ocasiones no es posible, por lo que los resultados del método de comparación (X) también presenta errores.

(\*) Westgard JO, Hunt MR. Use and interpretation of common statistical test in method-comparison studies. Clin Chem 1973; 19: 44-57.



Se representan los datos de un estudio en el que se tomó la tensión arterial sistólica en 30 pacientes utilizando dos métodos diferentes. Se representan gráficamente estos datos, indicando el coeficiente de correlación lineal de Pearson  $r=0,997$  ( $p<0,0001$ ) una asociación prácticamente lineal.

Para variables cuantitativas, es frecuente que el análisis de la concordancia entre dos medidas de la misma variable se realice ERRÓNEAMENTE mediante el cálculo del coeficiente de correlación de lineal ( $r$ ) de Pearson como índice de concordancia.

Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1: 307-310.

## Regresión lineal por mínimos cuadrados:

Variable Y	DIGITAL
Variable X	ANALÓGICO
Tamaño de muestra	30
Coefficiente de determinación R <sup>2</sup>	0,9940
Ecuación de regresión	$Y = 5,8672 + 0,9351 X$

Parámetro	Coefficiente	Error estándar	IC95%	t	p
$\beta_0$	5,8672	1,9830	1,8053-9,9291	2,9588	0,0062
$\beta$	0,9351	0,01369	0,9071-0,9632	66,295	< 0,0001

## Coeficiente de correlación de Pearson

- Mide la fuerza de asociación entre ambas variables (X-Y), es decir la intensidad de la relación lineal entre los dos métodos diagnósticos cuantitativos.
- No ofrece información sobre el grado de CONCORDANCIA, por tanto no puede detectar sesgo constante o proporcional.
- Es posible encontrar un alto coeficiente de correlación con una baja intensidad de concordancia entre métodos.

CORRELACIÓN  $\neq$  CONCORDANCIA

## Mito I: “La sonrisa de la esfinge”

¿Por qué has elegido el coeficiente de correlación de Pearson? ¿Se cumplen todos los requisitos para poder utilizarlo? ¿Te permite ese test decir que los resultados son intercambiables entre ambos instrumentos de medida?

El coeficiente de correlación de lineal ( $r$ ) de Pearson NO resulta una medida adecuada del grado de acuerdo o concordancia entre dos mediciones de un mismo parámetro (presión arterial), ya que si dos instrumentos miden sistemáticamente cantidades diferentes uno del otro, la correlación puede ser perfecta ( $r \approx 1$ ), a pesar de que la concordancia sea nula.

El coeficiente de correlación  $r$  de Pearson decía que la magnitud de la asociación de la medida de la presión arterial entre el método manual A y el método digital B es del 99,7 ( $r=0,997$ ) de la máxima posible.

## Coeficiente de correlación intraclassa

- Los valores del CCI pueden oscilar entre 0 y 1, de modo que la máxima concordancia posible corresponde a un valor de  $CCI=1$ .
- Cuando el valor  $CCI=0$  se obtiene cuando la concordancia observada es igual a la que se esperaría que ocurriera sólo por AZAR. En este caso la variabilidad observada se explicaría por las discrepancias entre los métodos de medición o de los diferentes criterios utilizados por los observadores en el experimento.

CORRELACIÓN  $\neq$  CONCORDANCIA

Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;8:307-10.

Pita Fernández S, Pértegas Díaz S. La fiabilidad de las mediciones clínicas: el análisis de concordancia para variables numéricas. Cad Aten Primaria 2004.

## Coeficiente de correlación intraclase

- La obtención del CCI que permite evaluar la concordancia general entre dos o más métodos u observaciones diferentes se basa en un modelo de análisis de la variancia (ANOVA) con medidas repetidas, pero su obtención directa no es posible con los paquetes estadísticos habituales.
- La variabilidad total observada en el CCI puede dividirse en tres componentes:
  - La variabilidad debida a las diferencias entre los pacientes ( $\sigma^2_P$ );
  - La variabilidad debida a las diferencias entre los observadores (médicos A y B) ( $\sigma^2_O$ ), y
  - La variabilidad (residual), inexplicable (aleatoria), asociada al error inherente a toda medición ( $\sigma^2_R$ ).
- El CCI se define como la proporción de variabilidad total debida a la variabilidad de los pacientes:

$$CCI = \frac{\sigma^2_P}{\sigma^2_P + \sigma^2_O + \sigma^2_R}$$

## Coeficiente de correlación intraclase

- Los valores del CCI pueden oscilar entre 0 y 1: donde el 0 indica ausencia de concordancia y el 1, la concordancia o fiabilidad absoluta de los resultados obtenidos.
  - Si bien el valor del CCI que define una fiabilidad satisfactoria es arbitrario (criterios de Fleiss):
    - CCI  $< 0,4$  representan BAJA fiabilidad.
    - CCI entre  $0,4 - 0,75$  representan una fiabilidad entre REGULAR y BUENA.
    - CCI  $> 0,75$  representan una fiabilidad EXCELENTE.
- Fleiss JL. The design and analysis of clinical experiments. Nueva York: John Wiley & Sons, Inc., 1986.*
- Los resultados ofrecidos por la  $r$  de Pearson para estos propósitos pueden resultar equívocos.
  - Se recomienda el uso del CCI para cuantificar la fiabilidad de las mediciones clínicas.

## Pregunta de investigación

¿Podemos sustituir en instrumento clásico analógico por el nuevo sistema digital para la medida de la tensión arterial?



¿Cuál es el test estadístico más acertado?

## Coeficiente de correlación intraclassa

### Ejemplo

$$CCI = \frac{k \cdot SS_{ENTRE} - SS_{TOTAL}}{(k-1) \cdot SS_{TOTAL}} = \frac{2 \times 73597,683 - 73940,183}{73940,183} = 0,991$$

- En este caso  $CCI=0.991$ , casi toda la variabilidad observada no depende de las diferencias en los métodos de medición empleados o los diferentes observadores (tan sólo el 0,9%).

- Es decir, sólo el 0,9% de la variabilidad observada es debida a las discrepancias o diferencias entre los métodos de medición o por el diferente criterios utilizados por los observadores.

Sistema analógico



Sistema digital

Pita Fernández S, Pértegas Díaz S. La fiabilidad de las mediciones clínicas: el análisis de concordancia para variables numéricas. Cad Aten Primaria 2004.

## Coeficiente de correlación intraclass

### Cálculo del CCI mediante programas estadísticos

Parámetro	Correlación intraclass	IC95%
Medidas individuales	0,9911	0,9398 a 0,9971
Medidas promedio	0,9955	0,9690 a 0,9986

## Valoración del Coeficiente de correlación intraclassa

Valoración de la concordancia según los valores de Coeficiente de Correlación Intraclassa (CCI)

CCI	Fuerza de la concordancia
> 0,90	Muy buena 
0,71 - 0,90	Buena
0,51 - 0,70	Moderada
0,31 - 0,50	Mediocre
< 0,30	Mala o nula

## Modelo de efectos aleatorios

Este modelo aleatorio es apropiado cuando los observadores implicados en la medición, las enfermeras A y B en el ejemplo, representan una «muestra» aleatoria de la población de posibles observadores (p. ej, otras enfermeras del hospital) que en el futuro harán uso del instrumento evaluado (el esfigmomanómetro).

Luis Prieto, Rosa Lamarca, Alfonso Casado. La evaluación de la fiabilidad en las observaciones clínicas: el coeficiente de correlación intraclase. Med Clin (Barc). 1998;110:142-5.

Se desea evaluar la concordancia entre las medidas de tensión arterial entre dos enfermeras A y B mediante la comparación de los resultados de dos mediciones consecutivas de la presión arterial sistólica en 10 pacientes, realizadas en el mismo esfigmomanómetro por dos enfermeras.

ENFERMERA A

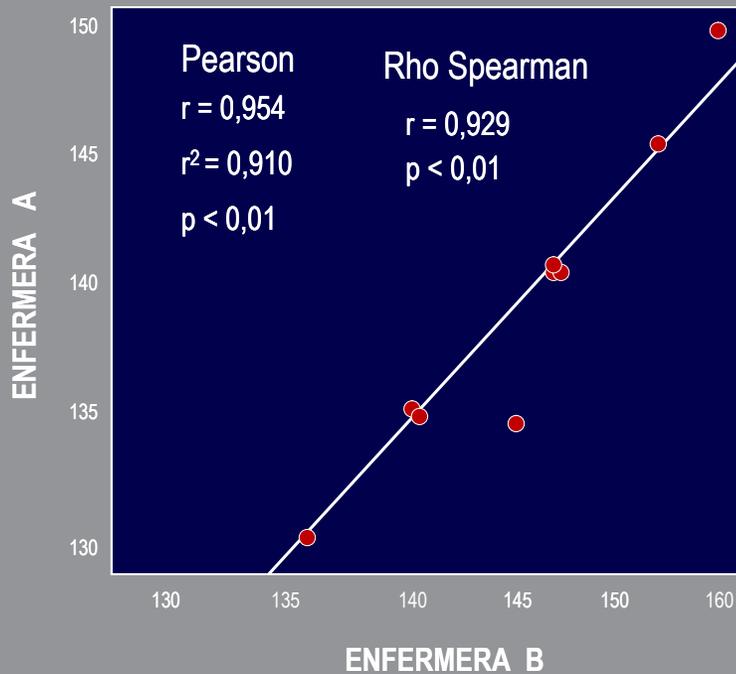


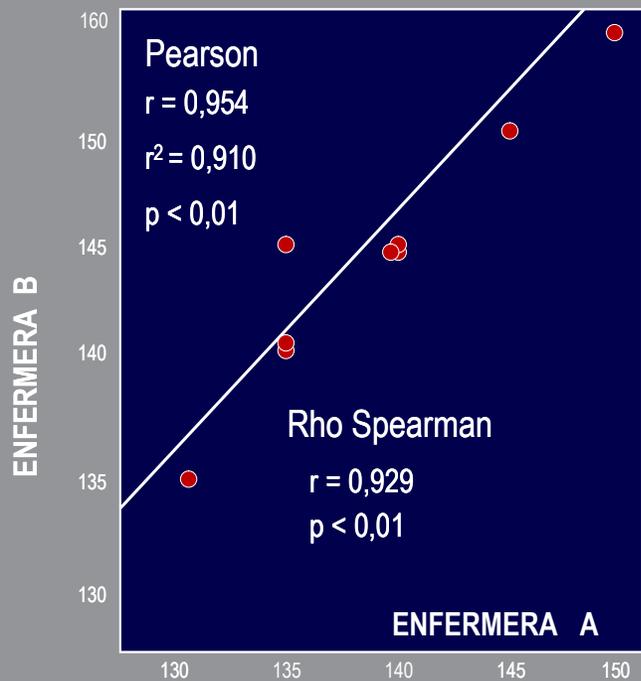
ENFERMERA B

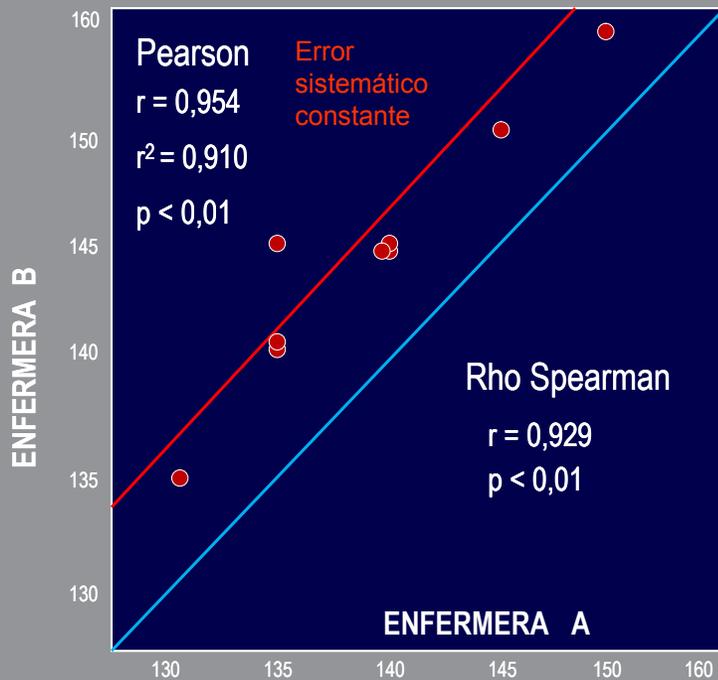
¿Cuál es el test estadístico más acertado?

Comparación de dos mediciones consecutivas de la presión arterial sistólica en 10 pacientes, realizadas en el mismo esfigmomanómetro por dos enfermeros diferentes.

Paciente	Enfermera A	Enfermera B	Diferencia (B-A)
1	135	140	5
2	140	145	5
3	130	135	5
4	145	150	5
5	140	145	5
6	150	160	10
7	140	145	5
8	135	140	5
9	140	145	5
10	135	145	10
Media	139	145	6
DE	5,68	6,67	2,11







## T de Student para datos apareados:

1. Mide la diferencia entre las medias de las mediciones realizadas por los dos métodos comparados.
2. Sólo puede detectar el sesgo o error sistemático constante.
3. No puede detectar el error sistemático PROPORCIONAL.

## Resultados

Variables	Media	N	Desviación estándar (s)	Error estándar de la media (EEM)
ENFERMERA A	139,0	10	5,678	1,795
ENFERMERA B	145,0	10	6,667	2,108

Variable	N	Correlación (r)	Significación
ENFERMERA A versus ENFERMERA B	10	0,954	0,000

Variables	Diferencias relacionadas			Intervalo de confianza de la diferencia		$t_{\text{calculada}}$	gl	Significación bilateral
	Media	s	EEM	Inferior	Superior			
Enfermera B - Enfermera A	6,00	2,108	0,667	4,492	7,508	9,000	9	0,000

$t_{\text{calculada}} > t_{\text{tabulada}} \rightarrow p < 0,05$  (rechazamos la hipótesis nula)

La  $t$  de Student para datos apareados permite detectar el sesgo o error sistemático CONSTANTE.

$$CCI = \frac{\sigma^2_P}{\sigma^2_P + \sigma^2_O + \sigma^2_R}$$

Fórmula simplificada del coeficiente de correlación intraclase (CCI)  
para un MODELO de EFECTOS ALEATORIOS

$$CCI = \frac{(DE_A)^2 + (DE_B)^2 - (DE_{B-A})^2}{(DE_A)^2 + (DE_B)^2 + (X_{B-A})^2 - (DE_{B-A})^2 / n}$$

Cálculo del CCI del ejemplo anterior:

$$CCI = \frac{(5,68)^2 + (6,67)^2 - (2,11)^2}{(5,68)^2 + (6,67)^2 + (6)^2 - (2,11)^2 / 10} = 0,64$$

## Coeficiente de correlación intraclass

### Cálculo del CCI mediante programas estadísticos

Parámetro	Correlación intraclass	IC95%
Medidas individuales	0,6436	- 0,0488 a 0,9240
Medidas promedio	0,7831	- 0,1027 a 0,9695

## Valoración del Coeficiente de correlación intraclassa

Valoración de la concordancia según los valores de Coeficiente de Correlación Intraclassa (CCI)

CCI	Fuerza de la concordancia
> 0,90	Muy buena
0,71 - 0,90	Buena
0,51 - 0,70	Moderada 
0,31 - 0,50	Mediocre
< 0,30	Mala o nula

## Conclusiones

- El valor del CCI obtenido es claramente inferior al valor 0,954 obtenido con la r de Pearson, demostrando que la existencia de una alta asociación entre dos variables no es necesariamente signo de una alta concordancia.
- Aplicando la fórmula a los datos anteriores, se obtiene un CCI igual a 0,64, lo que indica una concordancia moderada entre los resultados obtenidos por las enfermeras A y B.

CORRELACIÓN  $\neq$  CONCORDANCIA

## Modelo de efectos fijos

- Este modelo es apropiado cuando los observadores que intervienen en el estudio son los únicos que participarán en el mismo.
- Se utiliza cuando el objetivo es conocer la magnitud de las discrepancias entre dos médicos y determinar si el médico A puede ser sustituido por el médico B en esa tarea.
- Sólo estamos interesados en la comparación de los resultados obtenidos por estos 2 médicos y no desearemos generalizar los resultados al conjunto de médicos que miden la frecuencia cardíaca en el hospital.

Se desea valorar la concordancia de las mediciones de la FRECUENCIA CARDÍACA obtenidas en 10 pacientes por 2 médicos diferentes: un médico A, jefe del servicio, y un médico B, en período de formación. El objetivo es conocer la magnitud de las discrepancias entre ambos y determinar si el médico A puede ser sustituido por el médico B en esa tarea.

El pulso radial se siente mejor en la muñeca, debajo del pulgar.

JEFE DE SERVICIO

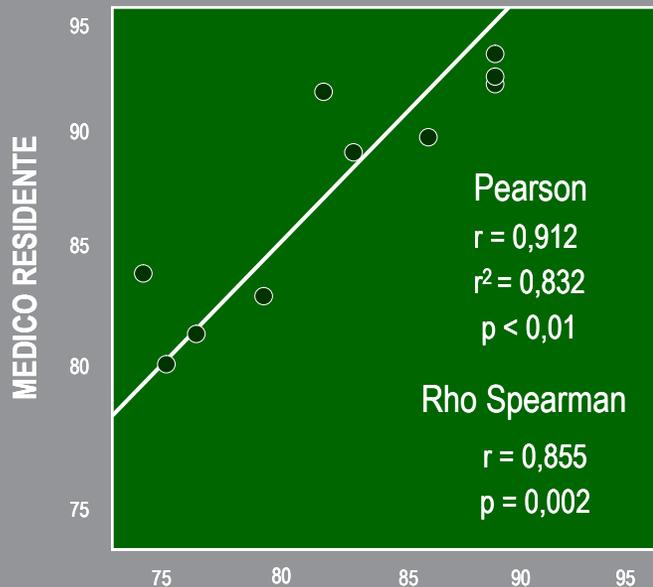


MÉDICO RESIDENTE

¿Cuál es el test estadístico más acertado?

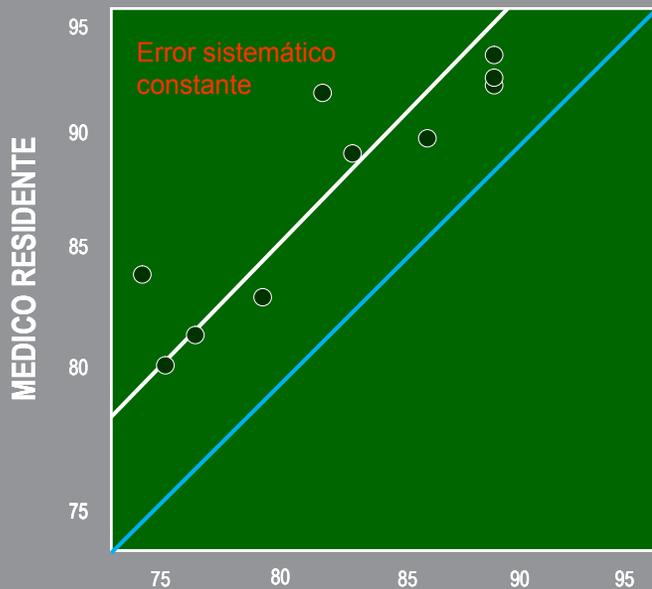
Comparación de dos mediciones consecutivas de la FRECUENCIA CARDÍACA en pulsaciones por minuto (ppm) en 10 pacientes realizada por dos médicos diferentes: un médico A, jefe del servicio, y un médico B en periodo de formación (modelo de EFECTOS FIJOS).

Paciente	Médico A	Medico B	Diferencia (B-A)
1	75	80	5
2	74	84	10
3	76	81	5
4	79	83	4
5	82	92	10
6	83	88	5
7	85	90	5
8	87	92	5
9	87	92	5
10	88	93	5
Media	81,6	87,5	5,9
DE	5,3	5,04	2,18



$$\text{RESIDENTE} = 16,7 + 0,868 * \text{JEFE}$$

JEFE DE SERVICIO



$$\text{RESIDENTE} = 16,7 + 0,868 * \text{JEFE}$$

JEFE DE SERVICIO

$$CCI = \frac{\sigma_P^2}{\sigma_P^2 + \sigma_O^2 + \sigma_R^2}$$

Fórmula simplificada del coeficiente de correlación intraclase (CCI)  
para un MODELO de EFECTOS FIJOS

$$CCI = \frac{(DE_A)^2 + (DE_B)^2 - (DE_{B-A})^2}{(DE_A)^2 + (DE_B)^2 + (X_{B-A})^2 / 2 - (DE_{B-A})^2 / 2n}$$

Cálculo del CCI del ejemplo anterior:

$$CCI = \frac{(5,3)^2 + (5,04)^2 - (2,18)^2}{(5,3)^2 + (5,04)^2 + (5,9)^2 / 2 - (2,18)^2 / 2 \times 10} = 0,69$$

## Valoración del Coeficiente de correlación intraclassa

Valoración de la concordancia según los valores de Coeficiente de Correlación Intraclassa (CCI)

CCI	Fuerza de la concordancia
> 0,90	Muy buena
0,71 - 0,90	Buena
0,51 - 0,70	Moderada 
0,31 - 0,50	Mediocre
< 0,30	Mala o nula

## Conclusiones

A pesar de que la  $r$  de Pearson indica una alta asociación ( $r = 0,912$ ), aplicando la fórmula a los datos anteriores, se obtiene un CCI igual a 0,69, lo que indica una concordancia moderada entre los resultados obtenidos por los médicos A y B.

CORRELACIÓN  $\neq$  CONCORDANCIA

## Limitaciones del coeficiente de correlación intraclase

A pesar de ser la medida de concordancia más adecuada para el caso de variables NUMÉRICAS, el CCI presenta ciertas limitaciones:

1. A la hora de interpretar los valores del CCI, la clasificación de la tabla 5 es subjetiva.
2. Dificultad inherente a su cálculo.
3. Se trata de una prueba PARAMÉTRICA, por lo que limita su uso a:
  - variables distribuidas según una normal.
  - igualdad de varianzas.
  - independencia entre los errores de cada observador.
4. El CCI depende en gran medida de la variabilidad de los valores observados: cuanto más homogénea sea la muestra estudiada, más bajo tenderá a ser el valor del CCI.
5. Lo que más ha limitado la difusión del uso del CCI en la literatura médica es la carencia de interpretación clínica completamente objetiva.

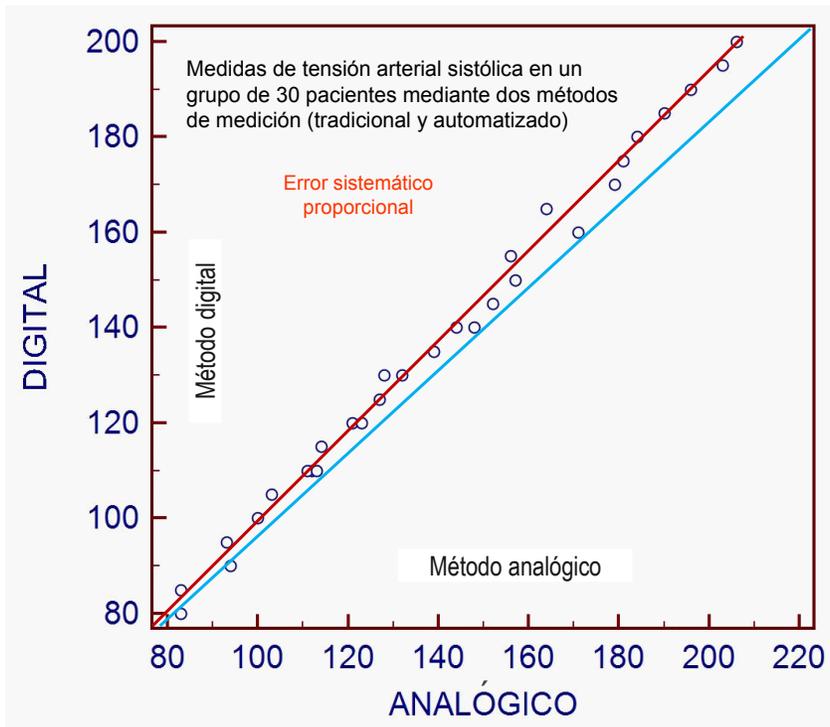
Prieto L, Lamarca R, Casado A. La evaluación de la fiabilidad en las observaciones clínicas: el coeficiente de correlación intraclase. Med Clin (Barc) 1998;110:142-5.

## Pregunta de investigación

¿Podemos sustituir en instrumento clásico analógico por el nuevo sistema digital para la medida de la tensión arterial?



¿Cuál es el test estadístico más acertado?



## Coeficiente de concordancia de Lin

- El coeficiente de correlación de las medidas de concordancia de LIN mide el grado de acuerdo entre dos variables cuantitativas continuas.
- Es útil para evaluar la reproducibilidad o fiabilidad entre dos observadores o evaluadores.

Escala de McBrige para el coeficiente de concordancia de Lin

Valor de la $\sigma$	Fuerza de acuerdo
< 0,90	Pobre
0,90 – 0,95	Moderada
0,95 – 0,99	Sustanciales
> 0,99	Casi perfecto

## Coeficiente de concordancia de Lin

Variable Y	DIGITAL
Variable X	ANALÓGICO
Tamaño de muestra	30
Coeficiente de CONCORDANCIA de Lin	➡ 0,9908
Intervalo de confianza al 95%	0,9832 a 0,9949
Pearson $\sigma$ (precisión)	0,9970
Factor de corrección del sesgo $C_b$ (exactitud)	0,9937

Se desea evaluar la concordancia entre las medidas de tensión arterial entre dos enfermeras A y B mediante la comparación de los resultados de dos mediciones consecutivas de la presión arterial sistólica en 10 pacientes, realizadas en el mismo esfigmomanómetro por dos enfermeras.

ENFERMERA A



ENFERMERA B

¿Cuál es el test estadístico más acertado?

## Coeficiente de concordancia de Lin

Variable Y	Enfermera A
Variable X	Enfermera B
Tamaño de muestra	10
Coeficiente de CONCORDANCIA de Lin	➡ 0,6190
Intervalo de confianza al 95%	0,3074 a 0,8107
Pearson $\sigma$ (precisión)	0,9542
Factor de corrección del sesgo C b (exactitud)	0,6487

## Pregunta de investigación

¿Podemos sustituir en instrumento clásico analógico por el nuevo sistema digital para la medida de la tensión arterial?



¿Cuál es el test estadístico más acertado?

## Método de regresión ortogonal de Deming

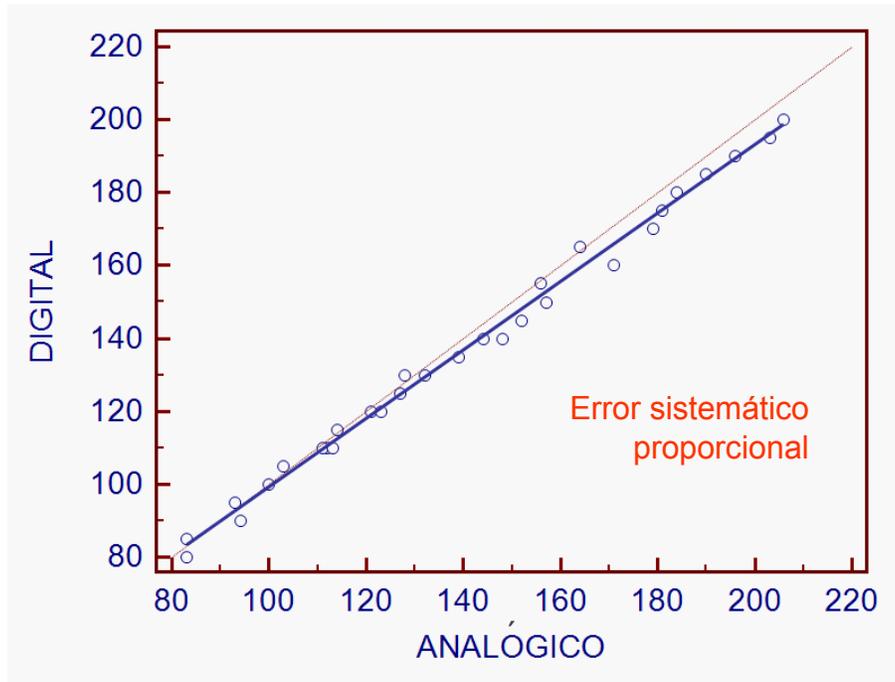
1. En el caso de que presenten errores aleatorios en el método de referencia (X) ha sido propuesto como más aceptable el método de regresión ortogonal establecido por Deming en 1943 (\*).
2. Es adecuado incluso si existen valores aberrantes (\*\*).
3. Si el error que presenta esta variable es proporcional, cuando la varianza de los resultados aumenta con la concentración, ha sido considerado adecuado el cálculo de la correlación lineal mediante el cálculo de mínimos cuadrados ponderado.
4. Mientras que consideran mejor el método de Deming si los errores en «x» son significativos frente al intervalo ensayado, cuando el intervalo de concentraciones es estrecho (v.g. en Potasio) (\*\*\*)).

(\*) Wakkers PJM, Hallendoorn HBA, Op de Weegh GJ, Heerspink W. Applications of statistics in clinical chemistry. A critical evaluation of regression lines. Clin Chim Acta 1975; 64: 173-84.

(\*\*) Cornbleet PJ, Gochman N. Incorrect least-squares regression coefficients in method-comparison analysis. Clin Chem 1979; 25: 432-38.

(\*\*\*) Zaydman L, Lasky FD, Van Nostrand C, Doptis PD. Recommended linear regression models for comparison of method in clinical Chemistry. Clin Chem 1991, 37: 960.

## Representación gráfica de Deming



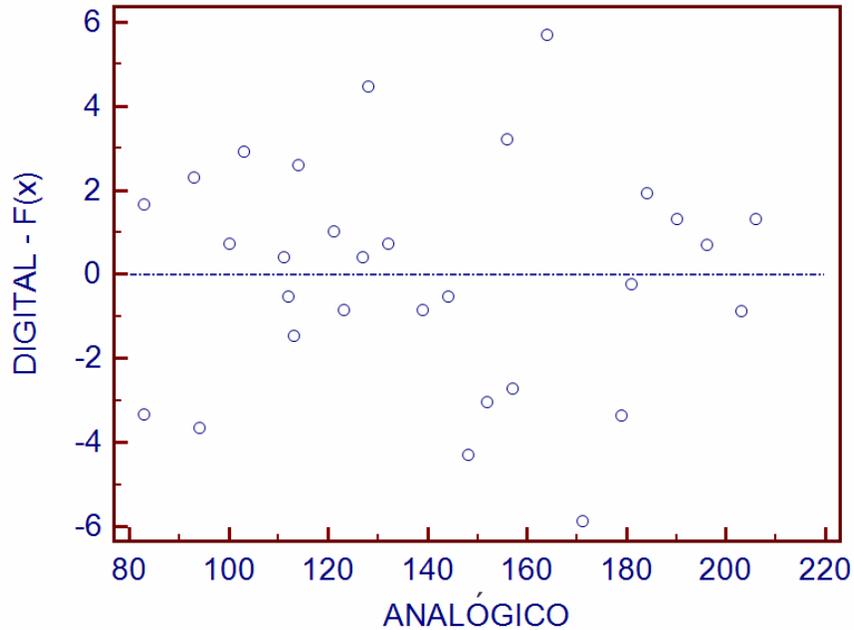
## Método de regresión ortogonal de Deming

Variable Y	DIGITAL
Variable X	ANALÓGICO
Tamaño de muestra	30
Ecuación de regresión	$Y = 5,4741 + 0,9379 X$ 

Método	Media	Coefficiente de variación (CV%)
X	140,23	26,24
Y	137,00	25,19

Parámetro	Coefficiente	Error estándar	IC95%
$\beta_0$	5,4741	1,7780	1,8377-9,1106
$\beta$	0,9379	0,01275	0,9124-0,9634

## Representación gráfica de los residuales de Deming



## Modelos de regresión de Passing-Bablok

- En los años 80-90, para evitar los inconvenientes del método simple de mínimos cuadrados, Passing y Bablok desarrollan el cálculo del método de regresión no paramétrica.
- Fue defendido como «método de elección debido a su robustez» y a permitir «obtener una estimación no sesgada de la recta de regresión».

Frey E, Fuentes-Arderiu X, Queraltó JM. Comparación estadística de métodos analíticos. Educación continuada Química Clínica. 1988; 1: 69-79.

# Modelos de regresión de Passing-Bablok

## Condiciones y limitaciones de aplicación

1. La razón de varianzas en «Y» y «X» iguale el cuadrado de la pendiente.
2. Ausencia de error en la variable PREDICTORA (independiente) «X».
3. Que el error en la variable RESULTADO «Y» debe ser homogéneo en todo el intervalo de concentraciones.

Zaydman L, Lasky FD, Van Nostrand C, Doptis PD. Recommended linear regression models for comparison of method in clinical Chemistry. Clin Chem 1991, 37: 960.

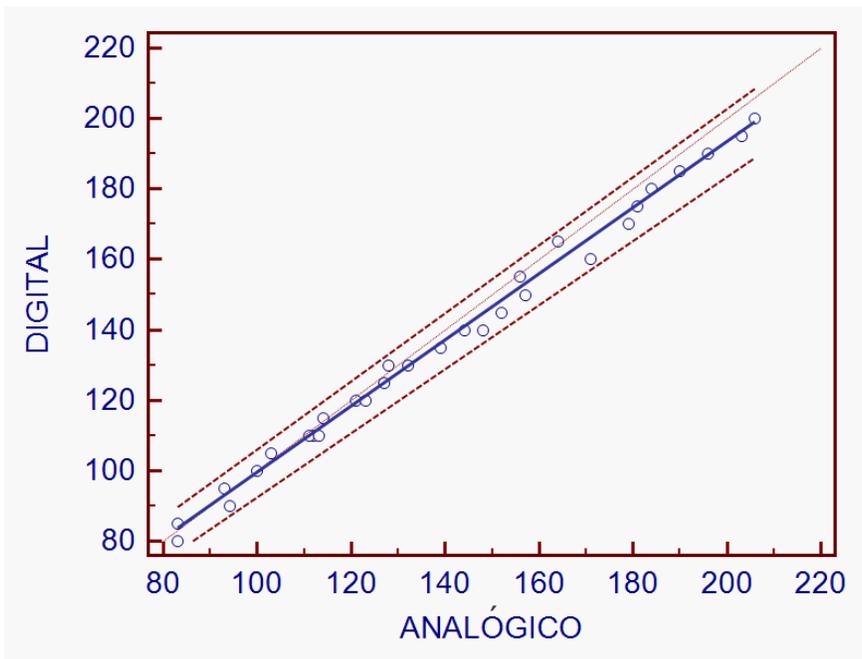
## Modelos de regresión de Passing-Bablok

Variable Y	DIGITAL
Variable X	ANALÓGICO
Tamaño de muestra	30
Ecuación de regresión	$Y = 5,9375 + 0,9375 X$

Parámetro	Coefficiente	IC95%
$\beta_0$	5,9375	1,4912-9,5455
$\beta$	0,9375	0,9091-0,9649

Test de linealidad de suma acumulada  $p < 0,10$  No existe desviación significativa de la linealidad





¿Se puede utilizar el método de Passing-Bablok?

$$\frac{S_y^2}{S_x^2} = [\beta_1]^2$$

$$\frac{1190,69}{1353,56} = 0,8796 = [0,9375]^2 = 0,8789$$

Sí.

## Pregunta de investigación

El objetivo de este estudio ha sido analizar la concordancia entre dos métodos enzimáticos de determinación de la AMILASA sérica perteneciente a una misma casa comercial y basados en el mismo principio físico-químico.

Amilasa antigua



Amilasa nueva

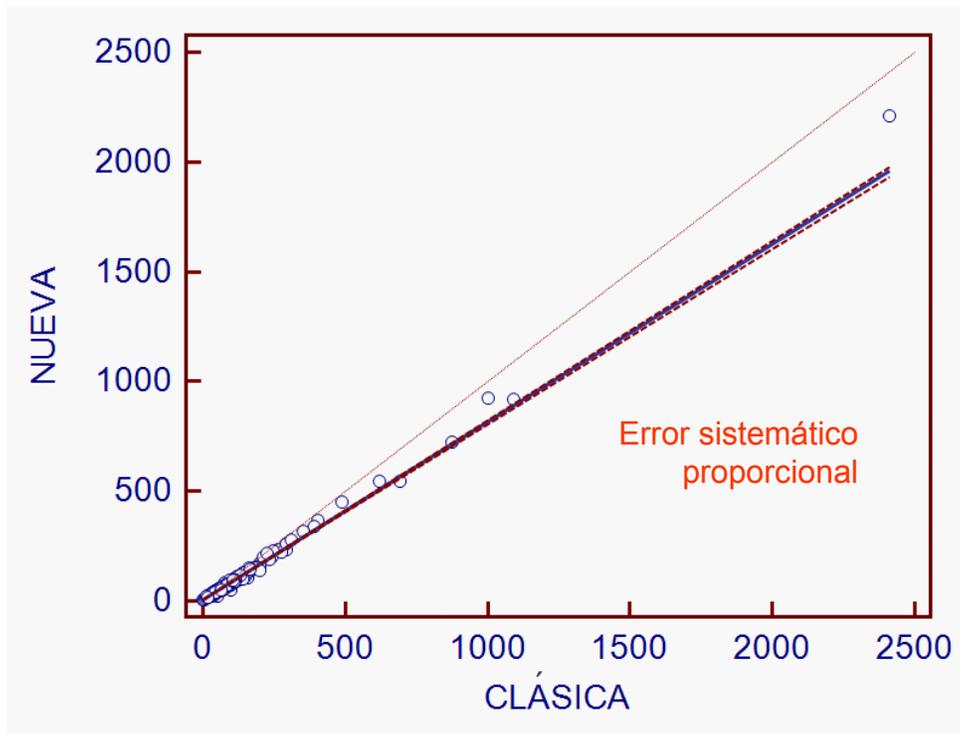
¿Cuál es el test estadístico más acertado?

## Modelos de regresión de Passing-Bablok

Variable Y	AMILASA NUEVA
Variable X	AMILASA CLÁSICA
Tamaño de muestra	548
Ecuación de regresión	$Y = 4,6190 + 0,8095 X$ 

Parámetro	Coefficiente	IC95%
$\beta_0$	4,6190	4,000-5,200
$\beta$	0,8095	0,8000-0,8182

Test de linealidad de suma acumulada  $p < 0,01$  Desviación significativa de la linealidad 



$$CV_a \text{ (nueva)} = 152\%$$

$$CV_a = \frac{SD}{\bar{X}} * 100$$

$$CV_a \text{ (clásica)} = 148\%$$

NO SE PUEDE UTILIZAR el método de Regresión ortogonal de Deming porque los  $CV > 100\%$ .

¿Se puede utilizar el método de Passing-Bablok?

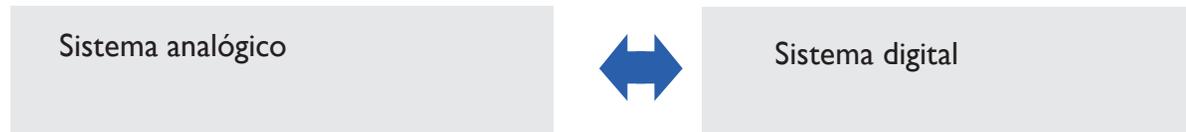
$$\frac{S_y^2}{S_x^2} = [\beta_1]^2$$

$$\frac{14508}{18695} = 0,7060 \quad \neq \quad [0,8095]^2 = 0,6553$$

No.

## Pregunta de investigación

¿Podemos sustituir en instrumento clásico analógico por el nuevo sistema digital para la medida de la tensión arterial?



¿Cuál es el test estadístico más acertado?

Pita Fernández S, Pértegas Díaz S. La fiabilidad de las mediciones clínicas: el análisis de concordancia para variables numéricas. Cad Aten Primaria 2004.

## Método de mínimos cuadrados ponderado

- En 1991 Zaydman et al., en su estudio sobre modelos de regresión lineal para comparación de métodos, rechazan el método de Passing-Bablok porque no cumple las condiciones anteriores.
- Obtienen los mejores resultados con el método de mínimos cuadrados ponderado cuando los errores en la variable PREDICTORA «X» son pequeños frente al intervalo de concentraciones como el medido (v.g. en Glucosa).

## Método de mínimos cuadrados ponderado

Resumen del modelo

R múltiple	→ ,997
R cuadrado	→ ,994
R cuadrado ajustado	→ ,994
Error típico de la estimación	,792
Valor de la función log-verosimilitud	-71,391

ANOVA

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Regresión	2912,687	1	2912,687	4643,382	,000
Residual	17,564	28	,627		
Total	2930,250	29			

Coeficientes

	Coeficientes no estandarizados		Coeficientes estandarizados		t	Sig.
	B	Error típico	Beta	Error típico		
(Constante)	5,845	1,923			3,039	,005
ANALOGICO	→ ,935	→ ,014	,997	,015	68,142	→ ,000

$$\frac{\beta}{EE\beta} = \frac{0,935}{0,014} = 66,76 \gg Z_{95\%} (1,96) \rightarrow p < 0,05$$

$$Z_{99,5\%} (3,29) \rightarrow p < 0,005$$

## Pregunta de investigación

El objetivo de este estudio ha sido analizar la concordancia entre dos métodos enzimáticos de determinación de la AMILASA sérica perteneciente a una misma casa comercial y basados en el mismo principio físico-químico.

Amilasa antigua



Amilasa nueva

¿Cuál es el test estadístico más acertado?

# Método de mínimos cuadrados ponderado

**Resumen del modelo**

R múltiple	,995
R cuadrado	,990
R cuadrado ajustado	,990
Error típico de la estimación	,184
Valor de la función log-verosimilitud	-1593,325

**ANOVA**

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Regresión	1868,764	1	1868,764	55027,383	,000
Residual	18,542	546	,034		
Total	1887,306	547			

**Coefficientes**

	Coefficients no estandarizados		Coefficients estandarizados		t	Sig.
	B	Error típico	Beta	Error típico		
(Constante)	5,060	,152			33,260	,000
CLASICA	,806	,003	,995	,004	234,579	,000

## Método gráfico de Bland y Altman

- Bland y Altman proponen su método de evaluación de las diferencias, con el atractivo de su sencillez de cálculo y su utilización frente al uso de los métodos de regresión anteriores.
- Representan gráficamente la diferencia entre los dos métodos ( $D = Y - X$ ) frente a la media de la suma de los dos resultados ( $A = X + Y / 2$ ).

Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *The Statistician* 1983; 32, 307-317.

López, F. Utilización adecuada de las pruebas estadísticas en los estudios de comparación de procedimientos. *Quim Clin* 2000; 19: 237.

## Pregunta de investigación

¿Podemos sustituir en instrumento clásico de gran tamaño por el sistema mini portátil para la medida del flujo respiratorio máximo (FEM)?

Clásico grande



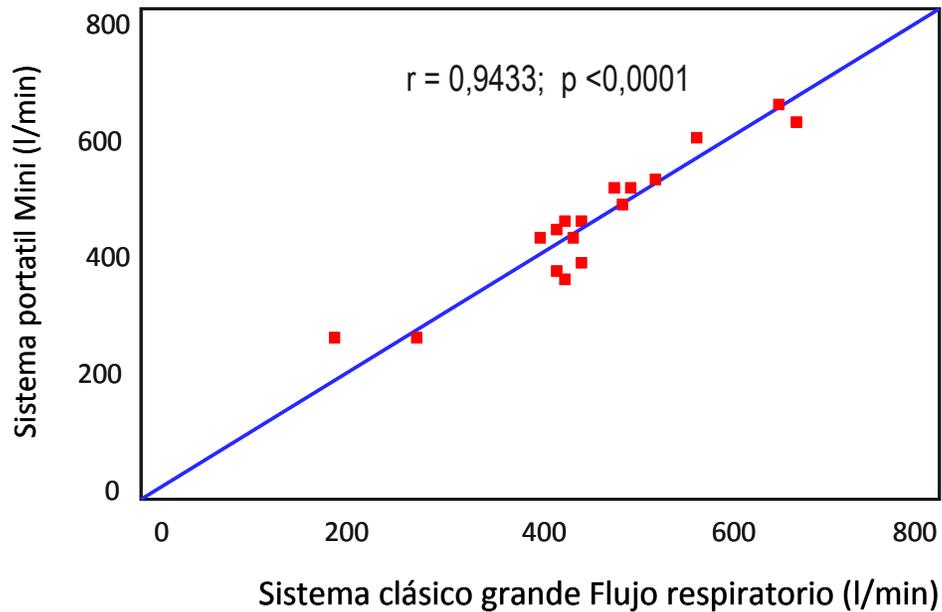
Mini portátil

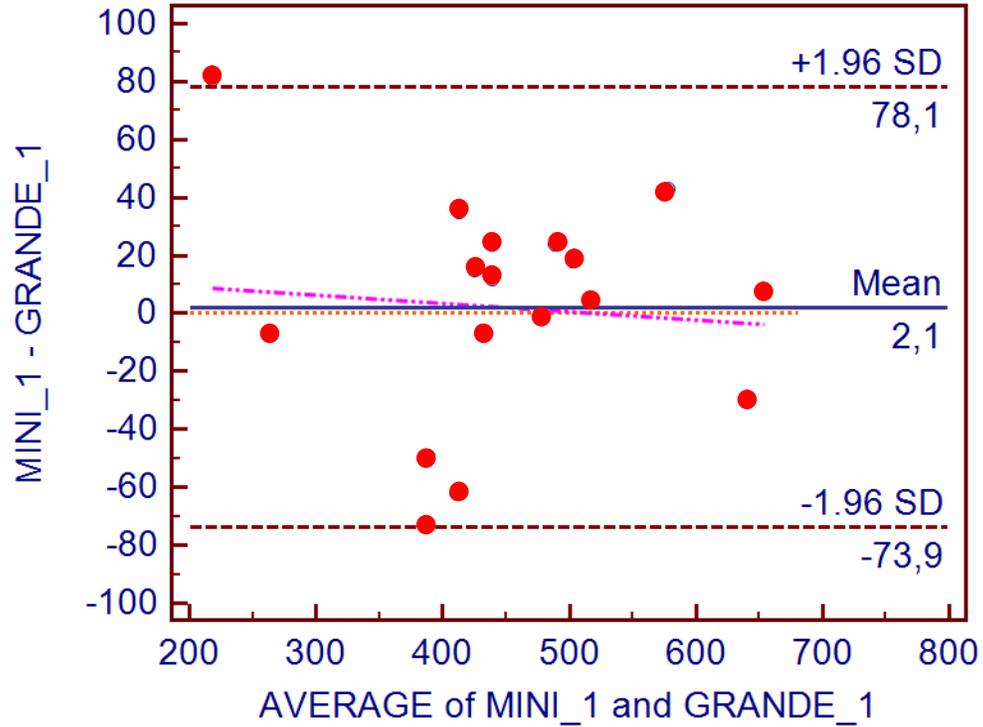
¿Cuál es el test estadístico más acertado?

	GRANDE 1ª Medida	GRANDE 2ª medida	MINI 1ª medida	MINI 2ª medida
1	494	490	512	525
2	395	397	430	415
3	516	512	520	508
4	434	401	428	444
5	476	470	500	500
6	557	611	600	625
7	413	415	364	460
8	442	431	380	390
9	650	638	658	642
10	433	429	445	432
11	417	420	432	420
12	656	633	626	605
13	267	275	260	227
14	478	492	477	467
15	178	165	259	268
16	423	372	350	370
17	427	421	451	443

Se midió el flujo respiratorio máximo (FEM) en una muestra de 17 individuos con un amplia gama de FEM. Se realizaron 2 mediciones en cada instrumento (grande clásico y Mini portátil). Sólo la 1ª medida de cada aparato se utilizó para la comparación de los métodos, la 2ª medición se utilizó en el estudio de la repetibilidad.

Generalmente los puntos de datos se agrupan cerca de la línea y es difícil evaluar las diferencias entre los métodos comparados





## Método gráfico de Bland y Altman: Interpretación

- En el caso de que no haya error sistemático los puntos se distribuirán de forma aleatoria a uno y otro lado de la recta correspondiente a la diferencia 0 entre medidas (línea horizontal naranja).
- La línea azul representa la media de las diferencias, que en nuestro ejemplo corresponde a 2,1 (error sistemático del segundo método respecto al primero).
- Las líneas marrones representan los límites de confianza definidos mediante  $\pm 2$  DE de la media de la diferencia, y se denominan límites de concordancia

Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *The Statistician* 1983; 32, 307-317.

Bland JM, Altman DG . Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986: 307-310.

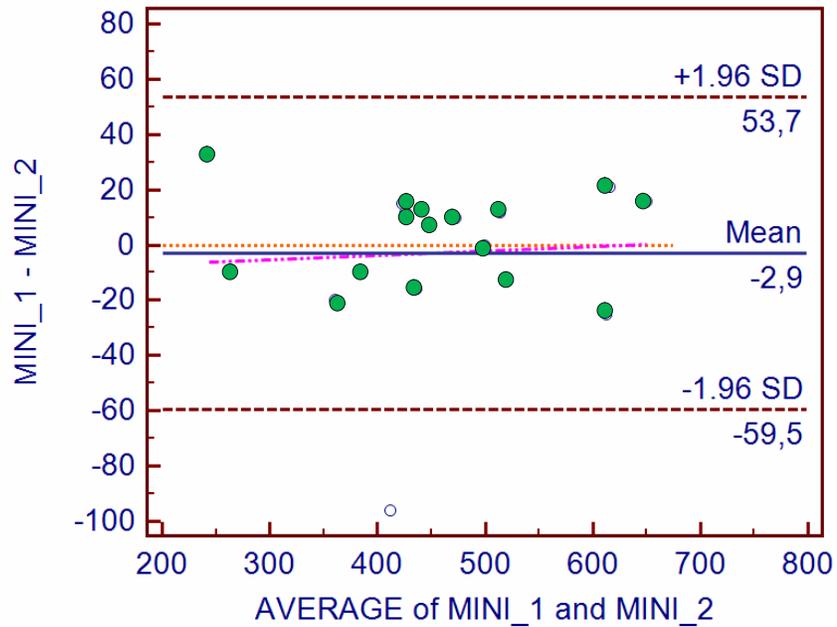
## Reflexiones del gráfico

- Se observa una considerable falta de concordancia entre los medidores grandes y mini, con discrepancias de hasta 80 l/min.
- Siempre que las diferencias  $d \pm 2s$  no sean clínicamente importantes podíamos asumir que los métodos de medición son intercambiables. Estas  $d \pm 2s$  son los “límites de acuerdo”.
- Para los datos de FEM se tiene:
  - $d - 2s = -2,1 - (2 * 38,8) = - 78,1 \text{ l / min}$
  - $d + 2s = -2,1 + (2 * 38,8) = - 73,9 \text{ l / min}$
- El medidor Mini ofrece datos de hasta 80 l / min por debajo o de 76 l / min por encima del medidor grande, lo que es inaceptable para los propósitos clínicos.
- Esta falta de concordancia NO es evidente en la figura de correlación lineal de FEM.

## Ensayo de repetibilidad

- La repetibilidad es relevante para el estudio de comparación de métodos porque la repetibilidad de los métodos de medida limita el grado de concordancia.
- Si un método presenta una repetibilidad pobre - es decir, existe una considerable variación en las mediciones repetidas en la misma muestra - el grado de concordancia entre ambos métodos es pobre también.
- La mejor manera de examinar la repetibilidad de un instrumento es tomar medidas repetidas en una serie de muestras.
- Se calcula la media y la desviación estándar de la diferencias medidas. La diferencia media debe ser cero, ya que se trata del mismo método. Si la diferencia de medias es significativamente diferente de cero, no podemos utilizar los datos para evaluar la repetibilidad, ya sea porque el conocimiento de la 1ª medida está afectando a la 2ª o el de la medida es modificar la cantidad).

## Ensayo de repetibilidad en el Mini



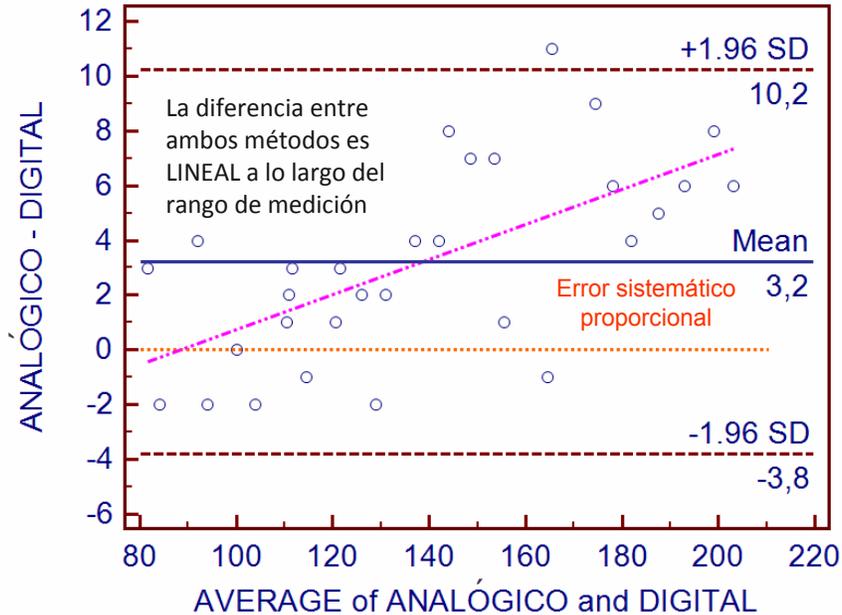
## Pregunta de investigación

¿Podemos sustituir en instrumento clásico analógico por el nuevo sistema digital para la medida de la tensión arterial?



¿Cuál es el test estadístico más acertado?

# Método gráfico de Bland y Altman



## Método gráfico de Bland y Altman: Interpretación

- Existe un importante error sistemático porque los puntos se distribuyen preferentemente por encima de la recta correspondiente a la diferencia 0 entre medidas (línea horizontal naranja).
- El error sistemático es proporcional porque la diferencia entre ambos métodos es LINEAL (línea rosa) a lo largo del rango de medición.
- La línea azul representa la media de las diferencias, que en nuestro ejemplo corresponde a 3,2 (error sistemático del segundo método –digital- respecto al primero –analógico-).
- Las líneas marrones representan los límites de confianza definidos mediante  $\pm 2$  DE de la media de la diferencia (de -3,8 a +10,2 ), y se denominan límites de concordancia.

## Pregunta de investigación

El objetivo de este estudio ha sido analizar la concordancia entre dos métodos enzimáticos de determinación de la AMILASA sérica perteneciente a una misma casa comercial y basados en el mismo principio físico-químico.

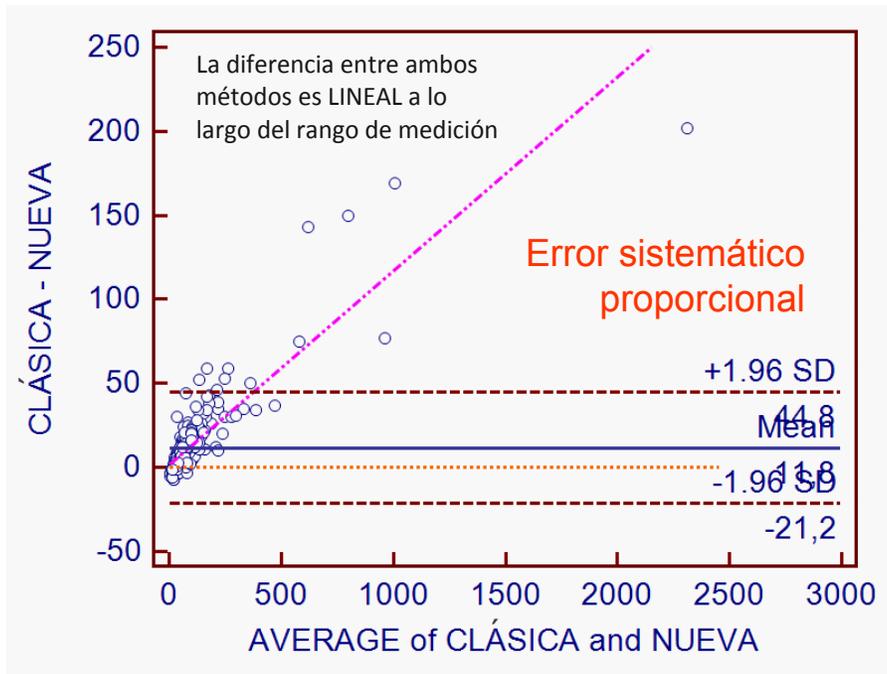
Amilasa antigua



Amilasa nueva

¿Cuál es el test estadístico más acertado?

# Método gráfico de Bland y Altman



## Método gráfico de Bland y Altman: Interpretación

- Existe un error sistemático porque casi todos los puntos se distribuyen preferentemente por encima de la recta correspondiente a la diferencia 0 entre medidas (línea horizontal naranja).
- El error sistemático es proporcional porque la diferencia entre ambos métodos es LINEAL (línea rosa) a lo largo del rango de medición.
- La línea azul representa la media de las diferencias, que en nuestro ejemplo corresponde a 11,8 (y nos mide el error sistemático del segundo método –amilasa nueva- respecto al primero –clásica-).
- Las líneas marrones representan los límites de confianza definidos mediante  $\pm 2$  DE de la media de la diferencia (de -21,2 a +44,8 ), y se denominan límites de concordancia.



# Evaluación de métodos diagnósticos cuantitativos

## Recomendaciones de Zaydman

Zaydman L, Lasky FD, Van Nostrand C, Doptis PD. Recommended linear regression models for comparison of method in clinical Chemistry. Clin Chem 1991, 37: 960.

## Recomendaciones de Zaydman

### Modelos de regresión de Passing-Bablok

No debe utilizarse porque NO cumple la condición:

1. La razón de varianzas en «Y» y «X» iguale el cuadrado de la pendiente.

$$\frac{S_y^2}{S_x^2} = [\beta_1]^2$$

Zaydman L, Lasky FD, Van Nostrand C, Doptis PD. Recommended linear regression models for comparison of method in clinical Chemistry. Clin Chem 1991, 37: 960.

## Recomendaciones de Zaydman

### Método de mínimos cuadrados (método regresión lineal simple)

No debe utilizarse porque NO cumple los siguientes requisitos:

1. Ausencia de error en la variable PREDICTORA (independiente) «X».
2. Que el error en la variable RESULTADO «Y» debe ser homogéneo en todo el intervalo de concentraciones.

Zaydman L, Lasky FD, Van Nostrand C, Doptis PD. Recommended linear regression models for comparison of method in clinical Chemistry. Clin Chem 1991, 37: 960.

## Recomendaciones de Zaydman

### Método de mínimos cuadrados ponderado

Debe utilizarse cuando:

Los errores en la variable predictora (independiente) «x» son pequeños frente al intervalo de concentraciones medido (glucemia).

### Método de regresión ortogonal de Deming

Debe utilizarse cuando:

Cuando los errores en la variable predictora «x» son significativos frente al intervalo ensayado, cuando el intervalo de concentraciones es estrecho (kaliemia).

Zaydman L, Lasky FD, Van Nostrand C, Doptis PD. Recommended linear regression models for comparison of method in clinical Chemistry. Clin Chem 1991, 37: 960.

TÉCNICAS Y PROCEDIMIENTOS

 Localizador web  
Artículo 160,261

## Determinación de óxido nítrico en aire espirado ( $FE_{NO}$ ) mediante un equipo portátil (NIOX-MINO<sup>®</sup> Aerocrine) en población sana

Ana María Fortuna, Teresa Feixas y Pere Casan

Unidad de Función Pulmonar. Departamento de Neumología. Hospital de la Santa Creu i Sant Pau. Facultad de Medicina. Universidad Autónoma de Barcelona. Barcelona. España.

La determinación de óxido nítrico en aire espirado ( $FE_{NO}$ ) es una técnica no invasiva que se utiliza en la evaluación clínica y el tratamiento del asma. El objetivo de este estudio ha sido determinar los valores de referencia de  $FE_{NO}$  en un grupo de voluntarios sanos mediante la utilización de un nuevo equipo portátil (NIOX-MINO<sup>®</sup> Aerocrine), así como determinar el grado de relación que presenta con el equipo de determinación habitual en nuestro laboratorio de función pulmonar (sensor de quimioluminiscencia N-6008<sup>®</sup> SIR). Según los resultados obtenidos, los valores de  $FE_{NO}$  que da el equipo portátil son siempre superiores a los que ofrece el sensor habitual, con un valor de corte de 34 ppb (media + 2 desviación estándar). Existe una relación directa y significativa de la determinación de  $FE_{NO}$  entre ambos equipos ( $r = 0,92$ ;  $p = 0,001$ ) con un factor de corrección de:  $FE_{NO}(\text{NIOX-MINO}^{\text{®}}) = 10 + 1,5 FE_{NO}(\text{N-6008}^{\text{®}})$ . La relación entre los valores de  $FE_{NO}$  y la edad, el sexo, el índice de masa corporal y los valores espirométricos no fue estadísticamente significativa.

**Palabras clave:** Óxido nítrico en aire espirado. Asma. Inflamación.

## Measurement of Fraction of Exhaled Nitric Oxide With the Portable NIOX-MINO Monitor in Healthy Adults

Measurement of the fraction of exhaled nitric oxide ( $FE_{NO}$ ) provides a noninvasive way to monitor asthma treatment in clinical practice. The aim of this study was to determine  $FE_{NO}$  reference values for measurements recorded with the portable NIOX MINO monitor in a group of healthy volunteers. We also assessed the association between values recorded by the portable monitor and the N-6008 chemiluminescence analyzer used in our pulmonary function laboratory. The  $FE_{NO}$  values obtained with the portable monitor were consistently higher than those recorded by the N-6008 analyzer; the cutoff value for the portable monitor was 34 ppb (mean + 2 SD). We detected a direct correlation ( $r=0.92$ ) between the  $FE_{NO}$  measurements recorded by the 2 monitors ( $P=.001$ ). The following equation expresses the relationship between measurements from the 2 devices:  $FE_{NO}(\text{NIOX MINO}) = 10 + [1.5 FE_{NO}(\text{N-6008})]$ . We did not observe statistically significant correlations between  $FE_{NO}$  measurements and age, sex, body mass index, or spirometry.

**Key words:** Fraction of exhaled nitric oxide. Asthma. Inflammation.

## Pregunta de investigación

El objetivo de este estudio ha sido analizar la concordancia entre dos dispositivos de medida del óxido nítrico (NO) exhalado, un Analizador de quimioluminiscencia estacionario (NIOX Aerocrine) analizador ideal, y un analizador portátil (NIOX-MINO), basado en análisis electroquímico.

Equipo estacionario  
N-6008 SIR



Equipo portátil  
NIOX-MINO

¿Cuál es el test estadístico más acertado?

Fortuna AM, Feixas T, Casan P. Determinación de óxido nítrico en aire espirado (FENO) mediante un equipo portátil (NIOX-MINO® Aerocrine) en población sana. Arch Bronconeumol. 2007;43:176-9.

## Análisis estadístico

“Para la descripción de los valores de FENO se facilitan valores medios  $\pm$  desviación estándar. La comparación entre valores medios se llevó a cabo mediante un test de Mann-Whitney. La comparación de los resultados individuales se analizó mediante la correlación de Pearson. En todos los casos la aproximación ha sido bilateral y el nivel de significación el habitual del 5% ( $\alpha = 0,05$ ). El programa informático empleado en el análisis fue el paquete estadístico SPSS versión 11”.

Fortuna AM, Feixas T, Casan P. Determinación de óxido nítrico en aire espirado (FENO) mediante un equipo portátil (NIOX-MINO® Aerocrine) en población sana. Arch Bronconeumol. 2007;43:176-9.

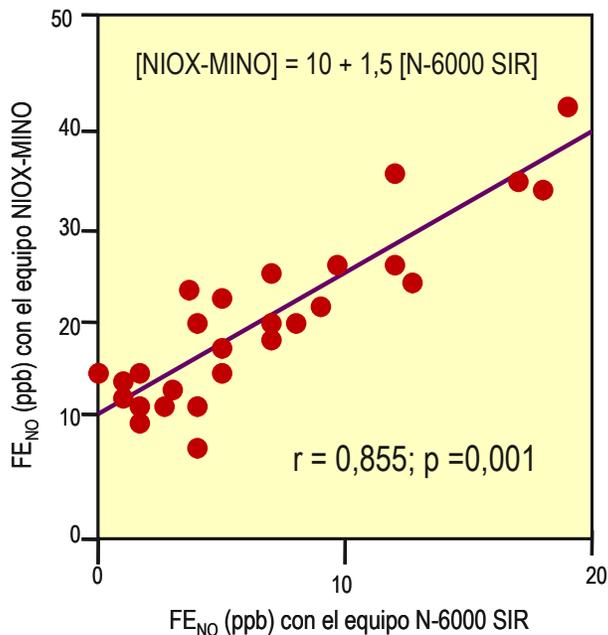


Fig. 3. Correlación de óxido nítrico en aire espirado (FENO) entre la determinación con los equipos N-6008® SIR (sensor de quimioluminiscencia) y NIOX-MINO® (método de electroquímica) en voluntarios sanos. La correlación de Pearson muestra la curva de regresión lineal entre ambos equipos ( $y = \text{FENO [ppb]} \text{ con equipo NIOX-MINO®}$ ;  $x = \text{FENO [ppb]} \text{ con el equipo N-6008® SIR}$ ).

## Resultados

“El valor medio ( $\pm$  desviación estándar) de FENO con el equipo NIOX-MINO® fue de  $20 \pm 7$  ppb (rango: 8-41ppb), y con el sensor N-6008, de  $7 \pm 5$  ppb (rango: 1-19 ppb)”.

“Se observó una correlación directa y estadísticamente significativa de la determinación de FENO entre ambos equipos ( $r = 0,92$ ;  $p = 0,001$ ). Los valores de FENO con el equipo NIOX-MINO® siempre fueron superiores a los obtenidos con el equipo N-6008®. Además, la curva de regresión lineal demostró un factor de corrección de:  $FENO (NIOX-MINO®) = 10 + 1,5 FENO (N-6008®)$  (fig. 3)”.

Fortuna AM, Feixas T, Casan P. Determinación de óxido nítrico en aire espirado (FENO) mediante un equipo portátil (NIOX-MINO® Aerocrine) en población sana. Arch Bronconeumol. 2007;43:176-9.

## Discusión

“Al comparar los 2 equipos y métodos se constata que existe una muy buena correlación de cada determinación de FENO para ambos equipos, con un factor de corrección entre los valores de los 2 equipos de  $\times 1,5 + 10$  para cada valor de FENO obtenido con el equipo N-6008 (figs. 2 y 3), lo que coincide con los valores obtenidos por otros autores”.

Fortuna AM, Feixas T, Casan P. Determinación de óxido nítrico en aire espirado (FENO) mediante un equipo portátil (NIOX-MINO® Aerocrine) en población sana. Arch Bronconeumol. 2007;43:176-9.

## CARTAS AL DIRECTOR



Localizador web

Artículo 217.504

### Metodología estadística en los estudios de comparación de métodos de medida

José-Francisco Pascual-Lledó,  
Carmen Contreras Santos  
y Carlos Martín Carrasco  
Unidad de Neumología.  
Hospital La Inmaculada.  
Huércal-Overa. Almería. España.



**Sr. Director:** Hemos leído con interés el artículo de Fortuna et al<sup>1</sup>, recientemente publicado, sobre la determinación de óxido nítrico en aire espirado (FE<sub>NO</sub>) con el equipo portátil NIOX-MINO® (Aerocrine, Solna, Suecia). Su trabajo se incluye dentro de los denominados estudios de comparación de métodos de medida o estudios de concordancia de métodos, en este caso, de medidas cuantitativas. Uno de los objetivos de este tipo de estudios es evaluar la fiabilidad de las medidas de nuevos equipos de medida<sup>2</sup>. Para ello los autores comparan las medidas tomadas con el equipo NIOX-MINO® Aerocrine con las obtenidas con un equipo estándar, que consideran de referencia, el sensor de quimioluminiscencia N-6008® (SIR, Madrid, España) y llevan a cabo el análisis estadístico mediante una comparación de valores medios, utilizando el test de Mann-Whitney, y mediante el cálculo del coeficiente de correlación de Pearson, para la comparación de resultados individuales. Los autores basan la conclusión de que el equipo es fiable en el hecho de que existe una muy buena correlación de cada determinación de FE<sub>NO</sub> para ambos equipos (r = 0,92; p = 0,001) y dan un intervalo de valores de referencia, con un punto de corte superior que estiman mediante la media más 2 desviaciones estándar.

## Mito I: “La sonrisa de la esfinge”

¿Por qué das un valor de corte, como valor de referencia, a partir de la media más 2 desviaciones estándar?

El cálculo de valores de referencia basado en la media  $\pm 2$  desviaciones estándar sólo puede aplicarse si las variables siguen distribuciones normales.

¿Se cumplen la normalidad de la variable cuantitativa (medida del NO en el aire exhalado) en los dos instrumentos comparados?

No aportas ningún test estadístico (analítico o gráfico) que demuestre que dicha variable sigue la distribución normal. En la tabla I aparecen otros trabajos (5 y 10) se han tenido que realizar transformaciones logarítmicas para normalizar la variable.

¿Seguro que sigue una distribución normal?

## Mito I: “La sonrisa de la esfinge”

¿Por qué has elegido el coeficiente de correlación de Pearson? ¿Se cumplen todos los requisitos para poder utilizarlo? ¿Te permite ese test decir que los resultados son intercambiables entre ambos instrumentos de medida?

El coeficiente de correlación de lineal ( $r$ ) de Pearson NO resulta una medida adecuada del grado de acuerdo o concordancia entre dos mediciones de un mismo parámetro (NO), ya que si dos instrumentos miden sistemáticamente cantidades diferentes uno del otro, la correlación puede ser perfecta ( $r \approx 1$ ), a pesar de que la concordancia sea nula.

El coeficiente de correlación  $r$  de Pearson decía que la magnitud de la asociación de la medida del NO entre el equipo NIOX-MINO® y el equipo N-6008® es del 92% ( $r=0,92$ ) de la máxima posible.

## Mito I: “La sonrisa de la esfinge”

¿Por qué has elegido el coeficiente de correlación de Pearson? ¿Se cumplen todos los requisitos para poder utilizarlo? ¿Te permite ese test decir que los resultados son intercambiables entre ambos instrumentos de medida?

Se ha demostrado que el coeficiente de correlación de Pearson para evaluar la concordancia entre medidas cuantitativas es insuficiente-inadecuado y que la interpretación de dicho coeficiente como una medida de concordancia es errónea. La igualdad de medias tan sólo garantiza que los 2 métodos se centran en el mismo valor, pero en ningún caso que todos sus valores sean iguales.

## Mito I: “La sonrisa de la esfinge”

¿Por qué utilizas el test de Mann-Whitney?

Los autores utilizan para la comparación de medias el test de Mann-Whitney, que es una prueba no paramétrica de comparación de medias para grupos independientes; sin embargo, al tratarse de medidas apareadas, deberían haber utilizado el test no paramétrico para la comparación de medias de medidas apareadas, que es la prueba de Wilcoxon.

## Mito I: “La sonrisa de la esfinge”

Afortunadamente, un buen estudio con un análisis deficiente siempre se puede volver a analizar. Hay paquetes estadísticos (MedCalc®) que incorporan directamente un procedimiento de comparación de métodos que incluyen el cálculo de test que determinan la concordancia entre medidas cuantitativas:

a) métodos analíticos, a través del cálculo de diferentes coeficientes que evalúan la concordancia, como el coeficiente de correlación intraclase o el coeficiente de correlación-concordancia de Lin, modelos de regresión (modelos de regresión de Deming y de Passing-Bablok), ANOVA de medidas repetidas y modelos de ecuaciones estructurales, y

b) métodos gráficos como el método de Bland y Altman (que calcula un intervalo de concordancia), el gráfico de distribución acumulativa empírica –“gráfico montaña” (mountain plot)– o el gráfico de la concordancia- supervivencia (survival-agreement plot).

## Mito I: “La sonrisa de la esfinge”

Como puedes observar en la tabla I he realizado una revisión bibliográfica más exhaustiva que la tuya en la literatura médica.

He encontrado en la página web <http://www.nioxmino.com/references.html> muestra diversos estudios que sí utilizan algunos de estos métodos (tabla I). Si hubieras usado los gráficos de Bland y Altman puede que hubieras encontrado (como los estudios 6 y 7 de la tabla) una relación entre los valores apareados de  $FE_{NO}$  con los 2 equipos de medida de tipo **heterocedástica**, es decir, a mayores valores de  $FE_{NO}$ , mayores son las diferencias entre las medidas.



**TABLA I**  
**Estudios de validación del medidor de óxido nítrico en aire exhalado y método estadístico utilizado**

Trabajo <sup>a</sup>	Coefficiente de correlación de Pearson	Coefficiente de correlación intraclase	Gráfico de Bland y Altman	Coefficiente de concordancia de Bland y Altman
1	0,97 y 0,98	Sí	Sí	Sí
2	0,97 y 0,98	No	No	No
3		No	No	Sí
4	0,992	No	Sí	Sí
5	0,977	No	Sí	No
6 y 7		Sí	Sí	Sí
8	0,99	No	No	No
9 y 10	0,94 y 0,96	No	Sí	No
11	0,94-0,99 <sup>b</sup>	No	No	No

## Mito I: “La sonrisa de la esfinge”

¿Por qué no indicas el intervalo de confianza del coeficiente de correlación de Pearson?

Los autores no aportan un intervalo de confianza para dicho coeficiente. Calculado por mí, dicho intervalo del 95% es de 0,83 a 0,96, lo que implica que valores posibles de dicho coeficiente podrían ser inferiores a 0,90, los cuales empíricamente se han clasificado como correlación sólo buena.

## Mito I: “La sonrisa de la esfinge”

¿Eres el único autor que ha investigado la correlación entre dichos instrumentos?

¿Por qué no comparas tus resultados con los publicados en la bibliografía?

Otros trabajos, encontrados en mi revisión bibliográfica, ninguno de los cuales han sido mencionados por los autores, ofrecen coeficientes de correlación de Pearson mejores que el tuyo (0,92) y mucho más cercanos a 1, como se indica en la tabla I.

TABLA I  
Estudios de validación del medidor de óxido nítrico en aire exhalado y método estadístico utilizado

Trabajo*	Coefficiente de correlación de Pearson	Coefficiente de correlación intraclass	Gráfico de Bland y Altman	Coefficiente de concordancia de Bland y Altman
1	0,97 y 0,98	Sí	Sí	Sí
2	0,97 y 0,98	No	No	No
3		No	No	Sí
4	0,992	No	Sí	Sí
5	0,977	No	Sí	No
6 y 7		Sí	Sí	Sí
8	0,99	No	No	No
9 y 10	0,94 y 0,96	No	Sí	No
11	0,94-0,99 <sup>b</sup>	No	No	No



\*Relación de trabajos:

- Alving K, Janson C, Nordvall L. Performance of a new hand-held device for exhaled nitric oxide measurement in adults and children. *Respir Res.* 2006;7:67 [consultado 01/07/2007]. Disponible en: [http://www.nioxmimo.com/pdffiles/Alving\\_Janson\\_Nordvall.pdf](http://www.nioxmimo.com/pdffiles/Alving_Janson_Nordvall.pdf) y en <http://respiratory-research.com/content/7/1/67>
- Nordvall SL, Janson C, Alving K. Agreement between NIOX<sup>®</sup> (stationary device) and NIOX MINO<sup>®</sup> (hand-held device) for fractional exhaled nitric oxide (FE<sub>NO</sub>) measurements in adults and children. *Actas de World Allergy Congress; 2005, junio 26-29; Munich. Poster session 1, Poster No. 164* [consultado 01/07/2007]. Disponible en: <http://www.nioxmimo.com/pdffiles/WAC-poster.pdf>
- Nordvall SL, Janson C, Alving K. Agreement between NIOX and the new hand-held NIOX MINO for FENO measurements in adults and children. *Actas de American Thoracic Society 2005 International Conference; 2005, mayo 20-25; San Diego. Poster Discussion Session. Poster 801, A309* [consultado 03/07/2007]. Disponible en: [http://www.filt.de/Hauptmenue/Aktuelles/Page10464/Abstracts\\_for\\_eNO\\_at\\_ATS\\_2005.pdf](http://www.filt.de/Hauptmenue/Aktuelles/Page10464/Abstracts_for_eNO_at_ATS_2005.pdf).
- Sovijärvi ARA, Järvinen H, Uusi-Simola J, Nikkinen P, Piirilä PL. Accuracy and repeatability of a hand-held nitric oxide analyzer in adults with respiratory symptoms. *European Respiratory Society Annual Congress; 2005, septiembre 17-21; Copenhagen. Poster 3670* [consultado 01/07/2007]. Disponible en: <http://www.nioxmimo.com/pdffiles/ers-poster.pdf>
- Vahlkvist S, Sinding M, Skamstrup K, Bisgaard H. Daily home measurements of exhaled nitric oxide an asthmatic children during natural birch pollen exposure. *J Allergy Clin Immunol.* 2006;117:1272-6.
- Turner SW, McGill C, Malik G. Validation of a hand-held exhaled nitric oxide analyzer for use in children. *Actas de American Thoracic Society 2006 International Conference; 2006, mayo 19-24; San Diego. Poster presentation* [consultado 01/07/2007]. Disponible en: <http://www.nioxmimo.com/pdffiles/MINO%20study.pdf>
- McGill C, Malik G, Turner SW. Validation of hand-held exhaled nitric oxide analyzer for use in children. *Pediatr Pulmonol.* 2006;41:1053-7.
- Torre O, Spencer A, Olivieri D, Pames PJ, Kharitonov SA. Feasibility and repeatability of fractional exhaled nitric oxide (FE<sub>NO</sub>) measurements using a hand-held NO monitoring device in asthma in general practice. *Actas de American Thoracic Society 2006 International Conference; 2006, mayo 19-24; San Diego. Volume 3, Abstracts Issue, p. A484* [consultado 01/07/2007]. Disponible en: [http://www.nioxmimo.com/pdffiles/ATS\\_Torre.pdf](http://www.nioxmimo.com/pdffiles/ATS_Torre.pdf)
- Nair A, Menzies D, Lipworth BJ. Portable exhaled nitric oxide measurement: comparison with the gold-standard technique. *Abstracts, British Thoracic Society Winter Meeting 2006, December 6-8, 2006, London. Thorax.* 2006;61 Suppl 2:75. *Actas de British Thoracic Society Winter Meeting 2006; 2006, diciembre 6-8; London. Poster P059* [consultado 01/07/2007]. Disponible en: <http://www.nioxmimo.com/pdffiles/BTS%20P059.pdf>
- Menzies D, Nair A, Lipworth BJ. Portable exhaled nitric oxide measurement: comparison with the "gold-standard" technique. *Chest.* 2007;131:410-4.
- Chen W, Purohit A, Barnig C, Casset A, De Blay F. Niox<sup>®</sup> and NioxMino<sup>®</sup>: comparison of exhaled NO in grass pollen allergic adult volunteers. *Allergy.* 2007;62:571-2.

<sup>b</sup>Intervalo de confianza del 95%, del coeficiente de correlación.

## Mito I: “La sonrisa de la esfinge”

Como has podido ver en la tabla I, aunque otros autores también han utilizado sólo el coeficiente de correlación de Pearson (trabajos 2, 8 y 11), este valor únicamente indica la fuerza de asociación entre 2 variables (la proximidad de los puntos de los 2 métodos de medida a la recta de regresión) y que dicha relación es lineal, pero no necesariamente que dicha recta sea la bisectriz; ni siquiera una correlación perfecta es sinónimo de concordancia perfecta.

## Mito I: “La sonrisa de la esfinge”

Aunque te puedo creer cuando describes las bondades de la medida de  $FE_{NO}$  con el nuevo equipo portátil (sencillez, rapidez, manejabilidad, coste, etc.) y su utilidad en el diagnóstico y seguimiento del paciente asmático, lo que en ningún caso, te puedo admitir del análisis estadísticos de tus datos es que dicho equipo sea concordante (fiable) con el usado como referencia, ni que los valores de referencia propuestos sean válidos sin aportar información acerca de la forma de la distribución de la variable.

An Pediatr (Barc). 2007; 67(6): 572-7



ANALES DE PEDIATRÍA

[www.elsevier.es/anpediatr](http://www.elsevier.es/anpediatr)



## ORIGINALES BREVES

# Concordancia entre dos dispositivos de medida de óxido nítrico exhalado

O. Sardón Prado<sup>a</sup>, A. Aldasoro Ruiz<sup>a</sup>, J. Korta Murua<sup>a</sup>, J. Mintegui Aramburu<sup>a</sup>, J.I. Emparanza Knorr<sup>b</sup> y E.G. Pérez-Yarza<sup>a</sup>

<sup>a</sup>Unidad de Neumología Infantil. Servicio de Pediatría. <sup>b</sup>Unidad de Epidemiología Clínica e Investigación. Hospital de Donostia. San Sebastián. España.

## Análisis estadístico

“En el análisis estadístico se ha estudiado la regresión de las medias de FENO obtenidas con ambos dispositivos de medida respecto a la diferencia de las mismas, mediante el método Bland-Altman, tanto en valores absolutos como relativos. Asimismo, se ha analizado el grado de acuerdo entre los dos dispositivos de medida, en relación con los valores de normalidad para ambos, mediante el estadístico kappa de Cohen”.

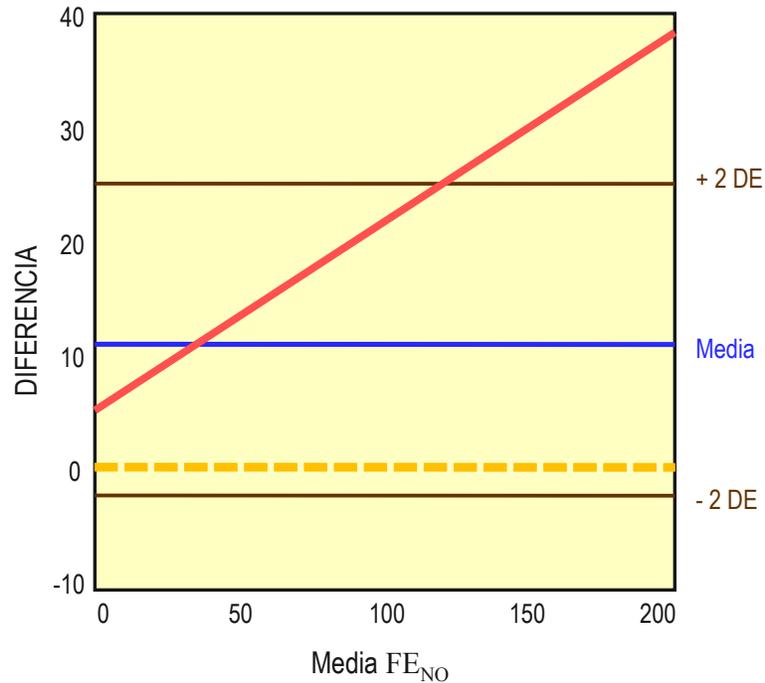


Figura 2: Análisis estadístico (Bland-Altman). Media FE<sub>NO</sub>: media de la fracción exhalada de óxido nítrico; DIF: diferencia de las medias de FE<sub>NO</sub> obtenidas con NIOX y con NIOX-MINO, en valores absolutos.

## Mito I: “La sonrisa de la esfinge”

- Existe un error sistemático porque casi todos los puntos se distribuyen preferentemente por encima de la recta correspondiente a la diferencia 0 entre medidas (línea horizontal naranja).
- El error sistemático es PROPORCIONAL porque la diferencia entre ambos métodos es LINEAL (línea rosa) a lo largo del rango de medición.
- La línea azul representa la media de las diferencias, que en nuestro ejemplo corresponde a 11 (y nos mide el error sistemático del segundo método –portátil-respecto al primero).
- Las líneas marrones representan los límites de confianza definidos mediante  $\pm 2$  DE de la media de la diferencia (de -3 a +25), y se denominan límites de concordancia.

## Análisis estadístico mediante el índice kappa de Cohen

NIOX-MINO	NIOX		TOTAL
	Anormal (+)	Normal (-)	
Anormal (+)	9	1	10
Normal (-)	2	18	20
Total	11	19	30

Punto de corte de  $FE_{NO}$ : media de la fracción exhalada de óxido nítrico.

Medida de acuerdo	Valor de Kappa	EEk	Z	Significación
Kappa de Cohen	+ 0,780	0,120	4,286	0,000
Nº de casos válidos	20			

## Resultados (fragmento del texto)

“La regresión de las medias obtenidas por ambos dispositivos de medida de FENO respecto a la diferencia de las mismas fue significativa, tanto en valores absolutos como relativos ( $p < 0,005$ ) (fig. 2). Asimismo, hubo acuerdo entre los valores obtenidos con ambos dispositivos de medida, kappa de Cohen 0,78 (fig. 3)”.

## Discusión (fragmento del texto)

“El objetivo de nuestro estudio ha sido analizar la concordancia entre dos dispositivos de medida para NO exhalado, sistema NIOX y sistema portátil NIOX-MINO. Según los datos obtenidos en nuestra muestra, observamos que existen diferencias significativas ( $p < 0,005$ ) en la regresión de las medias obtenidas por ambos dispositivos de medida respecto a la diferencia de las mismas. Por tanto, los dos sistemas estudiados **no son equivalentes**, por lo que **no son sustituibles**. De acuerdo con otros investigadores, los valores obtenidos con el sistema NIOX-MINO han sido significativamente superiores a los obtenidos con el sistema NIOX, tanto más cuanto mayor ha sido el valor de  $FE_{NO}$  obtenido (fig. 2)”.

## Discusión (conclusiones)

“En conclusión, la  $FE_{NO}$ , determinada mediante técnica electroquímica con sistemas portátiles, es válida en niños mayores de 5 años y ofrece la posibilidad de cuantificar de forma simple y no invasiva la inflamación eosinofílica de la vía aérea, que resulta de gran utilidad como complemento a la función pulmonar en el diagnóstico de asma y en el control del tratamiento en niños.”

## Introducción (fragmento del texto: objetivos)

“El objetivo de este estudio ha sido analizar la concordancia entre dos dispositivos de medida de NO exhalado, un analizador de quimioluminiscencia estacionario (NIOX, Aerocrine, Suecia), considerado como analizador”.

# Capítulo 12

## Publicación Bioestadística: Análisis multivariante.

**Dr. Jacobo Díaz Portillo**  
Coordinador de Formación Continuada  
Hospital Universitario de Ceuta

## Fases del análisis estadístico

1. ANÁLISIS DESCRIPTIVO



2. ANÁLISIS BIVARIANTE



3. ANÁLISIS MULTIVARIANTE

# Anova de 2 factores

Variable  
Predictora 1

Causa o Factor 1: Cualitativa



Variable RESULTADO

Interacción



Desenlace o efecto:  
Cuantitativa

Variable  
Predictora 2

Causa o Factor 2: Cualitativa



# Anova de 2 factores

Factor 1:

Variable predictora 1

Grupo cultural

F= 7,49 p<0,01



Interacción



F= 1,09 p<0,01



Factor 2:

Variable predictora 2

Tabaquismo materno



F= 0,59 N.S

Variable RESULTADO  
Cuantitativa

Desenlace o efecto:  
Peso del RN

## Análisis de la varianza (ANOVA) de dos factores

Variable dependiente cuantitativa: Peso del RN

Factor 1 cualitativa: Grupo cultural

Factor 2 cualitativa: Tabaquismo materno

Estudio de la variable resultado peso del RN. Efecto de las variables predictoras grupo cultural y hábito de tabaquismo en las madres.

Factor	Varianza	F	Significación
Grupo cultural	3,2631	7,49	$p < 0,01$
Tabaquismo	2,5776	0,59	N.S
Interacción (*)	4,7592	1,09	$p < 0,01$

(\*) Test ANOVA de dos factores

Las diferencias en el peso de los RN se mantuvieron sólo dependientes del grupo cultural o de la interacción de ambos factores, pero no del hábito de fumar. A ello pudo contribuir el bajo número de madres musulmanas fumadoras.

# Anova de 2 factores

Factor 1:

Variable predictora 1

Grupo cultural

F= 3,67 p<0,01



Variable RESULTADO  
Cuantitativa

F= 0,34 N.S



Desenlace o efecto:  
VCM



Factor 2:

Variable predictora 2

Tipo de lactancia

F= 1,03 p<0,01



## Análisis de la varianza (ANOVA) de dos factores

Variable dependiente o resultado cuantitativa: VCM

Factor 1 cualitativa (variable predictora principal): Grupo cultural

Factor 2 cualitativa (variable predictora secundaria): Tipo de lactancia

Estudio de la variable resultado VCM en lactantes. Efecto de las variables predictoras grupo cultural y tipo de lactancia.

Factor	Varianza	F	Significación
Grupo cultural	6,0017	3,673	p < 0,01
Lactancia	1,6904	1,034	p < 0,01
Interacción (*)	0,5579	0,341	N.S

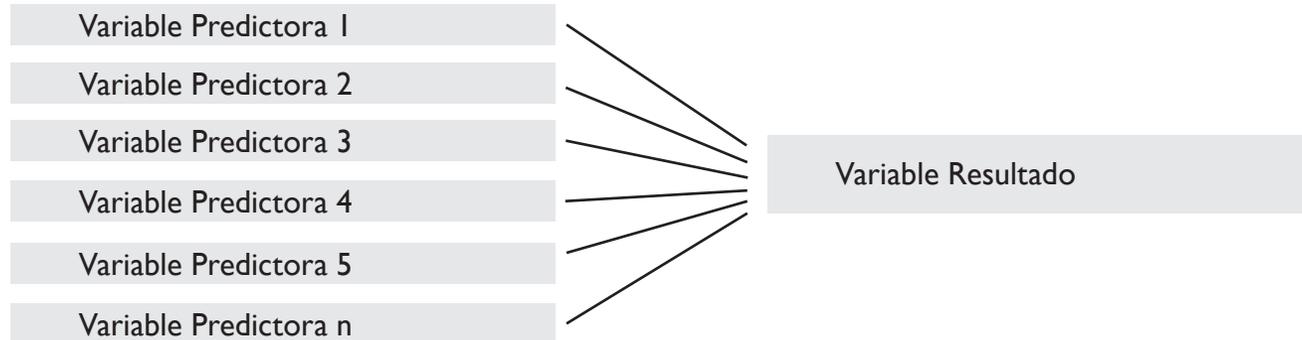
(\*) Test ANOVA de dos factores

Las diferencias en el VCM (microcitosis) se mantuvieron dependientes tanto del grupo cultural como del tipo de lactancia, pero no de la interacción de ambas variables.

Díaz Portillo J. «Influencia del grupo étnico en el patrón lipoproteico de una determinada población pediátrica de Ceuta». Tesis doctoral . Facultad de Medicina. Universidad de Cádiz . Noviembre 1997.

## Análisis multivariante

Análisis multivariante para la detección de potenciales factores de confusión



## Pruebas estadísticas de contraste de hipótesis más aplicables a cada situación

Variable Independiente o PREDICTORA	Variable Dependiente o RESULTADO	Pruebas de contraste de hipótesis	Observaciones y/o limitaciones del test
Cualitativa	Cualitativa	Ji cuadrado Prueba exacta de Fisher Test de McNemar Q de Cochran REGRESIÓN LOGÍSTICA	Si el tamaño muestral es grande Si el tamaño muestral es pequeño Dos grupos de muestras repetidas o apareadas Más de dos grupos de muestras repetidas o apareadas Varias variables predictoras cualitativas
Cualitativa	Cuantitativa	t de Student Análisis de la Varianza Test de Mann-Whitney Test de Wilcoxon Test de Kruskal-Wallis Test de Friedman	Compara medias entre dos grupos independientes Compara medias entre más de dos grupos independientes Dos grupos independientes No paramétricos Dos grupos medidas repetidas No paramétrico Más de dos grupos No paramétrico, medidas repetidas Más de dos grupos No paramétrico medidas repetidas
Cuantitativa	Cualitativa	Regresión logística simple Regresión logística multivariante	Una sola variable predictora cuantitativa Varias variable predictoras cuantitativa o cualitativas
Cuantitativa	Cuantitativa	Regresión lineal simple Correlación de Pearson Correlación de Spearman REGRESIÓN MÚLTIPLE	Paramétrico Paramétrico No paramétrico Varias variables predictoras cuantitativas
Cualitativa	Supervivencia	Kaplan-Meier Long-Rank Regresión de COX	Curvas de supervivencia Compara curvas de supervivencia Análisis Multivariante

	REGRESIÓN LINEAL MÚLTIPLE	REGRESIÓN LOGÍSTICA MULTIVARIANTE
Variable RESULTADO	Cuantitativa	Cualitativa (dicotómica)
Variables PREDICTORAS	Cualitativa Cuantitativa	Cualitativa Cuantitativa
ECUACIÓN	$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$	$\text{Log} \frac{P}{1 - P} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$

## Regresión logística multivariante

Es un modelo de regresión que permite estudiar si una variable resultado de tipo cualitativa nominal depende, o no, de otras variables predictoras. La naturaleza de las variable predictoras puede ser:

- Categórica DICOTÓMICA, es decir, sólo puede admitir dos posibles resultados.
- CUANTITATIVA (discreta o continua).

## Regresión logística multivariante

$$\text{Log} \frac{P}{1 - P} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

Variable Predictora 1

Variable Predictora 2

Variable Predictora 3

Variable Predictora 4

Variable Predictora 5

Variable Predictora n

Variable Resultado

Cualitativa dicotómica

## Ejemplos de posibles modelos de regresión logística

Variable RESULTADO	Variables PREDICTORAS
Curación (si/no)	tratamiento aplicado grado de cumplimiento género comorbilidad
HTA (si/no)	ingesta de sal antecedentes familiares edad
Morir por cáncer de mama	grado del tumor edad tipo histológico receptores hormonales
RN de bajo peso	HTA, edad, tabaquismo, peso materno, raza, complicaciones obstétricas

## Aplicación de los modelos de regresión logística

- En ENSAYOS CLÍNICOS en los que se evalúa un evento o resultado en relación a uno o más factores etiológicos, predictores o asociados (RR).
- En estudios de COHORTES en los que se evalúa un evento o resultado en relación a uno o más factores de riesgo (etiológicos, predictores o asociados) (RR).
- En estudios de CASOS Y CONTROLES en los que se evalúan diferentes factores en cada grupo, para encontrar asociaciones entre ellos y la enfermedad o caso (OR).
- En estudios TRANSVERSALES en los que se evalúan diferentes variables predictoras (factores de riesgo), para encontrar asociaciones entre ellas y la enfermedad en estudio (variable resultado).

## Objetivos de los modelos de regresión logística

1. ESTIMAR la relación entre una o más variables predictoras con la variable resultado.
  - Estudios etiológicos.
  - Confusión e interacción.
2. PREDECIR lo mejor posible la variable RESULTADO o respuesta, usando la combinación más eficiente de las variables PREDICTORAS incluidas en el estudio.

$$\text{Log} \frac{P}{1 - P} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

## Características de los modelos de regresión logística multivariante

- La variable RESULTADO es siempre cualitativa discreta

Si = 1

No = 0

- Las variables PREDICTORAS del modelo puede ser

Continuas

Discretas

- La ecuación inicial del modelo NO es una función lineal, sino EXPONENCIAL, si bien una sencilla transformación logarítmica (LOGIT) la convierte finalmente en una función lineal, que sigue la ecuación de la recta cuyas pendientes son los distintos coeficientes de regresión  $\beta_i$ .

$\beta$

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$$

$$P = \frac{e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}}$$



$$\text{Log} \frac{P}{1-P} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

## Análisis de regresión logística múltiple

$$\text{Log} \frac{P}{1 - P} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

$$P = \frac{e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}}$$

$$O.R. = e^{\beta_i (X_{i1} - X_{i2})}$$

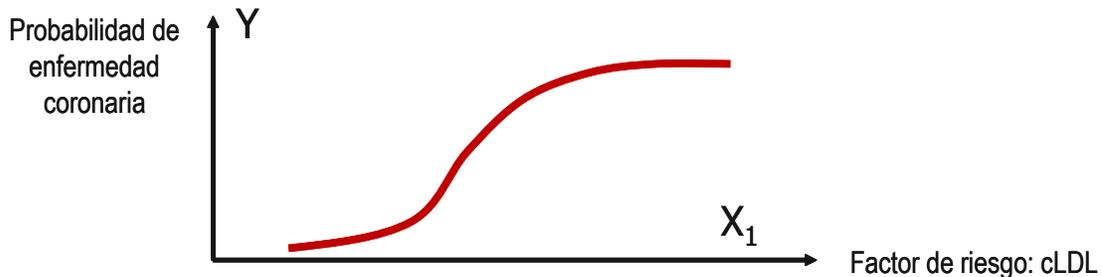
$$O.R. = e^{\beta_1 (x_{11} - x_{12}) + \beta_2 (x_{21} - x_{22}) + \dots + \beta_p (x_{p1} - x_{p2})}$$

# Ventajas de los modelos de regresión logística multivariante (I)

1. Es un modelo útil en muchas circunstancias de investigación en medicina...
  - En las que se quiere analizar o predecir una respuesta dicotómica de la variable resultado
    - Enfermo / sano
    - Vive / muere
    - Presente / ausente
    - Positivo / negativo
  - En función de su perfil de variables predictoras o factores de riesgo, que pueden ser
    - cualitativas: género, sexo, grupo cultural, estado civil, nivel de estudios, nivel socioeconómico,...
    - cuantitativas: edad, nº de cigarrillos, IMC, colesterol, TAS,...

## Ventajas de los modelos de regresión logística multivariante (2)

2. Es la mejor aproximación a la realidad biológica, ya que la influencia de una variable predictora o factor de riesgo sobre la probabilidad de que aparezca una enfermedad (variable resultado) suele asemejarse más a una curva sigmoidea (función exponencial) que a una recta. La curva sigmoidea es la representación habitual de una respuesta a diferentes dosis de un estímulo.



## Ventajas de los modelos de regresión logística multivariante (3)

3. El modelo matemático del modelo permite obtener directamente la ODDS RATIO (OR) o el RIESGO RELATIVO (RR) para cada factor de riesgo o variable PREDICTORA incluida en el modelo. El coeficiente  $\beta_i$  de cada variable predictora obtenido por regresión logística es el logaritmo neperiano de la OR o del RR del factor de riesgo.

$$OR = e^{\beta_1(x_{11}-x_{12}) + \beta_2(x_{21}-x_{22}) + \dots + \beta_p(x_{p1}-x_{p2})}$$

$$OR_1 = e^{\beta_1}$$

$$RR = e^{\beta_1(x_{11}-x_{12}) + \beta_2(x_{21}-x_{22}) + \dots + \beta_p(x_{p1}-x_{p2})}$$

$$RR_1 = e^{\beta_1}$$

## Funciones de los modelos de regresión logística multivariante

1. CUANTIFICAR la relación existente entre las distintas variables predictoras incluidas en el modelo y la variable resultado (respuesta).
2. CONTROLAR Y AJUSTAR aquellas variables predictoras que actúan como potenciales FACTORES MODIFICADORES de EFECTO, especialmente los FACTORES de CONFUSIÓN.
3. PREDECIR LA RESPUESTA de la variable resultado en función de las variables predictoras o factores de riesgo, clasificando a los individuos en una de las dos posibles respuestas en función de la probabilidad que arroje su perfil de variables predictoras incluidas en el modelo de regresión logística multivariante.

## VARIABLES EXPLICATIVAS O PREDICTORAS EN EL MODELO DE REGRESIÓN LOGÍSTICA

VARIABLES CUALITATIVAS	VARIABLES CUANTITATIVAS
El modelo asume estimaciones basadas en variables continuas	La OR y el RR son medidas que cuantifican el cambio en el riesgo cuando se pasa de un valor del factor a otro
Las variables cualitativas deben ser modeladas en 2 categorías (0,1) o transformadas en tantas variables dicotómicas como categorías (variables DUMMY)	Este aumento o disminución del riesgo es lineal y proporcional al cambio

## Regresión logística:

### Estimación de coeficientes $\beta_i$

$\beta_0, \beta_1, \beta_2, \dots, \beta_n$  son los coeficientes de regresión logística del modelo

- $\beta_0$  es el término independiente (o constant en SPSS)
- El resto de las  $\beta_i$  son los coeficientes de regresión logística para las diferentes variables predictoras incluidas en el modelo
- La evidencia de relación entre una variable predictora y la variable resultado es la existencia de un coeficiente de regresión logística significativamente diferente a cero  $\beta_i \neq 0$
- El signo de  $\beta_i$  tiene un significado importante porque
  - Si es positivo ( $\beta_i > 0$ ) significa que esa variable predictora AUMENTA la PROBABILIDAD del suceso (variable resultado o respuesta) que estamos analizando.
  - Si es negativo ( $\beta_i < 0$ ) su cálculo se lleva a cabo mediante la FUNCIÓN DE MÁXIMA VEROSIMILITUD.

## Regresión logística:

### Definición de las hipótesis nula $H_0$ y alternativa $H_a$

- La hipótesis conceptual es que “una o varias variables predictoras influyen significativamente en la probabilidad del suceso en estudio (que la variable predictora tome un valor u otro)”
- Las hipótesis operativas se reducen al contraste de los diferentes coeficientes  $\beta_i$  obtenidos en el análisis de los datos como estimadores de las variables predictoras de interés

Así, cuando sólo hay una variable predictora  $X_1$

$$H_0 \quad \rightarrow \quad \beta_1 = 0$$

$$H_a \quad \rightarrow \quad \beta_1 \neq 0$$

- Hay varios procedimientos para llevar a cabo este contraste.

## Regresión logística:

### Procedimientos para evaluar el contraste de hipótesis

1. Evaluación estadística del coeficiente  $\beta_i$
2. Evaluación estadística de WALD
  - El estadístico de WALD es el resultado de dividir el cuadrado del coeficiente de regresión logística  $\beta_i$  por el cuadrado de su error estándar  $EE\beta_i$
3. Evaluación estadística de la razón de verosimilitud ( $L_0$ )
4. Evaluación estadística del cambio del estadístico  $-2LL_0$ 
  - Este estadístico resulta ser menor de dos veces el logaritmo neperiano de la razón de verosimilitud ( $L_0$ )
5. Evaluación del estadístico  $Z^2$ , (Goodness of Fit)

## I. Evaluación estadística del coeficiente $\beta_i$

- El coeficiente de regresión logística  $\beta_i$  de la variable predictora  $i$  debe ser estadísticamente distinto de cero  $\beta_i \neq 0$  para que esta tenga influencia significativa sobre la variable resultado.
- La hipótesis  $H_0$  será que...  $\beta_i = 0$
- La hipótesis  $H_a$  será que...  $\beta_i \neq 0$
- Para tener en cuenta el error aleatorio ( $EE\beta_i$ ) con que el coeficiente  $\beta_i$  (obtenido en la regresión logística) estima al coeficiente de regresión poblacional  $\beta_\pi$  se lleva a cabo un contraste a través del estadístico  $t$  de Student con un número de grados de libertad igual al número de individuos de la muestra ( $N$ ) menos el número de variables predictoras ( $n$ ) introducidas en el modelo.

## Interpretación del coeficiente de regresión logística $\beta_i$

- Si su signo es positivo ( $\beta_i > 0$ ), la probabilidad de que se dé o aparezca la característica de la variable predictora aumentará según aumente el valor de  $x_{ij}$  (la variable resultado a quien afecta).
- Si  $\beta_i \neq 0$  es significativamente distinta de cero (test de contraste de hipótesis  $p < 0,05$ ), la variable predictora  $i$  influye significativamente en la probabilidad de que se dé o aparezca la característica de la variable resultado, y el modelo logístico tiene sentido.

## 2. Evaluación estadística de Wald

El estadístico de Wald ( $W$ ) es también muy utilizado para evaluar la significación estadística del modelo de regresión logística multivariante.

El estadístico  $W$  se obtiene dividiendo el cuadrado de  $\beta_i$  por el cuadrado del error estándar de  $\beta_i$

$$W = \frac{[\beta_i]^2}{[EE\beta_i]^2}$$

El estadístico de Wald se distribuye según una CHI CUADRADO.

La hipótesis a contrastar son las mismas que en el caso anterior

La hipótesis  $H_0$  será que...  $W = 0$

La hipótesis  $H_a$  será que...  $W \neq 0$

## Evaluación del modelo de Regresión logística multivariante mediante el programa estadístico SPSS

ESTADÍSTICO	HIPÓTESIS NULA	CONTRASTE	SIGNIFICACIÓN
$\beta_i$	$\beta_i = 0$	t de Student	$p < 0,05$
Wald $(\beta_i)^2 / (EE\beta_i)^2$	$\beta_i = 0$	CHI CUADRADO	$p < 0,05$
- 2 Log Likelihood (-2LL)	$L = 1$ ó $(-2LL) = 0$ “ajuste perfecto”	CHI CUADRADO con n-k-1 grados de libertad	Si rechaza la hipótesis nula ( $p < 0,05$ ) significa que el modelo de regresión no es significativo
Improvement o cambio en la razón de verosimilitud en cada paso	$(-2LL_{k-1}) - (-2LL_k)$	CHI CUADRADO con k grados de libertad	$p < 0,05$
$Z^2$ (Goodness of Fit)	$Z^2 \neq 0$	CHI CUADRADO	Si se rechaza la hipótesis nula ( $p < 0,05$ ) significa que el modelo de regresión no es significativo
Model Chi- Square	$(-2LL^0) - (-2LL^K)$	CHI CUADRADO con k grados de libertad	$p < 0,05$

## Evaluación del modelo de Regresión logística multivariante mediante el programa estadístico SPSS

ESTADÍSTICO	HIPÓTESIS NULA	CONSTANTE	SIGNIFICACIÓN
$\beta_i$	$\beta_i = 0$	t de Student	$p < 0,05$
Wald $\frac{[\beta_i]^2}{[EE\beta_i]^2}$	$W = 0$	CHI CUADRADO	$p < 0,05$

Si se rechaza la hipótesis nula ( $H_0$ ) ( $p < 0,05$ ) significa que el modelo de regresión es significativo.

¿Es posible predecir el estado de los depósitos medulares de hierro conociendo los niveles séricos de ferritina y el VCM del hemograma?

Variable predictora 1

VCM



Variable resultado cualitativa

Depósitos de hierro en  
Médula ósea

1 = Depósitos de hierro normales  
0 = Déficit de hierro medular



Variable predictora 2

FERRITINA  
(sérica)



## Análisis de regresión lineal múltiple

Posible variables predictoras asociadas a los depósitos de hierro en médula ósea		
Variables predictoras	Estimación $\beta$	Error estándar $EE\beta$
Constante	- 0,6799	
VCM	- 0,02223	0,03835
FERRITINA SÉRICA	0,1919	0,05525

Para que una variables PREDICTORAS esté relacionada estadísticamente con la variable resultado (depósitos de hierro en médula ósea) debe cumplirse siempre la condición de que...

$$\frac{\beta_{\text{Ferritina}}}{EE\beta_{\text{Ferritina}}} = \frac{0,1919}{0,05525} = 3,47 > Z_{95\%}(1,96) \rightarrow p < 0,05$$

$$\frac{\beta_{\text{VCM}}}{EE\beta_{\text{VCM}}} = \frac{-0,02223}{0,03835} = 0,06 \ll Z_{95\%}(1,96) \rightarrow p > 0,05$$

Método de introducción de las variables	REGRESIÓN LINEAL MÚLTIPLE	REGRESIÓN LOGÍSTICA MULTIVARIANTE
Forward (hacia adelante)	Introducir progresivamente en el modelo las variables predictoras estadísticamente significativas, comenzando por la que tenga mayor significación	
Backward (hacia atrás)	Comienza por estimar un modelo que contiene todas las variables predictoras para ir eliminando posteriormente una a una, de forma sucesiva, aquellas que no sean significativas	
Stepwise	Es una combinación de los anteriores	

## Análisis de Regresión Logística múltiple

¿Qué relación existe entre las variables predictoras VCM y Ferritina sérica y los depósitos de hierro medular?

Variables PREDICTORAS

Variable RESULTADO

VCM y Ferritina sérica



Depósitos de Fe medular

$$\log \frac{P}{1 - P} = \beta_0 + \beta X_1 + \beta_2 X_2$$

$$\log \frac{P}{1 - P} = -2,1290 + 0,1713 \text{ FERRITINA}$$

P = probabilidad de que existan depósitos de hierro medular

## Análisis de Regresión Logística múltiple

¿Qué relación existe entre las variables predictoras VCM y Ferritina sérica y los depósitos de hierro medular?

Variables PREDICTORAS

Variable RESULTADO

VCM y Ferritina sérica



Depósitos de Fe medular

$$OR = e^{\beta_i (X)}$$

$$OR = e^{0,1713 (X_a - X_b)}$$

$X_a$  = Niveles de ferritina sérica del paciente A

$X_b$  = Niveles de ferritina sérica del paciente B

¿Qué diferencia promedio de probabilidades de tener depósitos medulares normales existe entre un paciente A con 24 ng/ml de ferritina y otro B con 3 ng/ml?

Variables predictoras	$\beta_i$	Paciente A	PACIENTE B
Ferritina sérica (ng/ml)	0,1713	24 ng/ml	3 ng/ml

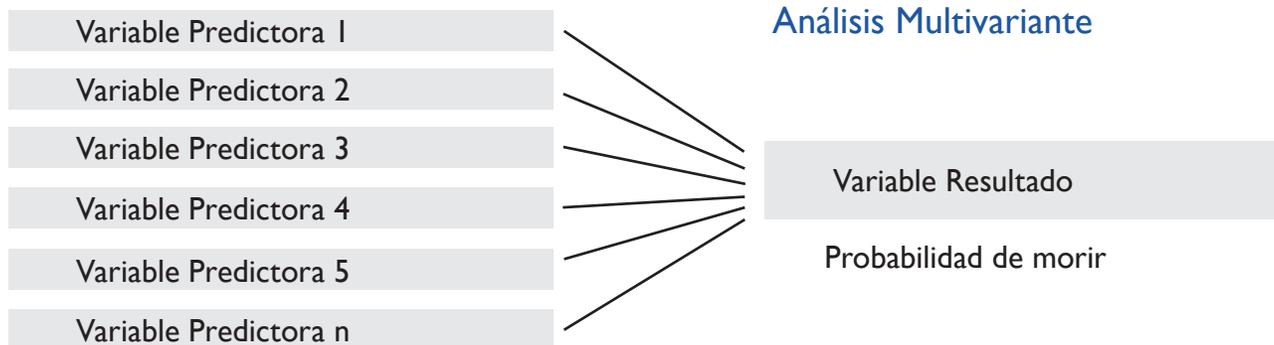
$$OR = e^{\beta_i (X_a - X_b)}$$

$$OR = e^{0,1713 (24 - 3)} = 36,6$$

El paciente A tiene una probabilidad de 37 veces de tener depósitos medulares de hierro normal frente al paciente B.

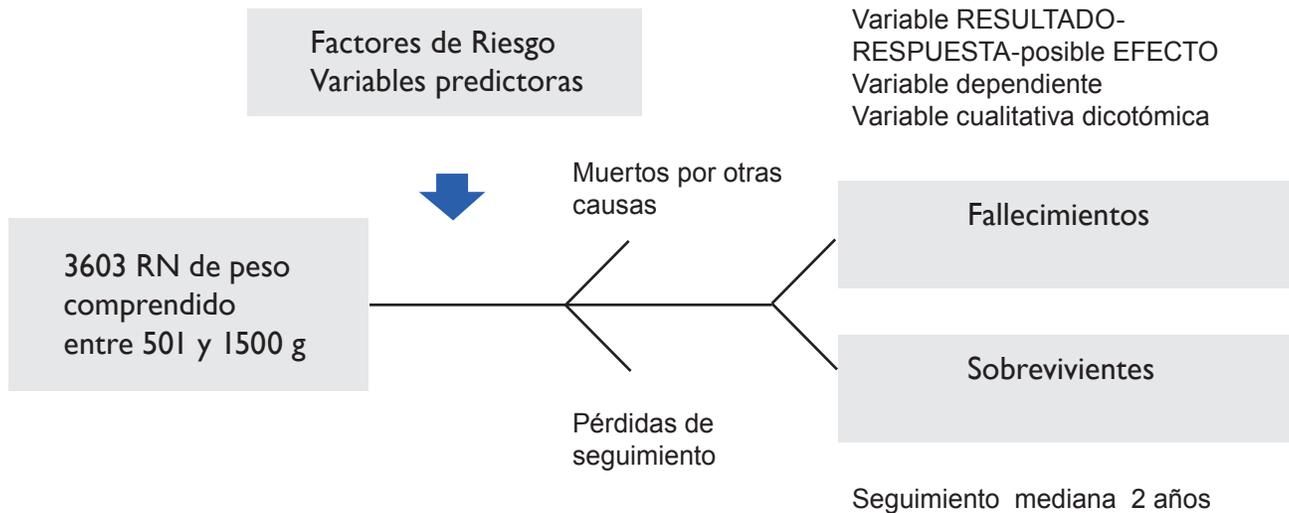
## Pregunta de investigación

- ¿Qué relación existe entre diversas posibles variables predictoras y la probabilidad de muerte del recién nacido de bajo peso?
- Estudio de COHORTES realizado por Horbar et al (1993) con 3603 RN de peso entre 501-1500 g (2 años de seguimiento).



Horbar JD, Onstad L, Wright E. Predicting mortality risk for infants weighing 501 to 1500 grams at birth: The National Institutes of Health Neonatal Research Network report. Crit Care Med 1993; 21:12–18.

## Estudio de cohortes prospectivo de Recién Nacidos



Horbar JD, Onstad L, Wright E. Predicting mortality risk for infants weighing 501 to 1500 grams at birth: The National Institutes of Health Neonatal Research Network report. *Crit Care Med* 1993; 21:12–18.

## Estudio de cohortes prospectivo de Recién Nacidos

Objetivo: Evaluar el riesgo de muerte en RN de bajo peso (entre 501 – 1500 g)

Variable resultado: Fallecimiento del RN (dicotómica)

Variable predictora [1]: peso en gramos del RN

Variable predictora [2]: Tamaño según edad gestacional

Variable predictora [3]: Raza

Variable predictora [4]: Sexo

Variable predictora [5]: tipo de parto (vaginal/cesárea)

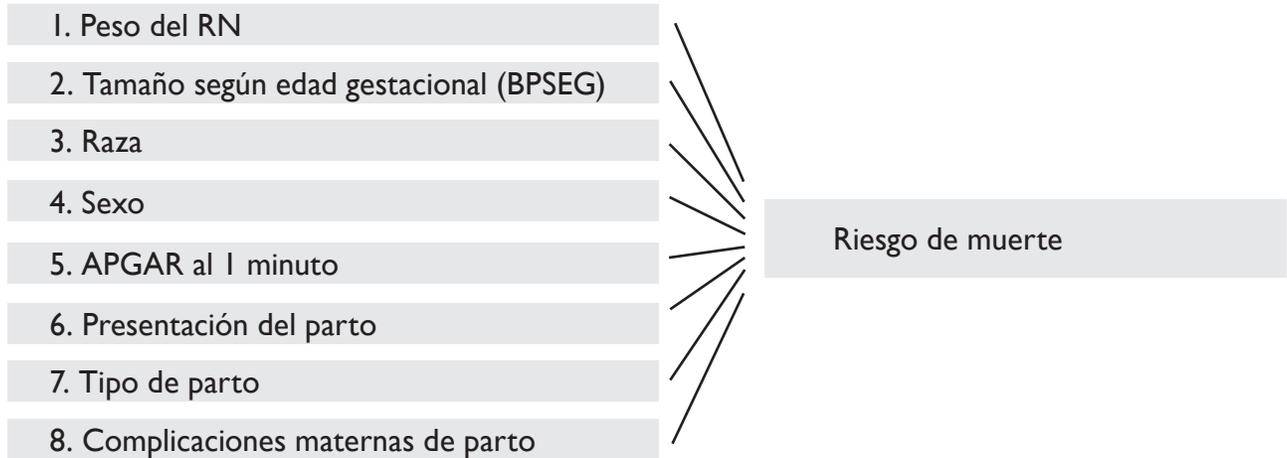
Variable predictora [6]: presentación del parto

Variable predictora [7]: Índice del Apgar al minuto

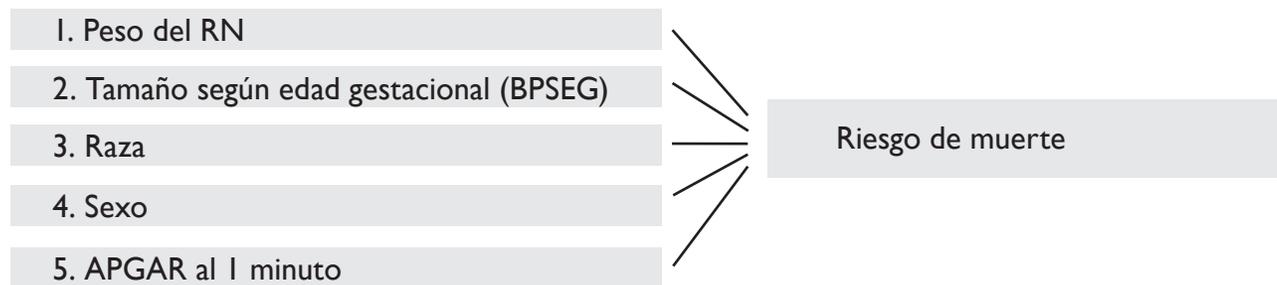
Variable predictora [8]: Presencia o no de complicaciones maternas (diabetes melitus, eclampsia, hipertensión o hemorragia) antes del parto

$$\text{Log} \frac{P}{1 - P} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

$$P = \frac{e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}}$$



## Variables predictoras significativas en la regresión logística multivariante



## Análisis de regresión logística múltiple

Estimación del modelo para predecir la mortalidad en recién nacidos con pesos entre 501 y 1500 gramos

Variables predictoras		
Constante	t de Student	0,029
Peso (en gr/100)	- 0,422	0,157
BPSEG	- 0,491	0,145
Raza	- 0,450	p < 0,05
Sexo	0,656	0,131
Apgar	0,971	0,131

$$\text{Log} \frac{P}{1 - P} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

$$O.R. = e^{\beta_1(x_{11}-x_{12}) + \beta_2(x_{21}-x_{22}) + \dots + \beta_p(x_{p1}-x_{p2})}$$

## Tipos de variables del estudio Horbar et al.

Variable resultado (cualitativa dicotómica):

- Muerte = 1
- Sobrevivió = 0

1. Variable predictora 1: Peso del RN (gramos/100)
2. Variable predictora 2: Bajo peso según la edad gestacional (BPSEG).
  - Peso inferior al P10 para su edad gestacional = 1
  - Peso igual o superior al P10 para su edad gestacional = 0
3. Variable predictora 3: RAZA
  - Raza negra = 1
  - Raza blanca = 0
4. Variable predictora 4: SEXO
  - Hombre = 1
  - Mujer = 0
5. Variable predictora 5: Índice de APGAR
  - Apgar  $\leq 3$  = 1
  - Apgar  $> 3$  = 0

Horbar JD, Onstad L, Wright E. Predicting mortality risk for infants weighing 501 to 1500 grams at birth: The National Institutes of Health Neonatal Research Network report. Crit Care Med 1993; 21:12–18.

¿Qué diferencia promedio de riesgo de morir existe entre un RN A con 850 gramos y otro B con 1250 gramos iguales para el resto de las variables predictoras ensayadas?

Variables predictoras	$\beta_i$	RN A	RN B
Peso en gramos/100	- 0,422	850 g [8,5]	1250 G [12,5]

$$RR = e^{\beta_1(x_{11}-x_{12}) + \beta_2(x_{21}-x_{22}) + \dots + \beta_p(x_{p1}-x_{p2})}$$

$$RR = e^{\beta_i (X_a - X_b)}$$

$$RR = e^{-0,422 (850/100 - 1250/100)} = 5.4$$

Un RN A de 850 g de peso al nacer presenta 5.4 veces más riesgo de morir que otro RN B de 1250 g a igualdad del resto de las variables predictoras ( $X_i$ ) incluidas en el modelo.

¿Qué diferencia promedio de riesgo de morir existe entre un RN A con SFA (Apgar < 3) con respecto a otro RN B sin SFA (Apgar > 3) con igualdad para el resto de las variables predictoras?

Variables predictoras	$\beta_i$	RN A SFA (+)	RN B SFA (-)
APGAR (>3/<=3) [1/0]	0,971	APGAR < 3 [1]	APGAR > 3 [0]

$$RR = e^{\beta_1(x_{11}-x_{12}) + \beta_2(x_{21}-x_{22}) + \dots + \beta_p(x_{p1}-x_{p2})}$$

$$RR = e^{\beta_i (X_a - X_b)}$$

$$RR = e^{0,971 (1-0)} = 2,62$$

Los RN con SFA presenta 2,62 veces más riesgo de morir que los RN sin SFA a igualdad del resto de las variables predictoras (Xi) incluidas en el modelo.

¿Qué diferencia promedio de riesgo de morir existe entre un RN A niño con respecto a otro RN B niña con igualdad para el resto de las variables predictoras incluidas en el modelo?

Variables predictoras	$\beta_i$	RN A niño	RN B niña
SEXO (hombre/mujer) [1/0]	0,656	hombre [1]	mujer [0]

$$RR = e^{\beta_1(x_{11}-x_{12}) + \beta_2(x_{21}-x_{22}) + \dots + \beta_p(x_{p1}-x_{p2})}$$

$$RR = e^{\beta_i (X_a - X_b)}$$

$$RR = e^{0,656(1-0)} = 1,93$$

Los RN varones presentan 1,93 veces más riesgo de morir (un 93%) que los RN mujeres a igualdad del resto de las variables predictoras ( $X_i$ ) incluidas en el modelo.

¿Qué diferencia promedio de riesgo de morir existe entre un RN N de raza negra con respecto a otro RN B de raza blanca con igualdad para el resto de las variables predictoras?

Variables predictoras	$\beta_i$	RN N	RN B
Raza ( $>1/\leq 1$ ) [1/0]	-0,450	Negra [1]	Blanco [0]

$$RR = e^{\beta_1(x_{11}-x_{12}) + \beta_2(x_{21}-x_{22}) + \dots + \beta_p(x_{p1}-x_{p2})}$$

$$RR = e^{\beta_i (X_n - X_b)}$$

$$RR = e^{-0,450 (1 - 0)} = 0.64$$

Los RN de raza blanca están a  $1/0.64 = 1.56$  veces (un 56%) más riesgo de morir que los RN de raza negra a igualdad del resto de las variables predictoras.

Los RN de raza negra presenta 0.64 veces más riesgo de morir que los RN de raza blanca a igualdad del resto de las variables predictoras incluidas en el modelo.

## ¿Qué diferencia promedio de riesgo de morir existe entre dos RN A y B con las siguientes características?

Variables predictoras	$\beta_i$	RN A	RN B
Peso en gramos/100	- 0,422	1150 g [11,5]	1400 g [14,0]
Bajo peso para su edad gestacional BPSEG <sup>a</sup> [1/ 0]	- 0,491	BPPSEG [1]	[0]
Raza (>1/<=1) [1/0]	- 0,450	Negra [1]	Blanco [0]
Sexo (mujer/hombre) [1/ 0]	0,650	Niña [0]	Niño [1]
Apgar (<=3/ >3) [1/ 0]	0,971	Apgar <=3 [1]	Apgar >3 [0]

$$RR = e^{\beta_1(x_{11}-x_{12}) + \beta_2(x_{21}-x_{22}) + \dots + \beta_p(x_{p1}-x_{p2})}$$

$$e^{-0.422(11.5-14.0) - 0.491(1-0) + 0.656(0-1) + 0.971(1-0)} = e^{0.429} = 1.54$$

El RN A (niña negra) presenta un 54% de riesgo de morir que el RN B (niño blanco) según el valor de las variables predictoras incluidas en el modelo.

## Cálculo de la probabilidad (P) de muerte individual

Variables predictoras	RN A	RN C
Peso en gramos/100	1150 g [11,5]	750 g [7,5]
Bajo peso para su edad gestacional BPSEG <sup>a</sup> [1/ 0]	BPPSEG [1]	[0]
Raza (>1/<=1) [1/0]	Negra [1]	Blanco [0]
Sexo (mujer/hombre) [1/ 0]	Niña [0]	Niño [1]
Apgar (<=3/ >3) [1/ 0]	Apgar <=3 [1]	Apgar >4 [1]

$$\text{RN A} \quad P = \frac{e^{2.606 - 0.422(11.5) - 0.491(1) - 0.45(1) + 0.656(0) + 0.971(1)}}{1 + e^{2.606 - 0.422(11.5) - 0.491(1) - 0.45(1) + 0.656(0) + 0.971(1)}} = 0.098$$

$$\text{RN C} \quad P = \frac{e^{2.606 - 0.422(7.5) - 0.491(0) - 0.45(0) + 0.656(1) + 0.971(1)}}{1 + e^{2.606 - 0.422(7.5) - 0.491(0) - 0.45(0) + 0.656(1) + 0.971(1)}} = 0.74$$

## Estudio de Marshall et al. (NEOCOSUR)

- **Objetivo:** Desarrollar una ecuación que permita predecir el riesgo de mortalidad neonatal en RN de muy bajo peso de nacimiento basado en variables predictoras conocidas en el **momento del nacimiento**, antes del ingreso a la unidad de cuidado intensivo neonatal.
- **Métodos:** Fueron incluidos en este estudio 1801 RN con peso al nacimiento de 500 a 1500 gramos.
- **Variables Predictoras:** edad materna, peso de nacimiento, edad gestacional, Apgar al 1 minuto, malformaciones congénitas mayores, sexo, gestación múltiple, uso de esteroides prenatales y peso pequeño para su edad gestacional (BPSEG) (definida como más bajo que Percentil 10 de peso para edad gestacional).

Marshall G, Tapia JL, D'Apremont I, Grandi C. Un nuevo score para predecir el riesgo de mortalidad neonatal del muy bajo peso de nacimiento en la red sudamericana NEOCOSUR. *Journal of Perinatology* 2005; 25:577–582.

## Estudio de Marshall et al (NEOCOSUR)

Objetivo: Predecir el riesgo de mortalidad neonatal en RN de muy bajo peso de nacimiento (entre 500 – 1500 g)

Variable resultado: Fallecimiento del RN (dicotómica)

Variable predictora [1]: Edad materna

Variable predictora [2]: Peso de nacimiento

Variable predictora [3]: Edad gestacional

Variable predictora [4]: Índice del Apgar al minuto

Variable predictora [5]: Malformaciones congénitas mayores

Variable predictora [6]: Sexo

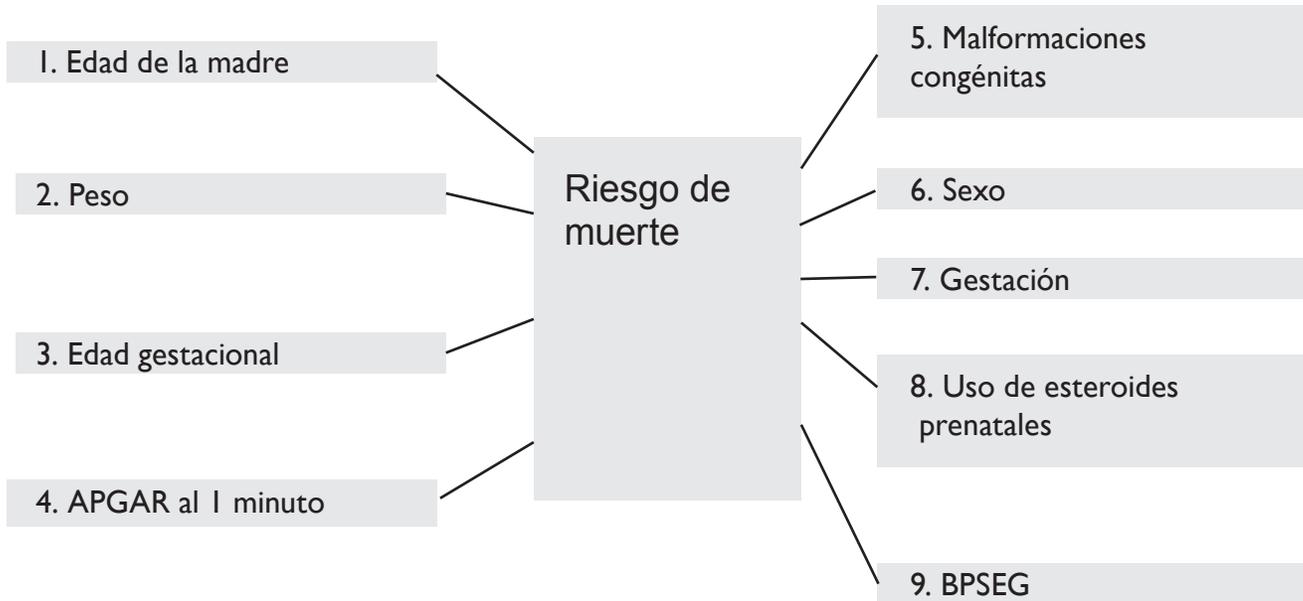
Variable predictora [7]: Gestación múltiple

Variable predictora [8]: Uso de esteroides prenatales

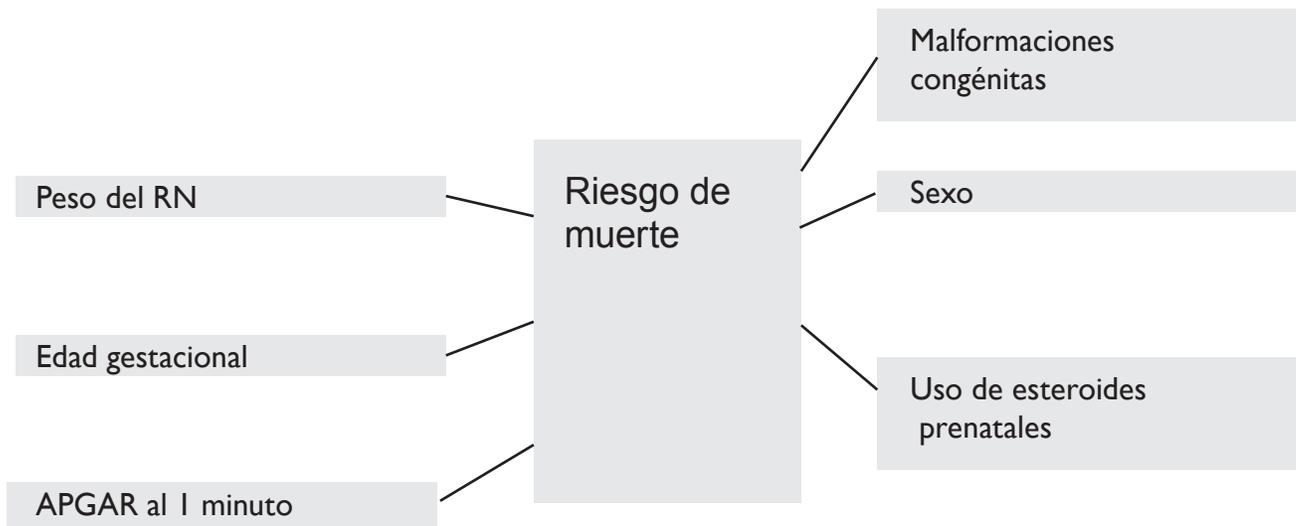
Variable predictora [9]: Peso pequeño para su edad gestacional (definida como más bajo que Percentil 10 de peso para edad gestacional) (BPSEG)

$$\text{Log} \frac{P}{1 - P} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

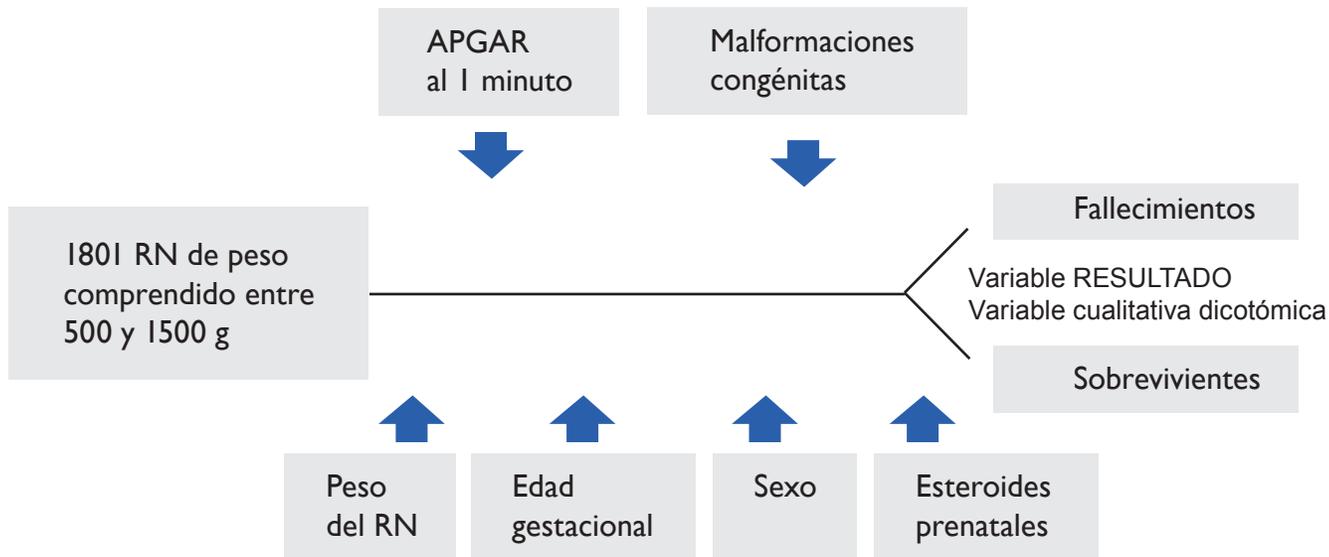
$$P = \frac{e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}}$$



## Variables predictoras significativas en la regresión logística multivariante



## Estudio de Marshall et al (NEOCOSUR)



Marshall G, Tapia JL, D'Apremont I, Grandi C. Un nuevo score para predecir el riesgo de mortalidad neonatal del muy bajo peso de nacimiento en la red sudamericana NEOCOSUR. *Journal of Perinatology* 2005; 25:577–582.

### Análisis de Regresión Logística Múltiple: Estimación del modelo para predecir la mortalidad en recién nacidos con pesos entre 500 y 1500 gramos

Variables predictoras	Estimación Coeficiente $\beta$	Error estándar $EE\beta$	Cociente $\beta / EE\beta$	OR	IC95%
Constante	8,378	0,99			
Peso del RN	- 0,331	0,04	8,27	0,72	0,64 – 0,79
Edad gestacional	- 0,132	0,04	3,3	0,88	0,80 – 0,95
Apgar al minuto	0,265	0,03	8,83	0,77	0,71 – 0,82
Malformaciones congénitas (Si =1; No = 0)	3,419	0,64	5,32	30,55	8,74 – 106,77
Uso de esteroides prenatales (Si =1; No = 0)	- 0,302	0,08	3,77	0,74	0,58 – 0,89
Sexo (mujer =1; hombre = 0)	- 0,474	0,16	2,96	0,62	0,30 – 0,93

$$\text{Log} \frac{P}{1 - P} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

$$O.R. = e^{\beta_1(x_{11}-x_{12}) + \beta_2(x_{21}-x_{22}) + \dots + \beta_p(x_{p1}-x_{p2})}$$

Se ha utilizado el modelo de regresión logística múltiple con un procedimiento paso a paso, descartándose 3 variables e incluyendo 6 variables predictoras que fueron estadísticamente significativos.

## Interpretación del modelo del estudio NEOCOSUR

La Tabla muestra los coeficientes del modelo de regresión logística múltiple  $\beta$ , el error estándar  $EE\beta$  y los odds ratios (OR) asociados. Estos coeficientes para las 6 variables predictoras seleccionadas en el modelo logístico permite evaluar el score NEOCOSUR y puede ser usado para calcular el riesgo de mortalidad del RN con muy bajo peso al nacimiento.

- Un aumento de 100 g en el peso de nacimiento redujo el riesgo de mortalidad intrahospitalaria en 28% (OR = 0,72).
- La adición de 1 punto en el score de Apgar al 1-minuto redujo el riesgo en 23% (OR = 0,77).
- Una semana adicional de edad de gestación redujo el riesgo en 12% (OR= 0,88).
- La presencia de malformación congénita tuvo el mayor efecto de entre todas las variables predictoras (factores de riesgo)(OR=30,55) aumentando notablemente el riesgo de mortalidad neonatal.

$$RRR = [ 1 - RR ] * 100$$

## Variables DUMMY (indicadoras)

Para la introducción de variables policotómicas en el análisis multivariante es necesario crear las llamadas variables dummy. Siempre se elegirá una **categoría de referencia** con la que se compararan el resto de las categorías. Si la variable tiene  $k$  categorías, se proponen  $k - 1$  variables dummy al modelo multivariante.

Hosmer and Lemeshow del Baystate Medical Center realizaron un estudio para conocer los principales factores de riesgo asociados con el bajo peso del RN (definido como peso  $< 2500$  g) en un grupo de 189 mujeres. Una de las 11 variables predictoras evaluadas fue la RAZA, la cual tenía 3 categorías (negra, blanca e hispana).

Para la variable RAZA se tienen que proponer dos variables dummy.

¿Cómo se formularon esas dos nuevas variables?

## Análisis bivalente previo

Relación entre la variable predictora RAZA y la prevalencia del bajo peso al nacer.

RAZA	Bajo peso del RN		Total	(%)
	SI	NO		
Blanca	23	73	96	24%
Hispana	25	42	67	37%
Negra	11	15	26	42%
TOTAL	59	130	189	

Test de  $\chi^2$  Pearson

$$\chi^2 = 5,005 \quad \rightarrow \quad p = 0,082$$

NO existe asociación

## Análisis bivalente previo

Relación entre la variable predictora RAZA y la prevalencia del bajo peso al nacer.

RAZA	Bajo peso del RN		Total	(%)
	SI	NO		
Blanca	23	73	96	24%
Otras razas	36	57	93	39%
TOTAL	59	130	189	

Test de  $\chi^2$  Pearson

$$\chi^2 = 4,787 \quad \rightarrow \quad p = 0,029$$

Existe asociación

## Análisis bivalente previo

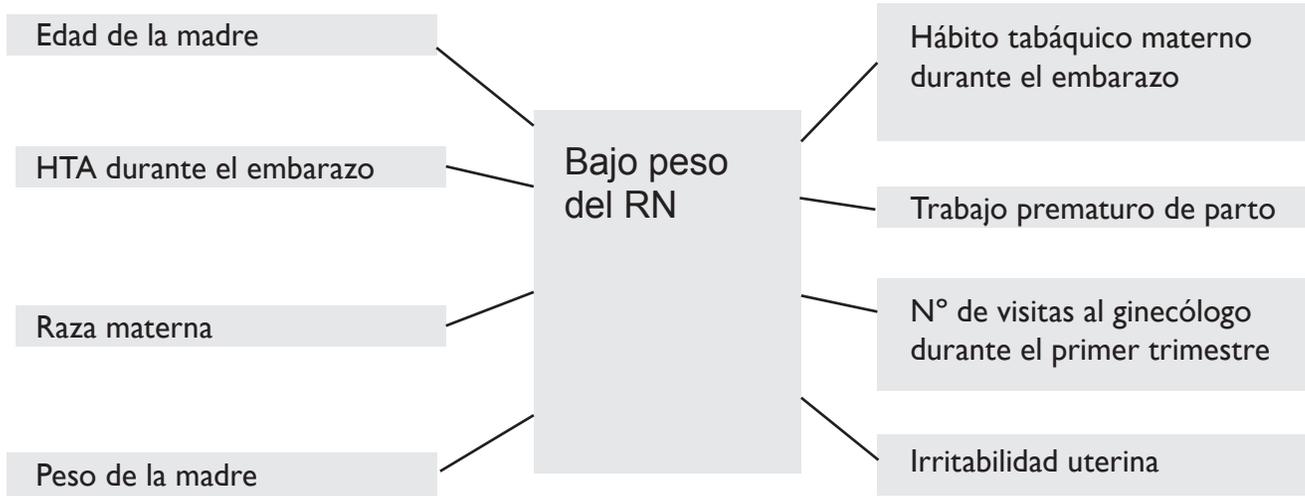
Relación entre la variable predictora RAZA y la prevalencia del bajo peso al nacer.

RAZA	Bajo peso del RN		Caso-control		
	SI	NO	Total	(%)	OR
Blanca	23	73	96	24%	1,00
Hispana	25	42	67	37%	1,89
Negra	11	15	26	42%	2,33
TOTAL	59	130	189		TOTAL

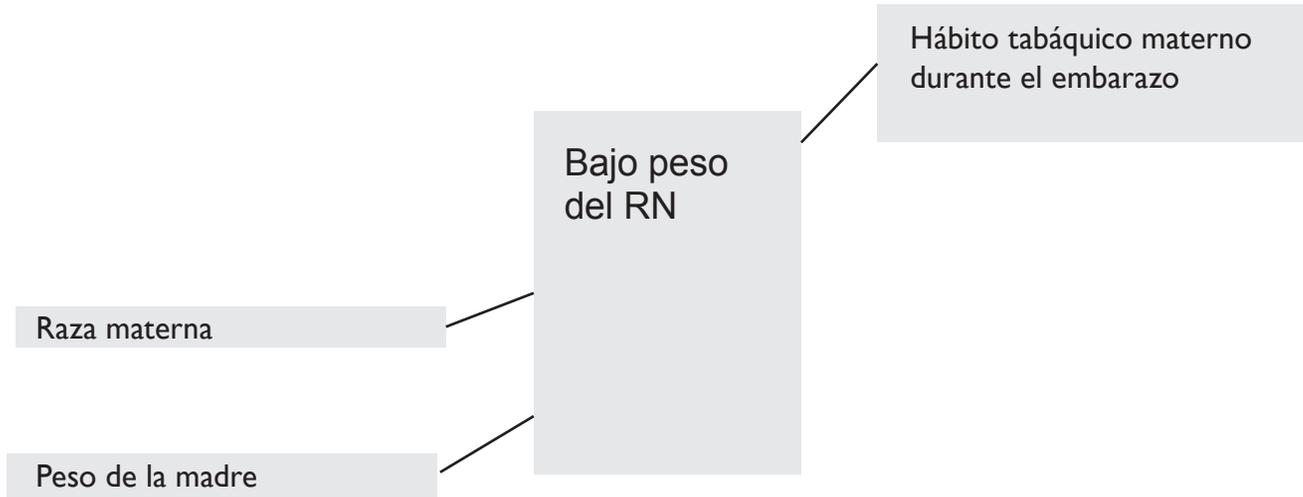
Test de  $\chi^2$  de tendencia lineal

$$\chi^2 = 4,68 \quad \rightarrow \quad p = 0,030$$

Existe asociación



## Variables predictoras significativas en el análisis multivariante



$$\text{Log} \frac{P}{1 - P} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

Hosmer and Lemeshow (2000) Applied Logistic Regression: Second Edition. Baystate Medical Center, Springfield, Massachusetts .

Modelo de regresión logístico para el BAJO PESO AL NACER				
Variables predictoras	Estimación $\beta$	Error estándar $EE\beta$	$\beta/EE\beta$	p
Constante	-0,109	0,882	0,123	0,901
Tabaquismo	1,060	0,378	2,804	0,005
Peso de la madre	-0,027	0,013	2,077	0,036
Raza negra (1ª dummy) (NEGRA=1, blanca =0, hispana=0)	1,290	0,511	2,525	0,012
Raza hispana (2ª dummy) (HISPANA=1, blanca =0, negra=0)	0,970	0,412	2,354	0,019

Para la variable RAZA se tienen que proponer dos variables dummy ( $k-1$ ). La categoría de referencia es la que no tiene variable dummy propia, en esta caso la raza blanca.

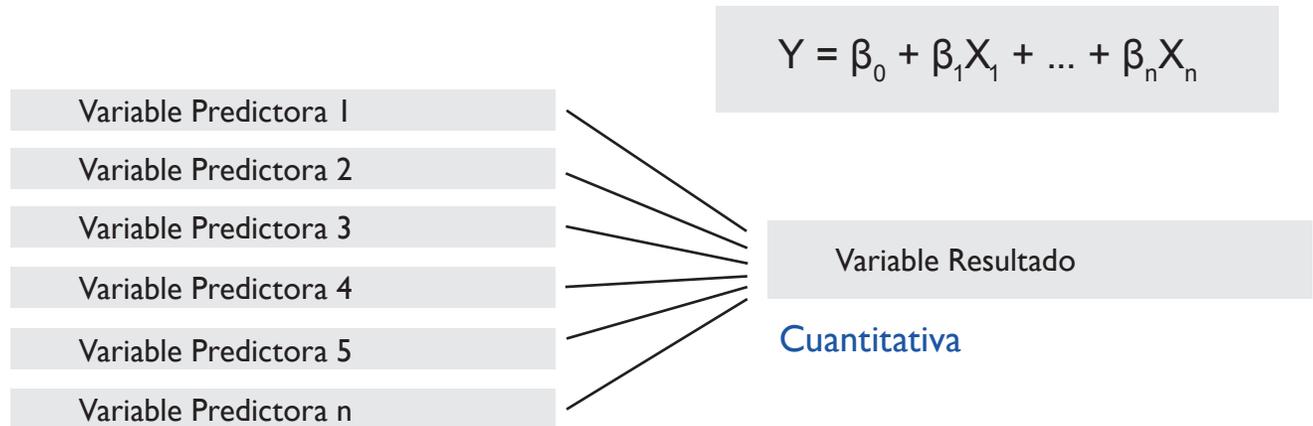
	REGRESIÓN LINEAL MÚLTIPLE	REGRESIÓN LOGÍSTICA MULTIVARIANTE
Variable RESULTADO	Cuantitativa	Cualitativa (dicotómica)
Variables PREDICTORAS	Cualitativa Cuantitativa	Cualitativa Cuantitativa
ECUACIÓN	$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$	$\text{Log} \frac{P}{1 - P} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$

## Regresión lineal múltiple

Se utiliza la regresión lineal múltiple cuando estudiamos la posible relación entre **varias variables predictoras**  $X_i$  (independientes o explicativas) y **una variable resultado** (respuesta, dependiente, explicada) que es siempre de tipo numérico (cuantitativa).

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

# Regresión lineal múltiple



## Utilidades de la regresión lineal múltiple

- Predicción de respuestas (variable resultado) a partir de un conjunto de variables explicativas (predictoras).
- Identificación de variables predictoras y construcción del modelo de regresión lineal múltiple.
- Detección de interacciones entre las distintas variables predictoras que afectan a la variable resultado.
- Identificación (y ajuste) de variables predictoras que actúan como **FACTORES de CONFUSIÓN**.

## Requisitos para el uso de la regresión lineal múltiple

### 1. Linealidad.

La variable RESULTADO depende linealmente de las variables PREDICTORAS.

Si encontramos respuestas no-lineales, esas variables predictoras se pueden transformar en lineales.

### 2. Normalidad de las variables cuantitativas.

## Modelo de la regresión lineal múltiple

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

- Y es la **variable resultado** o dependiente (debe ser siempre numérica).
- Los términos  $X_i$  son las **variables predictoras** o independientes o explicativas, las cuales pueden ser
  - Cuantitativas (discretas o continuas).
  - Cualitativas dicotómicas; en este caso deben transformarse en cuantitativas ( 0,1 ).
  - Cualitativas policotómicas; en este caso deben transformarse en cuantitativas ( 0,1 ), son las variables dummy.
- Los términos  $\beta_i$  son los **coeficientes de regresión** del modelo y los calcula el programa estadístico.

## Pregunta de investigación

¿Existe relación entre la intensidad del estudio y el rendimiento académico en el examen del curso de Bioestadística?

Variable PREDICTORA I  
Cuantitativa continua

Horas de estudio



Variable RESULTADO  
Cuantitativa discreta

Rendimiento académico

¿Cuál es el test estadístico más acertado?

Modelo de regresión lineal simple para la variable resultado número de RESPUESTAS correctas del examen ajustado por la variable predictora número de horas de ESTUDIO

## Resumen del modelo

Modelo	R	R <sup>2</sup>	R <sup>2</sup> corregido	Error estándar de la estimación
Regresión	0,629	0,396	0,321	6,506

## ANOVA

Modelo	Suma de cuadrados	g.l	Media Cuadrática	F	p
Regresión	222,234	1	222.234	5,250	0,051
Residual	338,666	8	42,333		
Total	560,900	9			

## Coeficientes

Variables	Coeficiente $\beta$	$\epsilon\epsilon\beta$	t	p
Constante	9,662	2,496	3,871	0,005
➔ Horas de estudio	0,167	0,073	2,291	0,051

# Análisis multivariante

Variable predictora 1: Cuantitativa

Horas de estudio



Variable resultado:  
Cuantitativa

Interacción



Rendimiento académico

Variable predictora 2: Cuantitativa

Coficiente intelectual



## Pregunta de investigación

¿Existe relación entre el número de horas de estudio, el coeficiente intelectual del alumno y el rendimiento académico en el examen del curso de Bioestadística?

Nº del Alumno	Horas de estudio ( $X_1$ )	CI del alumno estudio ( $X_2$ )	Nº respuestas correctas (Y)
1º	5	104	6
2º	9	125	16
3º	17	115	18
4º	1	100	1
5º	2	100	3
6º	21	140	21
7º	3	112	7
8º	29	130	20
9º	7	125	15
10º	100	122	22

¿Cuál es el test estadístico más acertado?

Modelo de regresión lineal MÚLTIPLE para la variable resultado número de RESPUESTAS correctas del examen ajustado por las variables predictoras número de horas de ESTUDIO y coeficiente intelectual CI

## Resumen del modelo

Modelo	R	R <sup>2</sup>	R <sup>2</sup> corregido	Error estándar de la estimación
Regresión	0,948	0,899	0,870	2,842

## ANOVA

Modelo	Suma de cuadrados	g.l	Media Cuadrática	F	p
Regresión	504,363	2	252,182	31,224	0,000
Residual	56,537	7	8,077		
Total	560,900	9			

## Coeficientes

Variables	Coefficiente $\beta$	$\epsilon\epsilon\beta$	t	p
Constante	- 41,153	8,667	- 4,748	0,002
➔ Horas de estudio	0,096	0,034	2,836	0,001
➔ Coeficiente intelectual	0,445	0,075	5,910	0,0025

## Mito I: “La sonrisa de la esfinge”

Para la realización del modelo de la regresión lineal múltiple es necesario que las variables incluidas cumplan la normalidad.

¿Qué test de normalidad has utilizado?

Con independencia de la normalidad, los test paramétricos NO se recomiendan cuando el tamaño muestras es pequeño ( $n \leq 10$ ), y/o cuando la variables cuantitativa es discreta.

¿Estás seguro que se cumplen estas condiciones?

## Realidad I: “Respuesta de Edipo”

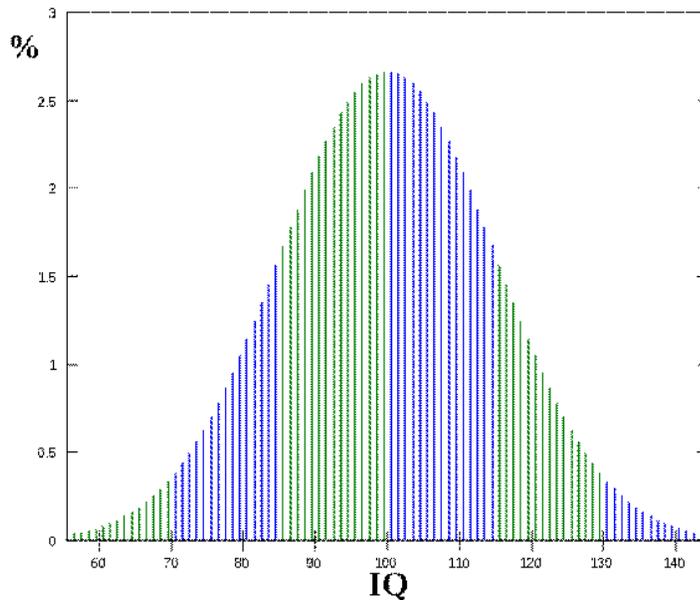
Test de Kolmogorov-Smirnov			
Variables	Estadístico	N	Significación
Horas de estudio	0,881	10	0,420
Coficiente intelectual	0,648	10	0,795
Rendimiento académico	0,439	10	0,991

$H_0$ :  $p > 0,05$  test paramétricos

$H_a$ :  $p < 0,05$  test NO paramétricos

El test de normalidad ensayado da un valor de  $p$  NO significativo ( $p > 0,05$ ) en las 3 variables evaluadas. Por tanto no hay evidencias suficientes para rechazar la hipótesis nula de normalidad. Podemos utilizar test paramétricos.

## Realidad I: “Respuesta de Edipo”



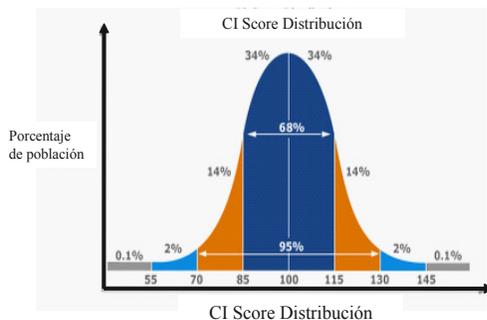
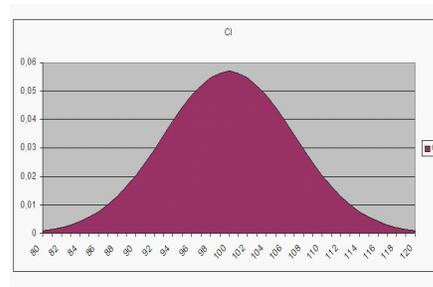
Distribución paramétrica del coeficiente intelectual en la población

# Realidad I: “Respuesta de Edipo”

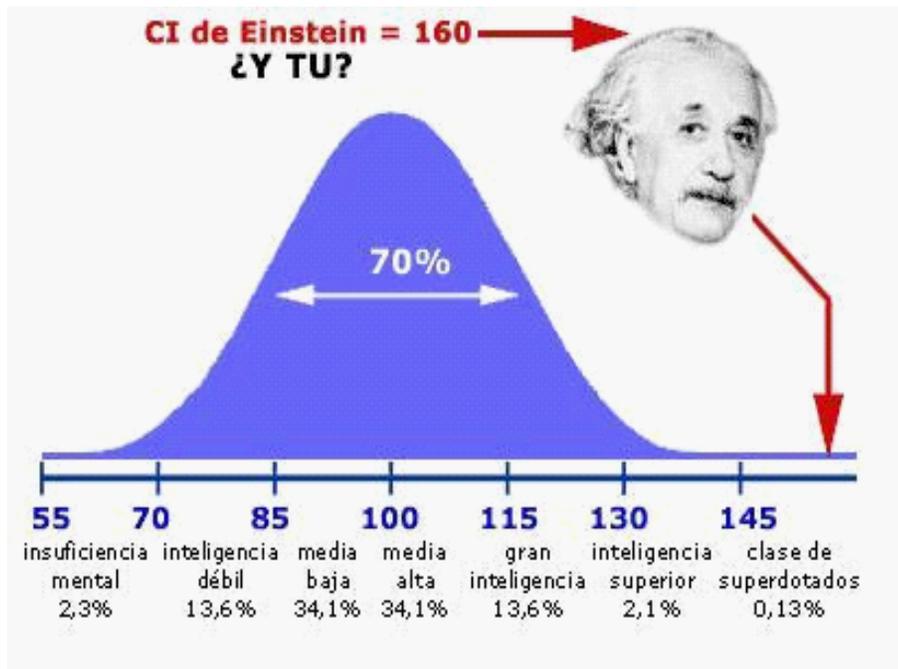
## CURVA NORMAL DE IQ (COEFICIENTE INTELECTUAL)



Standard Deviations	-4	-3	-2	-1	0	1	2	3	4
Wechsler IQ	40	55	70	85	100	115	130	145	160
Stanford-Binet IQ	36	52	68	84	100	116	132	148	164



Distribución paramétrica del coeficiente intelectual en la población



Distribución paramétrica del coeficiente intelectual en la población

## Mito I: “La sonrisa de la esfinge”

Para la aplicación del modelo de la regresión lineal múltiple es necesario que **NO** exista **COLINEALIDAD** entre las variables predictoras.

Variable PREDICTORA 1  
Cuantitativa continua

Horas de estudio



Variable PREDICTORA 2  
Cuantitativa discreta

Coficiente intelectual

¿Existe COLINEALIDAD?

## Colinealidad en la regresión lineal múltiple

- Es uno de los problemas más importantes que se puede encontrar en un análisis de regresión. Si en un modelo de Regresión Lineal Múltiple alguna variable predictora es combinación lineal de otras variables predictoras, el modelo es irresoluble.
- La colinealidad aparece cuando alguna variable predictora es “casi” combinación lineal de otra u otras variables predictoras, es decir sus coeficientes de correlación simple o múltiple entre las variables predictoras son cercanos a 1, aunque no llegan a dicho valor.
- Como medida de la misma hay varios estadísticos propuestos, los más sencillos son los coeficientes de determinación ( $r^2$ ) de cada variable predictoras con todas las demás.

## Diagnóstico de colinealidad (I):

- FACTOR DE INFLACIÓN DE LA VARIANZA (FIV)

FIV > 10: Colinealidad (\*)

$$FIV = \frac{1}{1 - R^2_j}$$

- TOLERANCIA (T)

T < 0,1: Colinealidad (\*)

$$T = \frac{1}{FIV} = 1 - R^2_j$$

(\*) Regla empírica de Kleinbaum

## Diagnóstico de colinealidad (I):

Modelo	R	R <sup>2</sup>	R <sup>2</sup> corregido	Error estándar de la estimación
Regresión	0,948	0,899	0,870	2,842

Variable resultado: Número de respuestas correctas

$$FIV = \frac{1}{1 - R^2_j}$$

$$FIV_{(\text{estudio-CI})} = 9,90$$

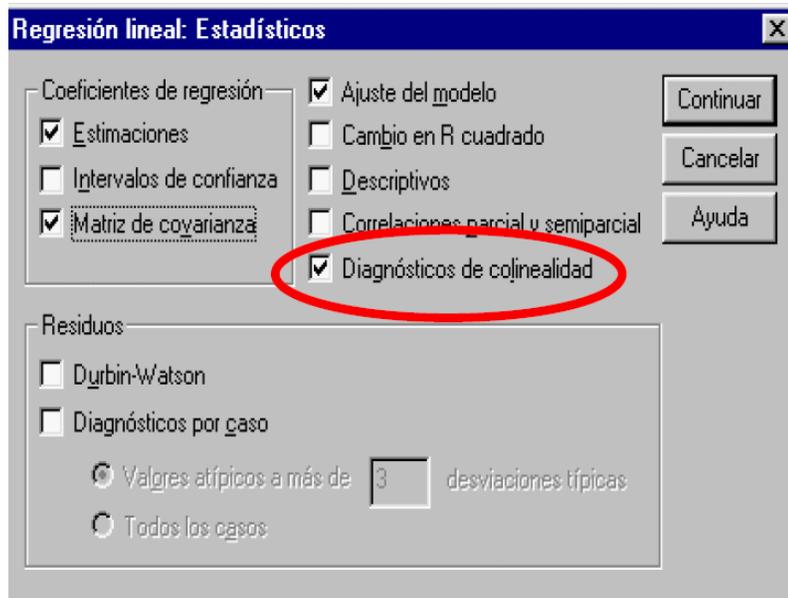
$$T = \frac{1}{FIV} = 1 - R^2_j$$

$$T_{(\text{estudio-CI})} = 0,101$$

NO hay colinealidad

## Diagnóstico de colinealidad (2) SPSS

Ruta con SPSS: Opción: Análisis, Regresión lineal, Estadísticos...



## Diagnóstico de colinealidad (2)

ÍNDICE DE CONDICIÓN: Regla empírica (Belsley):

- Índices de condición  $< 5$  : NO existe colinealidad.
- Índices de condición entre 5 y 10: colinealidad débil.
- Índices de condición entre 10 y 30: colinealidad moderada.
- Índices de condición  $> 30$  colinealidad fuerte (grave).

Modelo	Dimensión	Autovalor	Índice de condición	Proporciones de la varianza		
				Constante	Horas de Estudio	CI
1	1	2,456	1,000	0,00	0,06	0,00
	2	0,536	2,136	0,00	0,84	0,00
	3	0,005	21,608	1,00	0,10	1,00

Variable resultado: Número de respuestas correctas

## Diagnóstico de colinealidad (2)

- Hay un índice de condición alto (21,6) y asociado con el mismo está el término constante y el coeficiente intelectual (CI) con proporción de varianza alta.

- Nos indica, por tanto, que el CI sólo es colineal con la constante, pero no con las horas de estudio.

Modelo	Dimensión	Autovalor	Índice de condición	Proporciones de la varianza		
				Constante	Horas de Estudio	CI
1	1	2,456	1,000	0,00	0,06	0,00
	2	0,536	2,136	0,00	0,84	0,00
	3	0,005	21,608	1,00	0,10	1,00

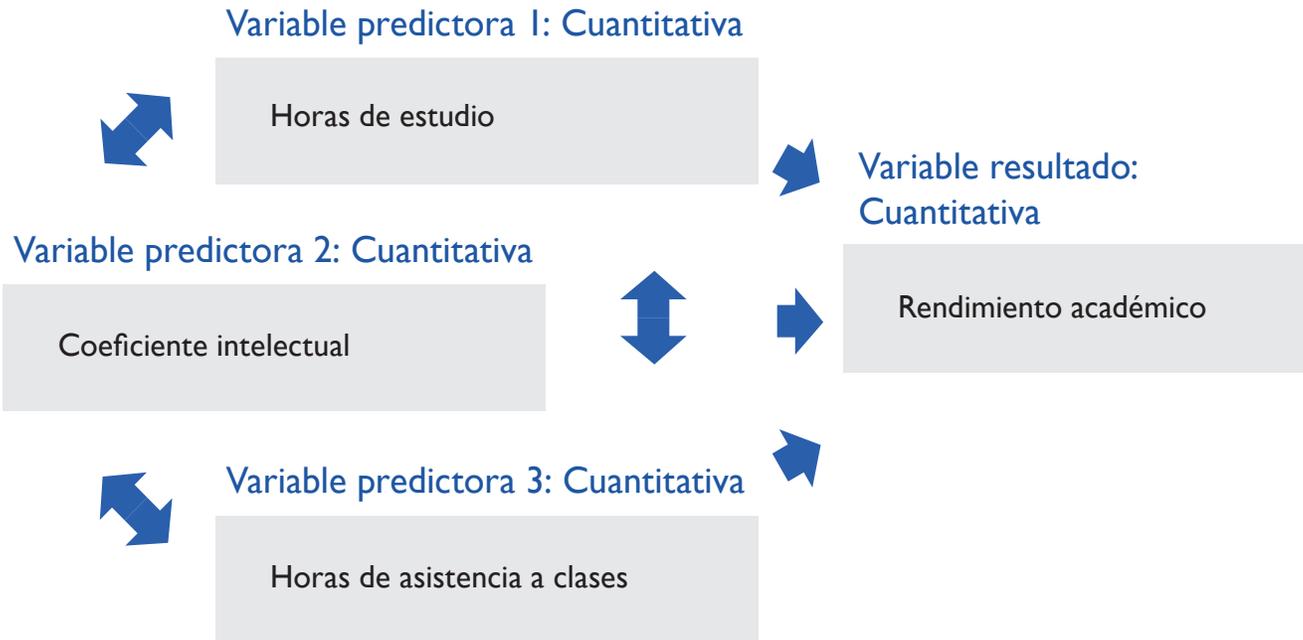
Variable resultado: Número de respuestas correctas

NO hay colinealidad

## Pregunta de investigación

¿Existe relación entre el número de horas de estudio, el coeficiente intelectual del alumno, asistencia a clases (horas) y el rendimiento académico en el examen del curso de Bioestadística?

Nº del Alumno	Horas de estudio ( $X_1$ )	CI del alumno ( $X_2$ )	Horas de Asistencia a clase ( $X_3$ )	Nº respuestas correctas (Y)
1º	5	104	23	6
2º	9	125	92	16
3º	17	115	78	18
4º	1	100	15	1
5º	2	100	20	3
6º	21	140	100	21
7º	3	112	35	7
8º	29	130	95	20
9º	7	125	75	15
10º	100	122	100	22



Modelo de regresión lineal MÚLTIPLE para la variable resultado número de RESPUESTAS correctas del examen ajustado por las variables predictoras número de horas de ESTUDIO, coeficiente intelectual CI y asistencia a clase

Resumen del modelo				
Modelo	R	R <sup>2</sup>	R <sup>2</sup> corregido	Error estándar de la estimación
Regresión	0,984	0,969	0,954	1,696

ANOVA					
Modelo	Suma de cuadrados	g.l	Media Cuadrática	F	p
Regresión	543,632	3	181,211	62,963	0,000
Residual	17,268	6	8,077		
Total	560,900	9			

#### Coefficientes

Variables	Coefficiente $\beta$	$\epsilon\beta$	t	p
Constante	- 2,593	11,651	0,0223	0,831
Horas de estudio	0,028	0,104	1,007	0,353
Coeficiente intelectual	0,022	0,123	0,177	0,865
➔ Asistencia a clase (horas)	0,196	0,053	3,694	0,010

Variables	Estadísticos de colinealidad	
	Tolerancia (T)	FIV
Horas de estudio	0,477	2,098
Coeficiente intelectual	0,117	8,552
Asistencia a clase (horas)	0,89	11,241

Obsérvese que los mayores valores de la tolerancia (T) y del factor de inflación corresponden a la variable predictora ASISTENCIA a clase.

- FACTOR DE INFLACIÓN DE LA VARIANZA (FIV)

FIV > 10: Colinealidad (\*)

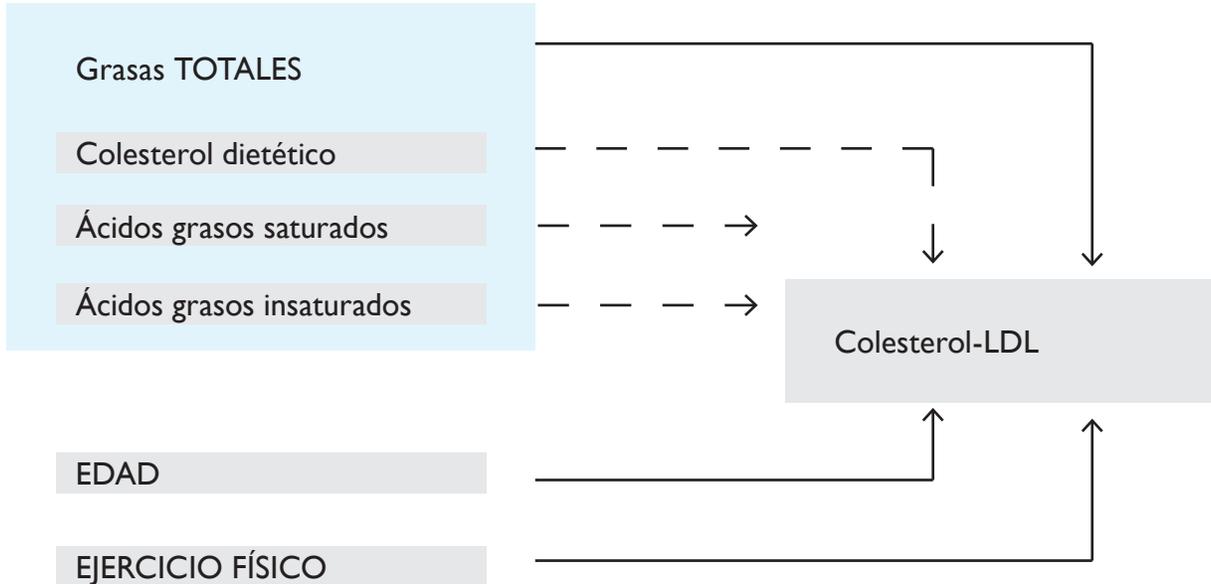
- TOLERANCIA (T)

T < 0,1: Colinealidad (\*)

Autovalor	Índice de condición	Proporciones de la varianza			
		Constante	Horas de ESTUDIO	Coefficiente intelectual	ASISTENCIA a clases
3.366	1,000	0,00	0,02	0,00	0,00
0,538	2,501	0,00	0,45	0,00	0,00
0,095	5.957	0,00	0,24	0,00	0,14
0,001	64,725	0,99	0,29	1,00	0,85

Variable resultado: Rendimiento académico (respuestas acertadas)

Hay un índice de condición alto (64,72) y asociado con el mismo hay dos variables (el término constante, coeficiente intelectual, y ASISTENCIA a clase) con proporción de varianza alta. Nos indica, por tanto, que la asistencia a clase es COLINEAL con el coeficiente intelectual y con la constante.



Realizando los análisis de regresión de cada una de las variables predictoras del modelo con todas las demás, se obtienen los siguientes coeficientes de determinación, tolerancia y factores de inflación de la varianza que indican graves problema de COLINEALIDAD.

Modelo	Coeficientes no estandarizados		t	Significación	Estadísticos de colinealidad	
	$\beta$	$EE\beta$			Tolerancia	FIV
Constante	258,484	130,837	1,976	0,70		
Grasa TOTAL	- 5,714	4,776	-1,196	0,253	0,20	49,978
Colesterol dieta	0,067	0,036	1,851	0,087	0,038	26,530
Edad	0,443	2,867	0,155	0,879	0,098	10,225
Ejercicio	- 3,115	45,953	- 0,068	0,947	0,143	7,005
Ácidos grasos saturados	0,0722	0,075	0,966	0,351	0,066	15,242
Ácidos grasos Insaturados	- 0,0746	1,150	- 0,065	0,949	0,130	7,683
Variable resultado colesterol-LDL						

Obsérvese que el factor de inflación mayor corresponde a la variable grasas TOTAL, resultado esperable debido a que se han creado tres variables más (colesterol dietético, ácidos grasos saturados y ácidos grasos insaturados) a partir de ella.

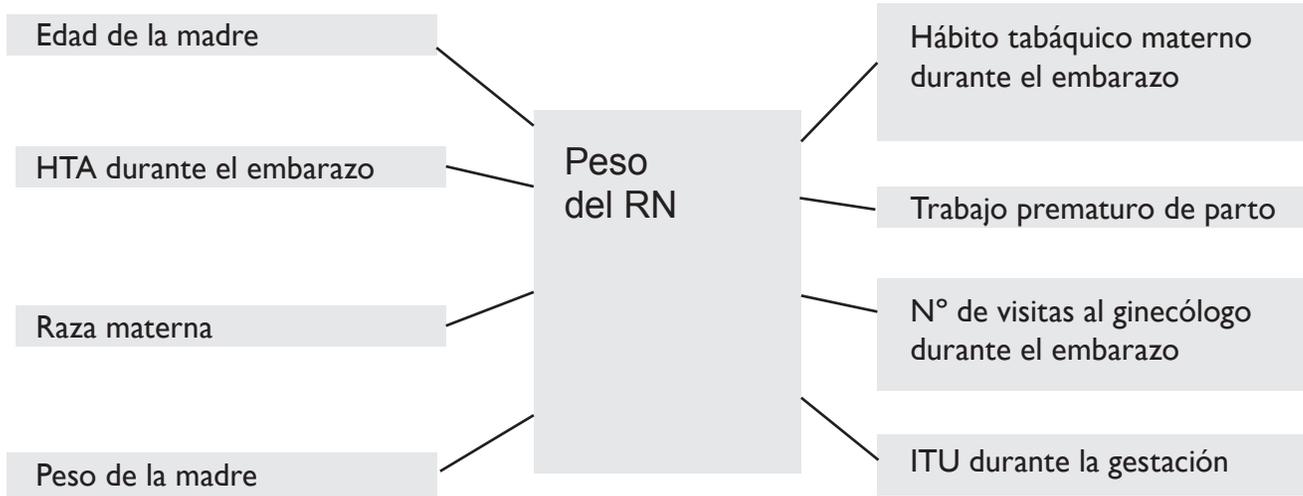
Autovalor	Índice de condición	Proporciones de la varianza						
		Constante	Edad	Grasas totales	Ejercicio	Colesterol dietético	Ácidos grasos insaturados	Ácidos grasos saturados
5,276	1,000	0,00	0,00	0,00	0,00	0,00	0,00	0,00
1,055	2,236	0,00	0,00	0,00	0,03	0,00	0,03	0,00
0,519	3,190	0,00	0,01	0,00	0,02	0,02	0,00	0,00
0,100	7,250	0,03	0,01	0,00	0,06	0,01	0,20	0,04
0,0354	12,200	0,06	0,00	0,00	0,78	0,03	0,64	0,04
0,01207	20,908	0,02	0,41	0,05	0,02	0,31	0,05	0,23
0,00205	50,781	0,89	0,58	0,95	0,09	0,64	0,07	0,69

Variable resultado: Colesterol-LDL

Autovalor	Índice de condición	Proporciones de la varianza						
		Constante	Edad	Grasas totales	Ejercicio	Colesterol dietético	Ácidos grasos insaturados	Ácidos grasos saturados
5,276	1,000	0,00	0,00	0,00	0,00	0,00	0,00	0,00
1,055	2,236	0,00	0,00	0,00	0,03	0,00	0,03	0,00
0,519	3,190	0,00	0,01	0,00	0,02	0,02	0,00	0,00
0,100	7,250	0,03	0,01	0,00	0,06	0,01	0,20	0,04
0,0354	12,200	0,06	0,00	0,00	0,78	0,03	0,64	0,04
0,01207	20,908	0,02	0,41	0,05	0,02	0,31	0,05	0,23
0,00205	50,781	0,89	0,58	0,95	0,09	0,64	0,07	0,69

Variable resultado: Colesterol-LDL

Hay un índice de condición alto (50,781) y asociado con el mismo hay cinco variables (el término constante, EDAD, GRASAS TOTALES, COLESTEROL dietético, y ácidos grasos SATURADOS) con proporción de varianza alta. Nos indica, por tanto, que Grasas TOTALES es COLINEAL con el COLESTEROL dietético, ácidos grasos SATURADOS, EDAD y con la constante. Al existir esta última colinealidad, el centrado de variables podría mejorar el problema, se podría también renunciar a estudiar los términos no lineales COLESTEROL dietético y ácidos grasos SATURADOS.



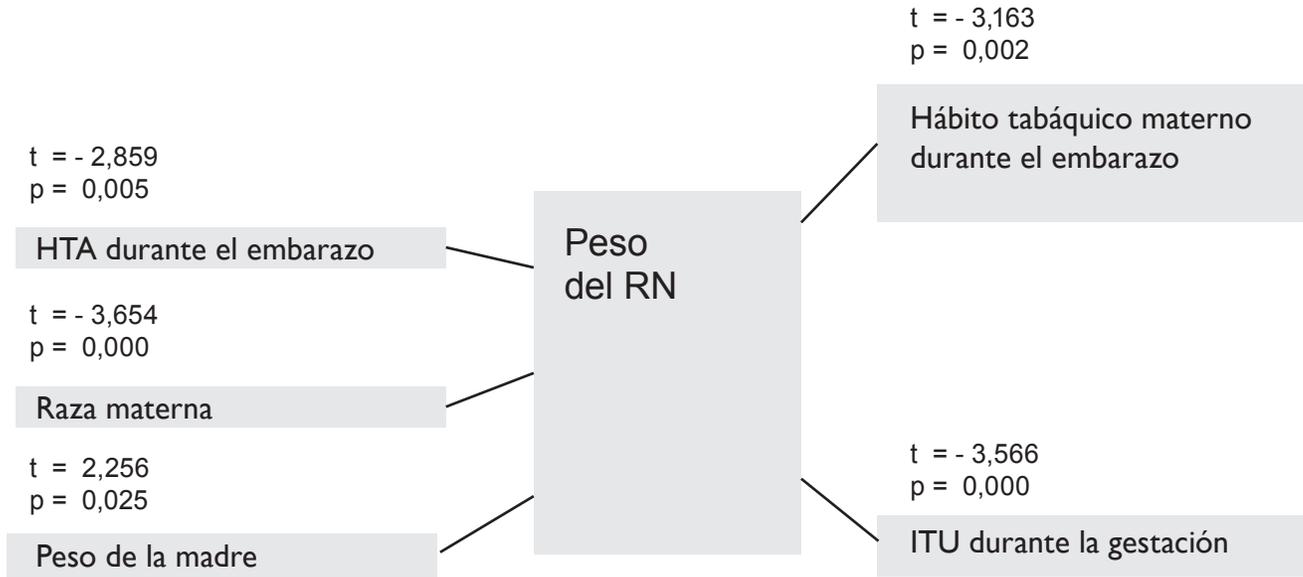
# Interpretación práctica de los resultados de la regresión lineal múltiple

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	2975,284	315,998		9,416	,000	2351,748	3598,820
	Edad de la madre	-1,919	9,544	-,014	-,201	,841	-20,752	16,914
	➤ Hábito tabáquico materno	-338,413	106,977	-,227	-3,163	➤ ,002	-549,503	-127,322
	➤ Hipertensión durante el embarazo	-573,555	200,591	-,192	-2,859	➤ ,005	-969,369	-177,742
	➤ Infección urinaria en la gestación	-488,353	136,959	-,239	-3,566	➤ ,000	-758,604	-218,102
	➤ Raza materna (blanca vs otras)	-380,931	104,245	-,262	-3,654	➤ ,000	-586,630	-175,232
	➤ Peso de la madre en Kg	8,055	3,570	,155	2,256	➤ ,025	1,010	15,100
	Nº de visitas en el embarazo categorizada	-11,248	99,741	-,008	-,113	,910	-208,059	185,563
	Trabajo prematuro de parto	-203,230	136,333	-,102	-1,491	,138	-472,247	65,787

a. Dependent Variable: Peso del RN

Modelo de regresión logístico para el BAJO PESO AL NACER				
VARIABLES predictoras	Estimación $\beta$	Error estándar $EE\beta$	$\beta/EE\beta$ t	p
Constante	2975,28	315,998	9,416	0,000
Edad de la madre	-1,919	9,544	-0,201	0,841
Hábito tabáquico materno	◀ -338,413	106,977	-3,163	0,002
Hipertensión durante el embarazo	◀ -573,555	200,591	-2,859	0,005
Infección urinaria durante el embarazo	◀ 488,353	136,959	-3,566	0,000
Raza materna (blanca versus otras)	◀ -380,931	104,245	-3,654	0,000
Peso de la madre en Kg	◀ 8,055	3,570	2,256	0,025
Nº visitas al ginecólogo categorizada	-11,248	99,741	-0,113	0,910
Trabajo prematuro de parto	-203,230	136,333	-1,491	0,138



## Interpretación práctica de los resultados de la regresión lineal múltiple

- La aceptación de la  $H_0$  ( $p > 0,05$ ) implica que la variable resultado  $Y$  no está influenciada por las variables predictoras  $X_i$  incluidas en el modelo.
- La **aceptación de la  $H_a$**  ( $p < 0,05$ ) supone que hay algún tipo de influencia de algunas de las variables predictoras  $X_i$  sobre la variable resultado.
- La significación estadística se realiza mediante el test de contraste de hipótesis del análisis de la varianza (ANOVA).

ANOVA<sup>a</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	24747080	8	3093384,997	7,407	,000 <sup>a</sup>
	Residual	75169973	180	417610,959		
	Total	99917053	188			

← Significación estadística del modelo de regresión

a. Predictors: (Constant), Trabajo prematuro de parto, Hipertensión durante el embarazo, Raza materna (blanca vs otras), N° de visitas en el embarazo categorizada, Infección urinaria en la gestación, Peso de la madre en Kg, Edad de la madre, Hábito tabáquico materno

b. Dependent Variable: Peso del RN

## Interpretación práctica de los resultados de la regresión lineal múltiple

- Los programas estadísticos como el SPSS ofrecen una estimación de los distintos coeficientes de regresión  $\beta_i$ , junto al:
  - Error estándar de los coeficientes de regresión  $EE\beta_i$ .
  - Significación estadística (valor de p) de  $\beta_i$  cada variable predictora.
  - Un intervalo de confianza (IC95%) de los coeficientes de regresión  $\beta_i$ .
- Las variables predictoras con significación estadística ( $p < 0,05$ ) deben quedar en el modelo multivariante y en la ecuación que predice a la variable resultado.
- Las variables predictoras sin significación estadística ( $p > 0,1$ ) deben ser excluidas del modelo final.

$\beta_i$  $EE\beta_i$ Coefficients<sup>a</sup>IC95% $\beta_i$ 

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	2975,284	315,998		9,416	,000	2351,748	3598,820
	Edad de la madre	-1,919	9,544	-,014	-,201	,841	-20,752	16,914
	Hábito tabáquico materno	-338,413	106,977	-,227	-3,163	,002	-549,503	-127,322
	Hipertensión durante el embarazo	-573,555	200,591	-,192	-2,859	,005	-969,369	-177,742
	Infeccion urinaria en la gestación	-488,353	136,959	-,239	-3,566	,000	-758,604	-218,102
	Raza materna (blanca vs otras)	-380,931	104,245	-,262	-3,654	,000	-586,630	-175,232
	Peso de la madre en Kg	8,055	3,570	,155	2,256	,025	1,010	15,100
	Nº de visitas en el embarazo categorizada	-11,248	99,741	-,008	-,113	,910	-208,059	185,563
	Trabajo prematuro de parto	-203,230	136,333	-,102	-1,491	,138	-472,247	65,787

a. Dependent Variable: Peso del RN

## Interpretación práctica de los resultados de la regresión lineal múltiple

- El término estadístico R cuadrado ( $R^2$ ) es el porcentaje de reducción de la incertidumbre sobre la variabilidad de la variable RESULTADO (respuesta) cuando son conocidas las variables PREDICTORAS. Cuanto más se acerque a 1 más poder explicativo tiene el modelo.
- La R cuadrado “corregida” es un resultado ajustado, que penaliza la introducción de variables innecesarias en el modelo, es decir, disminuye su valor cuando se introducen variables que no actúan sobre la variable resultado.

$R^2$  (R Square o bondad del ajuste)

Model Summary

$R^2$  corregida (ajustada)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,498 <sup>a</sup>	,248	,214	646,228

- a. Predictors: (Constant), Trabajo prematuro de parto, Hipertensión durante el embarazo, Raza materna (blanca vs otras), N° de visitas en el embarazo categorizada, Infección urinaria en la gestación, Peso de la madre en Kg, Edad de la madre, Hábito tabáquico materno

# Interpretación práctica de los resultados de la regresión lineal múltiple

- La tabla o matriz de correlaciones es de gran interés para identificar correlaciones lineales entre parejas de variables cuantitativas.
  - Interesante si ocurre entre la variable resultado y alguna de las variables predictoras.
  - Problemático si ocurre entre dos variables predictoras porque indica COLINEALIDAD.
- La matriz está formada por todos los coeficientes de correlación lineal de Pearson (r) para cada par de variables cuantitativas del estudio.

Coefficient Correlations<sup>a</sup>

Model		Trabajo prematuro de parto	Hipertensión durante el embarazo	Raza materna (blanca vs otras)	Nº de visitas en el embarazo categorizada	Infección urinaria en la gestación	Peso de la madre en Kg	Edad de la madre	Hábito tabáquico materno
1	Correlations	1,000	-.050	-.126	.074	-.174	.113	-.150	-.218
	Trabajo prematuro de parto	1,000							
	Hipertensión durante el embarazo	-.050	1,000	-.053	-.065	.092	-.240	.043	-.025
	Raza materna (blanca vs otras)	-.126	-.053	1,000	-.164	-.014	.045	.192	.370
	Nº de visitas en el embarazo categorizada	.074	-.065	-.164	1,000	-.042	.028	.165	-.199
	Infección urinaria en la gestación	-.174	.092	-.014	-.042	1,000	.091	.061	-.020
	Peso de la madre en Kg	.113	-.240	.045	.028	.091	1,000	-.175	.023
	Edad de la madre	-.150	.043	.192	.165	.061	-.170	1,000	.094
	Hábito tabáquico materno	-.218	-.025	.370	-.199	-.020	.023	.094	1,000
	Covariances								
	Trabajo prematuro de parto	18586,737	-1365,589	-1791,262	1009,836	-3254,760	55,082	-195,718	-3181,035
	Hipertensión durante el embarazo	-1365,589	40236,901	-1118,575	-1306,095	2519,927	-171,932	81,982	-525,956
	Raza materna (blanca vs otras)	-1791,262	-1118,575	10866,947	-1707,605	-200,116	16,866	190,755	4124,674
	Nº de visitas en el embarazo categorizada	1009,836	-1306,095	-1707,605	9948,197	-574,052	9,929	156,628	-2124,941
	Infección urinaria en la gestación	-3254,760	2519,927	-200,116	-574,052	18757,679	44,682	80,006	-292,270
	Peso de la madre en Kg	55,082	-171,932	16,866	9,929	44,682	12,745	-5,795	8,723
	Edad de la madre	-195,718	81,982	190,755	156,628	80,006	-5,795	91,092	95,883
	Hábito tabáquico materno	-3181,035	-525,956	4124,674	-2124,941	-292,270	8,723	95,883	11444,113

a. Dependent Variable: Peso del RN

## Mito I: “La sonrisa de la esfinge”

En la regresión lineal múltiple realizado se han incluido inicialmente 8 posibles variables predictoras de la variable resultado (peso del RN). De ellas sólo 5 han quedado en el modelo final.

¿Estás seguro que se ha realizado una completa y actualizada revisión bibliográfica?

¿Por qué no se ha tenido en cuenta en el modelo de regresión la variable DIABETES GESTACIONAL que es ampliamente conocido que aumenta el peso del RN causando macrosomías?

## Estudio transversal en pacientes con diabetes mellitus

Objetivo: Evaluar la influencia de determinadas variables en el control metabólico de la diabetes mellitus tipo 2.

Variable resultado: Niveles de hemoglobina glicosilada

Variable predictora [1]: Tipo de tratamiento

Variable predictora [2]: Sexo del paciente

Variable predictora [3]: Años desde el diagnóstico de la enfermedad

Variable predictora [4]: Número de enfermedades relacionadas con la diabetes

Variable predictora [5]: Acceso al dietista

Variable predictora [6]: Médico interesado por la diabetes (actitud y conocimiento)

Variable predictora [7]: Nivel socioeconómico

Variable predictora [8]: Estilos de vida (tabaco, alcohol...)

Variable predictora [9]: Edad del médico

Variable predictora [10]: Sexo del médico

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

## Análisis bivalente:

¿Existe relación entre el tipo de tratamiento y los niveles de hemoglobina glicosilada?

Variable PREDICTORA I  
Cualitativa policotómica

Tipo de tratamiento  
- Dieta  
- Insulina  
- ADO



Variable RESULTADO  
Cuantitativa continua

%HbA1c

¿Cuál es el test estadístico más acertado?

Análisis de la Varianza para un factor (ANOVA)  $p < 0,05$

## Análisis bivariante:

¿Existe relación entre el sexo del paciente y los niveles de hemoglobina glicosilada?

Variable PREDICTORA 2  
Cualitativa dicotómica

Sexo del paciente



Variable RESULTADO  
Cuantitativa continua

%HbA1c

¿Cuál es el test estadístico más acertado?

T de Student para muestras independientes  $p < 0,05$

## Análisis bivalente:

¿Existe relación entre los años desde el diagnóstico de la enfermedad y los niveles de HbA1c?

Variable PREDICTORA 3  
Cuantitativa continua

Años desde el diagnóstico  
de la enfermedad



Variable RESULTADO  
Cuantitativa continua

%HbA1c

¿Cuál es el test estadístico más acertado?

Coefficiente de correlación de Pearson  $p < 0,05$

## Análisis bivariante:

¿Existe relación entre los años desde el diagnóstico de la enfermedad y los niveles de HbA1c?

Variable PREDICTORA 3  
Cualitativa dicotómica

Años desde el diagnóstico  
de la enfermedad  
> 7 años  
< 7 años



Variable RESULTADO  
Cuantitativa continua

%HbA1c

¿Cuál es el test estadístico más acertado?

T de Student para muestras independientes  $p < 0,05$

## Análisis bivariante:

¿Existe relación entre el número de enfermedades relacionadas con la diabetes y los niveles de HbA1c?

Variable PREDICTORA 4  
Cuantitativa discreta

Número de enfermedades  
relacionadas con la DM



Variable RESULTADO  
Cuantitativa continua

%HbA1c

¿Cuál es el test estadístico más acertado?

Coefficiente de correlación de Spearman  $p < 0,05$

## Análisis bivariante:

¿Existe relación entre el número de enfermedades relacionadas con la diabetes y los niveles de HbA1c?

Variable PREDICTORA 4  
Cualitativa dicotómica

Número de enfermedades relacionadas con la DM  
- Más de una  
- Una o menos de una



Variable RESULTADO  
Cuantitativa continua

%HbA1c

¿Cuál es el test estadístico más acertado?

T de Student para muestras independientes  $p < 0,05$

## Análisis bivalente:

¿Existe relación entre el acceso al dietista y los niveles de hemoglobina glicosilada?

Variable PREDICTORA 5  
Cualitativa dicotómica

Acceso al DIETISTA



Variable RESULTADO  
Cuantitativa continua

%HbA1c

¿Cuál es el test estadístico más acertado?

T de Student para muestras independientes  $p < 0,05$

## Análisis bivalente:

¿Existe relación entre la actitud y conocimiento del médico por la DM y los niveles de HbA1c?

Variable PREDICTORA 6  
Cualitativa dicotómica

Médico interesado  
por la DM  
- Si  
- No



Variable RESULTADO  
Cuantitativa continua

%HbA1c

¿Cuál es el test estadístico más acertado?

T de Student para muestras independientes  $p < 0,05$

## Análisis bivalente:

¿Existe relación entre el nivel socioeconómico del paciente y los niveles de HbA1c?

Variable PREDICTORA 7  
Cualitativa policotómica

Nivel socioeconómico



Variable RESULTADO  
Cuantitativa continua

%HbA1c

¿Cuál es el test estadístico más acertado?

Análisis de la Varianza para un factor (ANOVA)  $p > 0,1$

## Análisis bivalente:

¿Existe relación entre el estilo del vida del paciente y los niveles de hemoglobina glicosilada?

Variable PREDICTORA 8  
Cualitativa dicotómica

Estilos de vida  
- Adecuados  
- Inadecuados  
(tabaco, alcohol, etc)



Variable RESULTADO  
Cuantitativa continua

%HbA1c

¿Cuál es el test estadístico más acertado?

T de Student para muestras independientes  $p > 0,1$

## Análisis bivalente:

¿Existe relación entre el sexo del médico y los niveles de hemoglobina glicosilada?

Variable PREDICTORA 9  
Cualitativa dicotómica

Sexo del Médico



Variable RESULTADO  
Cuantitativa continua

%HbA1c

¿Cuál es el test estadístico más acertado?

T de Student para muestras independientes  $p > 0,1$

## Análisis bivariante:

¿Existe relación entre el tipo de tratamiento y los niveles de hemoglobina glicosilada?

Variable PREDICTORA I0  
Cuantitativa continua

Edad del Médico

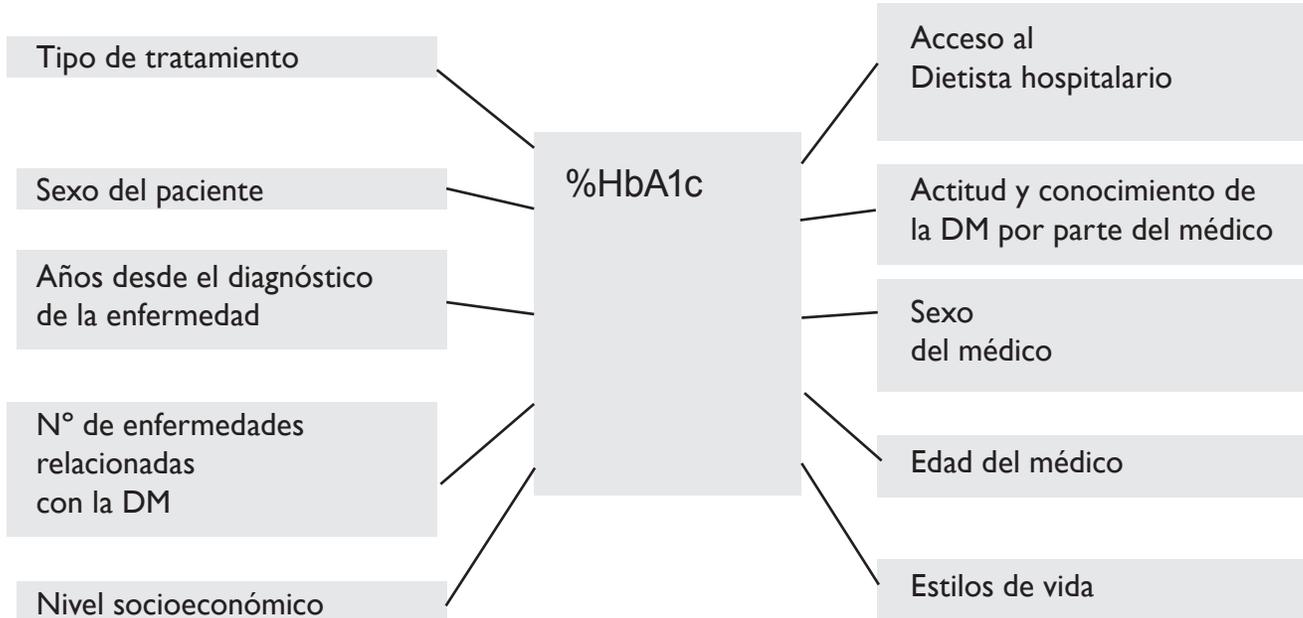


Variable RESULTADO  
Cuantitativa continua

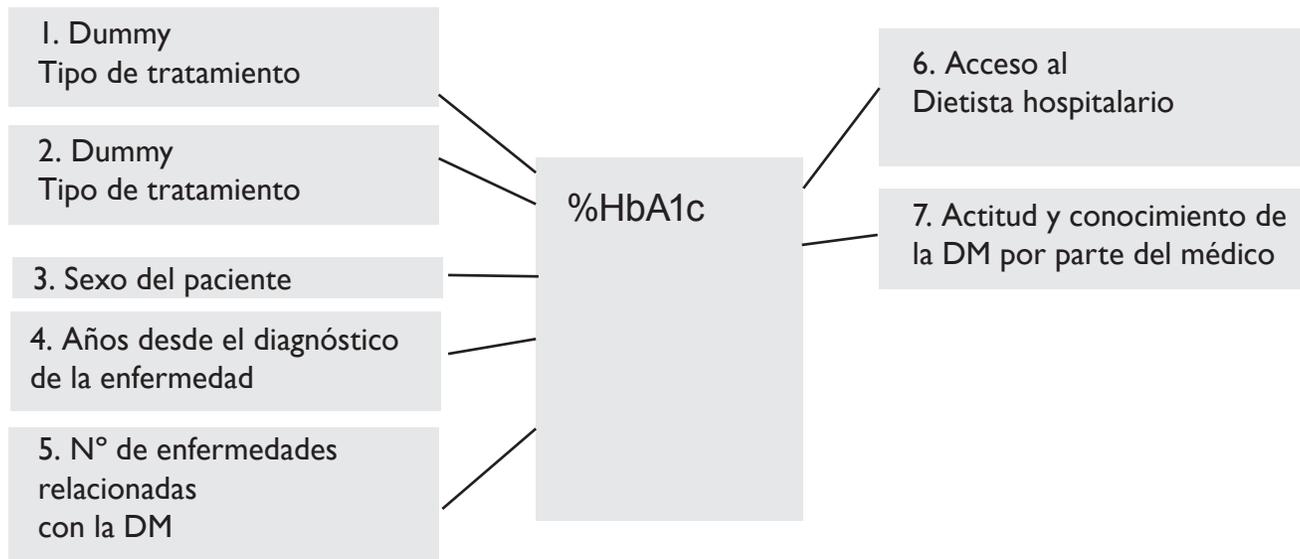
%HbA1c

¿Cuál es el test estadístico más acertado?

Coefficiente de correlación de Pearson  $p > 0,1$



## Análisis de regresión lineal múltiple



$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6$$

## Análisis de regresión lineal múltiple

### 1.- Tipo de tratamiento

Insulina

Antidiabéticos orales (ADO)

Dieta

Variables predictoras (X)

### 2.- Sexo del paciente (mujer [1]; hombre [0])

### 3.- Años desde el diagnóstico de la enfermedad

> 7 años: [1]

< 7 años: [0]

Variable resultado (Y):

HbA1c(%)

### 4.- Número de enfermedades relacionadas con la diabetes

- Más de una: [1]

- Una o menos de una: [0]

### 5.- Acceso al dietista (si [1]; no [0])

### 6.- Médico interesado por la diabetes (si [1]; no [0])

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6$$

## Análisis de regresión lineal múltiple

Variable resultado (Y):  
HbA1c(%)

Variables PREDICTORAS  
que estaban relacionadas  
estadísticamente con la  
variable resultado HbA1c  
en el análisis univariante

Coefficiente de regresión  
estandarizado  
(t calculada)

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

Variables predictoras (X)

Variables asociadas al nivel de hemoglobina glicosilada		
Variables predictoras	Estimación $\beta$	Error estándar $EE\beta$
Constante	10,51	0,59
1ª dummy Antidiabéticos orales / insulina	0,29	0,37
2ª dummy Dieta / insulina	-0,97	0,46
Nº enfermedades relacionadas con la diabetes ( $> 1 / \leq 1$ )	0,20	0,08
Sexo (mujer/hombre)	0,72	0,28
Años desde el diagnóstico ( $> 7 / \leq 7$ )	0,04	0,02
Acceso al dietista hospitalario (si/no)	-1,06	0,43
Médico interesado en la diabetes (si/no)	- 0,86	0,38

$$\frac{\beta_{\text{sexo}}}{EE\beta_{\text{sexo}}} = \frac{0,72}{0,28} = 2,57 > Z_{95\%} (1,96) \Rightarrow p < 0,05$$

## Mito I: “La sonrisa de la esfinge”

En la tabla de la regresión lineal múltiple presentada se han incluido sólo el valor de  $\beta$  y su error estándar  $EE\beta$ , pero no se ha incluido el IC95% de  $\beta$  ni la significación estadística  $p$  de cada variable predictora. Tampoco veo la tabla donde se indica el  $R^2$  (R Square o bondad del ajuste del modelo multivariante).

¿Estás seguro que las 6 variables predictoras incluidas en el modelo multivariante presentan significación estadística ( $p < 0,05$ )?

La variable 1ª dummy ADO/insulina presenta un valor de  $\beta$  (0,29) inferior a su error estándar  $EE\beta$ , (0,37) por lo que su cociente vale 0,78, cifra muy inferior al 1,96 ( $Z_{95\%}$ )

¿Por qué has incluido en el modelo de regresión la variable ADO/insulina?

## Análisis de regresión lineal múltiple

Variable resultado (Y):  
HbA1c(%)

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

Variables predictoras (X)

¿Cuál de las distintas variables predictoras incluidas en el modelo parece afectar más a la variable resultado (HbA1c)?

Variables asociadas al nivel de hemoglobina glicosilada		
Variables predictoras	Estimación $\beta$	Error estándar $EE\beta$
Constante	10,51	0,59
1ª dummy Antidiabéticos orales / insulina	0,29	0,37
2ª dummy Dieta / insulina	-0,97	0,46
Nº enfermedades relacionadas con la diabetes (> 1 / ≤ 1)	0,20	0,08
Sexo (mujer/hombre)	0,72	0,28
Años desde el diagnóstico (> 7 / ≤ 7)	0,04	0,02
Acceso al dietista hospitalario (si/no)	-1,06	0,43
Médico interesado en la diabetes (si/no)	- 0,86	0,38

$$E(HbA_1) = 10.51 + 0.29(\text{Ant. orales/insulina}) - 0.97(\text{Dieta/insulina}) + 0.20(\text{Nº enferm.}) + 0.72(\text{Sexo}) + 0.04(\text{Años diagnost.}) - 1.06(\text{Acc.dietit.}) - 0.86(\text{Interés})$$

## Análisis de regresión lineal múltiple

Variable resultado (Y):  
HbA1c(%)

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

Variables predictoras (X)

Variables asociadas al nivel de hemoglobina glicosilada				
Variables predictoras	Estimación $\beta$	Error estándar $EE\beta$	Cociente $\beta / EE\beta$	Significación $p$
Constante	10,51	0,59		
1ª dummy Antidiabéticos orales / insulina	0,29	0,37	0,78	$p > 0,1$
2ª dummy Dieta / insulina	-0,97	0,46	2,11	$p < 0,05$
Nº enfermedades relacionadas con la diabetes ( $> 1 / \leq 1$ )	0,20	0,08	2,5	$p < 0,05$
Sexo (mujer/hombre)	0,72	0,28	2,57	$p < 0,01$
Años desde el diagnóstico ( $> 7 / \leq 7$ )	0,04	0,02	2,00	$p < 0,05$
Acceso al dietista hospitalario (si/no)	-1,06	0,43	2,47	$p < 0,02$
Médico interesado en la diabetes (si/no)	- 0,86	0,38	2,26	$p < 0,05$

$$E(HbA_1) = 10.51 + 0.29(\text{Ant. orales/insulina}) - 0.97(\text{Dieta/insulina}) + 0.20(\text{Nº enferm.}) \\ + 0.72(\text{Sexo}) + 0.04(\text{Años diagnost.}) - 1.06(\text{Acc.dietit.}) - 0.86(\text{Interés})$$

¿Qué diferencia promedio de HbA1c existe entre un paciente A mujer (sexo=1) y un paciente B hombre (sexo=0) y a igualdad del resto de las variables predictoras incluidas en el modelo?

Variables predictoras	$\beta_i$	Paciente A	Paciente B
Sexo (mujer=1, hombre =0)	+0,72	mujer [1]	hombre [0]

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

$$\text{HbA1c (A)} = 10,51 + 0,72 (1) = 11,23$$

$$\text{HbA1c (B)} = 10,51 + 0,72 (0) = 10,51$$

Las mujeres presentan 0,72 unidades más de HbA1c que los hombres a igualdad del resto de las variables predictoras incluidas en el modelo.

¿Qué diferencia promedio de HbA1c existe entre un paciente A con acceso al dietista hospitalario (1) de un paciente B sin acceso (0) y a igualdad del resto de las variables predictoras incluidas en el modelo?

Variables predictoras	$\beta_i$	Paciente A	Paciente B
Acceso al dietista hospitalario (si =1, no =0)	-1,06	SI [1]	NO [0]

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

$$\text{HbA1c (A)} = 10,51 - 1,06 (1) = 9,45$$

$$\text{HbA1c (B)} = 10,51 - 1,06 (0) = 10,51$$

Los pacientes CON acceso al dietista por termino medio presentan 1,06 unidades menos de HbA1c que los pacientes SIN dietista hospitalario a igualdad del resto de las variables predictoras incluidas en el modelo.

¿Qué diferencia promedio de HbA1c existe entre un paciente A tratado con ADO (1) de un paciente N tratado con insulina (0) a igualdad del resto de las variables predictoras incluidas en el modelo?

Variables predictoras	$\beta_i$	Paciente A	Paciente N
ADO/insulina (ADO=1, insulina=0)	+ 0,29	ADO [1]	insulina [0]

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

$$\text{HbA1c (A)} = 10,51 + 0,29 (1) = 10,8$$

$$\text{HbA1c (N)} = 10,51 + 0,29 (0) = 10,51$$

Los pacientes en tratamiento con ADO por termino medio presentan 0,29 unidades más de HbA1c que los pacientes tratados con insulina a igualdad del resto de las variables predictoras incluidas en el modelo.

¿Qué diferencia promedio de HbA1c existe entre un paciente A tratado sólo con DIETA (1) de un paciente B tratado con insulina (0) a igualdad del resto de las variables predictoras incluidas en el modelo?

Variables predictoras	$\beta_i$	Paciente A	Paciente N
Dieta/insulina (dieta=1, insulina=0)	- 0,97	dieta [1]	insulina [0]

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

$$\text{HbA1c (D)} = 10,51 - 0,97 (1) = 9,54$$

$$\text{HbA1c (N)} = 10,51 - 0,97 (0) = 10,51$$

Los pacientes en tratamiento sólo con dieta por término medio presentan 0,97 unidades menos de HbA1c que los pacientes tratados con insulina a igualdad del resto de las variables predictoras incluidas en el modelo.

## ¿Qué diferencia promedio de HbA1c existe entre dos pacientes A y B con las siguientes características?

Variables predictoras	Paciente A	Paciente B
Antidiabéticos orales/insulina [1/ 0]	ADO [1]	Insulina [0]
Dieta/insulina [1/ 0]	Dieta [1]	Insulina [0]
Nº enfermedades relacionadas con la diabetes (>1/<=1) [1/ 0]	3 [1]	1 [0]
Sexo (mujer/hombre)	Hombre [0]	Mujer [1]
Años desde el diagnóstico (> 7 / ≤ 7)	8 años [1]	3 años [0]
Acceso al dietista hospitalario (si/no) [1/ 0]	Si [1]	No [0]
Médico interesado en la diabetes (si/no)	Si [1]	Si [1]

HbA1c A = 10,51 + 0,29 (1) – 0,97 (0) + 0,20 (1) + 0,72 (0) + 0,04 (1) – 1,06 (1) – 0,86 (1) = 9,12%

HbA1c B = 10,51 + 0,29 (0) – 0,97 (1) + 0,20 (0) + 0,72 (1) + 0,04 (0) – 1,06 (0) – 0,86 (1) = 9,40%

Diferencia HbA1c (A-B) = - 0,28%

Dr. Jacobo Díaz Portillo  
Coordinador de Formación Continuada  
Hospital Universitario INGESA de Ceuta



*Rev Diagn Biol 2000; 49: 125-137*

## **PREMIO MENARINI-A.E.B.M.**

## **ORIGINAL**

# Lipoproteína (a) y diabetes mellitus tipo 2: factores que influyen en las concentraciones plasmáticas y utilidad como marcador de complicaciones tardías.

**A. García\***, **J. Díaz\*\***, **L. Escobar\*\*\***, **M. Aguilar\*\*\*\***, **A. Mangas\*\*\*\***

\*Servicio de Endocrinología. Hospital del INSALUD. Ceuta.

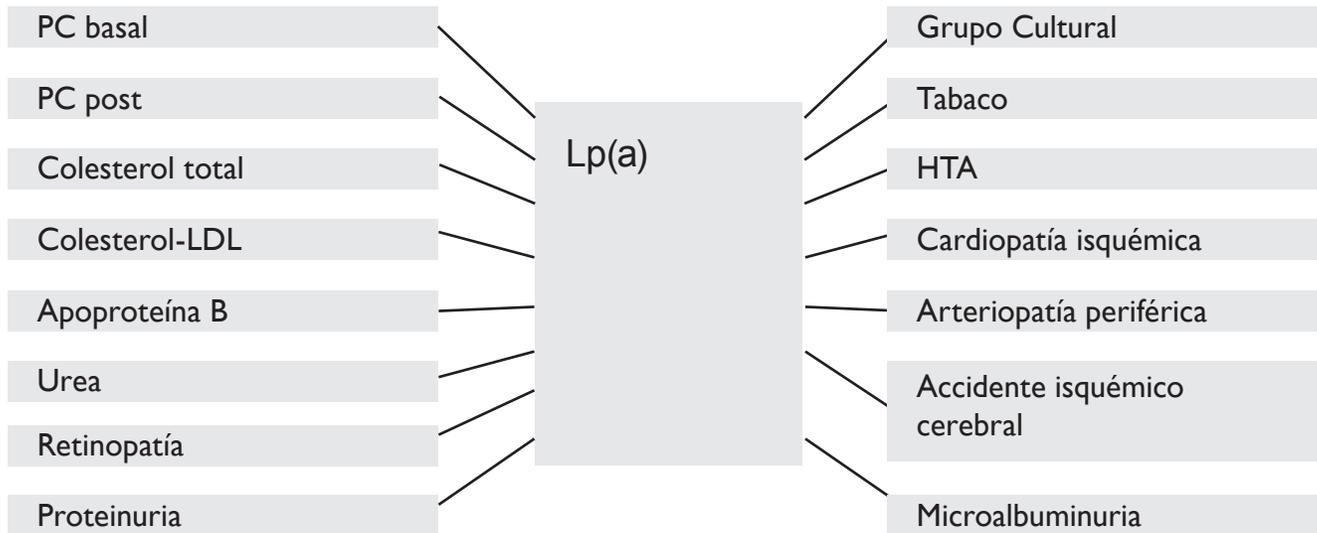
\*\*Servicio de Análisis Clínicos. Hospital del INSALUD. Ceuta.

\*\*\*Servicio de Endocrinología. Hospital Universitario Puerta del Mar de Cádiz.

\*\*\*\*Servicio de Medicina Interna. Hospital Universitario Puerta del Mar de Cádiz.

*Palabras clave:* diabetes mellitus tipo 2, lipoproteína (a), control glucémico, tipo de tratamiento, obesidad, hipertensión, hábito tabáquico, dislipemias, tiempo de evolución, complicaciones vasculares, macroangiopatía, cardiopatía isquémica, retinopatía, microalbuminuria, nefropatía.

*Key Words:* Type 2 diabetes mellitus, lipoproteína (a), glycaemic control, mode of treatment, obesity, hypertension, tobacco consumption, dyslipidemias, duration of diabetes, vascular diabetic complications, macroangiopathy, ischemic coronary disease, retinopathy, microalbuminuria, nephropathy.



## Análisis de regresión lineal múltiple

1.- Grupo cultural

Cristianos: [1]

Musulmanes: [0]

2.- Tabaco (si [1]; no [0])

3.- Hipertensión arterial

si: [1]

no: [0]

4.- Cardiopatía isquémica

si: [1]

no: [0]

5.- Arteriopatía periférica (si [1]; no [0])

6.- Accidente isquémico cerebral (si [1]; no [0])

7.- Retinopatía (si [1]; no [0])

8.- Microalbuminuria (si [1]; no [0])

9.- Proteinuria (si [1]; no [0])

Variables predictoras (X)  
CUALITATIVAS

Variable resultado (Y):  
Lipoproteína a

314 pacientes con  
DM tipo 2

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

Amor García, Jacobo Díaz, Luís Escobar, Manuel Aguilar, Alipio Mangas.

LIPOPROTEÍNA (a) Y DIABETES MELLITUS TIPO 2: FACTORES QUE INFLUYEN EN LAS CONCENTRACIONES PLASMÁTICAS Y UTILIDAD COMO MARCADOR DE COMPLICACIONES TARDÍAS (Premio AEBM). Revista: Revista de Diagnóstico Biológico 2000; 49: 125-137.

## Análisis de regresión lineal múltiple

- 1.- Edad
- 2.- Peso
- 3.- IMC
- 4.- ICC
- 5.- Tensión arterial sistólica (TAS)
- 6.- Tensión arterial diastólica (TAD)
- 7.- Glucemia
- 8.- Péptido C basal (PCB)
- 9.- Péptido C postprandial (PCP)
- 10.- IPC      IPC = PCP - PCB
- 11.- HbA1c
- 12.- Colesterol total
- 13.- Triglicéridos
- 14.- Colesterol-HDL
- 15.- Colesterol-LDL
- 16.- Apoproteína A1
- 17.- Apoproteína B
- 18.- Urea
- 19.- Creatinina

Variables predictoras (X)  
CUALITATIVAS

Variable resultado (Y):  
Lipoproteína a

314 pacientes con  
DM tipo 2

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

Amor García, Jacobo Díaz, Luís Escobar, Manuel Aguilar, Alipio Mangas.  
LIPOPROTEÍNA (a) Y DIABETES MELLITUS TIPO 2: FACTORES QUE INFLUYEN EN LAS  
CONCENTRACIONES PLASMÁTICAS Y UTILIDAD COMO MARCADOR DE COMPLICACIONES  
TARDÍAS (Premio AEBM). Revista: Revista de Diagnóstico Biológico 2000; 49: 125-137.

## Análisis de regresión lineal múltiple

Variables asociadas al nivel de lipoproteína (a)					
Variables predictoras	Estimación $\beta$	Error estándar $EE\beta$	$\beta/EE\beta$	Test F	p
<b>CUALITATIVAS</b>					
Grupo cultural	- 5,1428	2,7270	1,88	3,5564	p > 0,1
Tabaco	0,1952	3,1477	0,062	0,0038	NS
HTA	2,4388	2,5480	0,960	0,9161	NS
Cardiopatía isquémica	8,4219	2,8256	2,98	8,9028	p < 0,01
Arteriopatía periférica	3,1005	3,375	0,919	0,8440	NS
AIC	2,041	3,114	0,656	0,4298	NS
Retinopatía	- 0,8838	2,6850	0,329	0,1084	NS
Microalbuminuria	0,4571	2,6964	0,169	0,0287	NS
Proteinuria	12,905	4,1180	3,13	9,821	p < 0,005

$\beta / EE\beta$ : Coeficiente de regresión estandarizado

## Análisis de regresión lineal múltiple

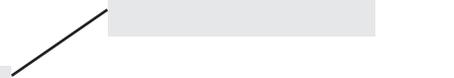
Variables asociadas al nivel de lipoproteína (a)					
Péptido C basal	- 1,863	0,9581	1,94	3,7807	0,051
Péptido C post	- 0,027	0,5067	0,053	0,0029	NS
Colesterol total	0,0638	0,0573	1,113	1,2420	NS
Colesterol-LDL	- 0,0518	0,0628	0,825	0,679	NS
Apoproteína B	0,0242	0,0747	0,324	0,1052	NS
Urea	0,1340	0,0704	1,90	3,6205	p > 0,1
AIC	2,041	3,114	0,656	0,4298	NS

$\beta$  / EE $\beta$ : Coeficiente de regresión estandarizado

Lp(a)

Cardiopatía isquémica

Proteinuria



## Discusión (parte final)

“Por lo tanto, la determinación de Lp(a) presenta utilidad como marcador de las complicaciones tardías de la DM tipo 2, y sus concentraciones plasmáticas están influenciadas de forma directa e independiente por la cardiopatía isquémica y la nefropatía. Cabe destacar que a pesar de que el diseño del estudio es transversal y no permite inferir una relación causal, si tenemos en cuenta el **condicionamiento genético de la Lp(a)**, los resultados sugieren que **la elevación de los niveles plasmáticos de Lp(a) predisponen al enfermo diabético a la aparición de complicaciones vasculares**, principalmente la cardiopatía isquémica y la nefropatía diabética. A pesar de tratarse de un estudio con elevado número de pacientes, dada la gran variabilidad interindividual de las concentraciones de Lp(a), estas asociaciones deberían confirmarse en estudios longitudinales más amplios y más a largo plazo que los publicados en la actualidad”.

Amor García, Jacobo Díaz, Luís Escobar, Manuel Aguilar, Alipio Mangas.

LIPOPROTEÍNA (a) Y DIABETES MELLITUS TIPO 2: FACTORES QUE INFLUYEN EN LAS CONCENTRACIONES PLASMÁTICAS Y UTILIDAD COMO MARCADOR DE COMPLICACIONES TARDÍAS (Premio Nacional AEBM). Revista: Revista de Diagnóstico Biológico 2000; 49: 125-137.

## Mito I: “La sonrisa de la esfinge”

Según se desprende en el final de la discusión del estudio, se sugiere que la Lp(a), por su condicionamiento genético, podría actuar como variable predictora principal de la aparición de algunas complicaciones vasculares como la cardiopatía isquémica y la nefropatía diabética en los pacientes con diabetes tipo 2. Sin embargo en el análisis multivariante la has considerado con variable resultado.

¿Estás seguro que la lipoproteína a es realmente la variable resultado?

¿No crees que la Lp(a) podría ser la variable predictora principal y la aparición de algún tipo complicaciones vasculares la variable resultado?

Variable PREDICTORA  
Cuantitativa continua

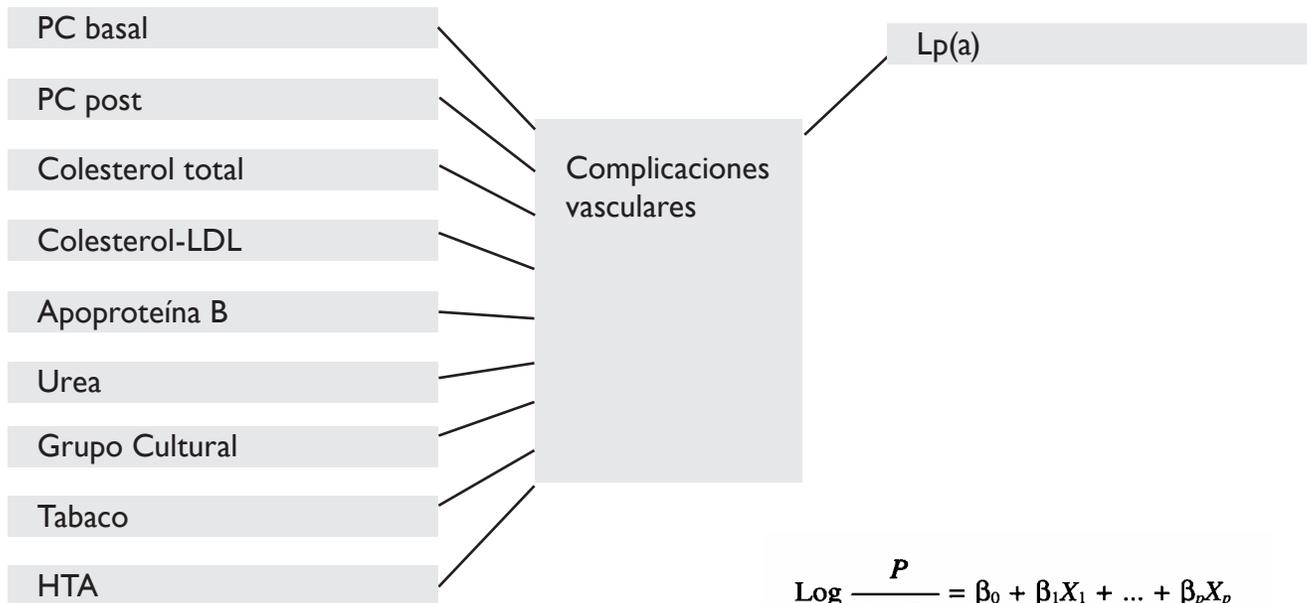
Lp(a)



Variable RESULTADO  
Cualitativa dicotómica (sí o no)

Complicaciones vasculares  
- Cardiopatía isquémica  
- Proteinuria

$$\text{Log} \frac{P}{1 - P} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$



$$\text{Log} \frac{P}{1 - P} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

Problema clínico: En Atención Primaria la eficiencia de la determinación de la excreción urinaria de albúmina en orina de 24 horas (EUA) para el diagnóstico de la nefropatía diabética es baja.

Objetivo del estudio: Introducir el cociente albúmina/creatinina en una muestra de orina matinal (CAC) como alternativa a la EUA.

Pregunta de investigación: ¿Podemos sustituir el cociente albúmina/creatinina en orina matinal (CAC) por la excreción urinaria de albúmina (EUA) en orina de 24 horas en pacientes con sospecha de nefropatía diabética? ¿Qué relación existe entre el CAC y la EUA en orina de 24 horas en dichos pacientes?.

¿Cuál es el test estadístico más acertado?

## Pregunta de investigación

¿Qué relación existe entre el cociente albúmina/creatinina en orina matinal (CAC) y la excreción urinaria de albúmina (EUA) en orina de 24 horas?

Variable PREDICTORA  
Cuantitativa continua

CAC



Variable RESULTADO  
Cuantitativa continua

EUA

REGRESIÓN LINEAL SIMPLE

$$Y = \beta_0 + \beta_1 * X$$

Coefficiente de correlación

$$r = 0,893 \text{ (89,3\%)}$$

## Coeficiente de correlación de Pearson

Correlación global EUA versus CAC  $r = 0,893$ ;  $R^2 = 0,806$ ;  $p < 0,01$

COEFICIENTE DE PEARSON: La magnitud de la asociación entre la EUA y la CAC es del 89,3% de la posible.

COEFICIENTE DE DETERMINACIÓN: El 80,6% de la variabilidad de la EUA puede ser explicada por la distinta CAC, resta por explicar un 19,4% que es achacable a otras variables PREDICTORAS distintas de la CAC.

¿Cuáles son esas variables predictoras?

Influencia del SEXO	
EUA versus CAC	Correlación de Pearson $r$
Sexo masculino	0,753
Sexo femenino	0,979

Las diferencias entre los coeficientes de correlación para cada sexo (masculino versus femenino) se pueden analizar por el método de OLKIN.

El test de Kaiser-Meyer-Olkin (KMO), que mide la idoneidad de los datos para realizar un análisis factorial comparando los valores de los coeficientes de correlación observados con los coeficientes de correlación parcial (grupos comparados). Si la suma de los cuadrados de los coeficientes de correlación parcial entre todos los pares de variables es pequeña en comparación con la suma de los coeficientes de correlación al cuadrado, esta medida tiende a uno. Para Kaiser los resultados del modelo factorial serán excelentes si el índice KMO está comprendido entre 0,9 y 1; buenos, si está comprendido entre 0,8 y 0,9; aceptables, si se encuentra entre 0,7 y 0,8; mediocres o regulares, cuando resulte entre 0,6 y 0,7; malos, si está entre 0,5 y 0,6; e inaceptables o muy malos cuando sea menor que 0,5.

$$KMO = \frac{(r_{\text{global}})^2}{(r_{\text{global}})^2 + (r_{\text{hombres}})^2}$$

El coeficiente de correlación parcial es un indicador de la fuerza de las relaciones entre dos variables eliminando la influencia del resto. Si las variables comparten factores comunes, el coeficiente de correlación parcial entre pares de variables deberá ser bajo, puesto que se eliminan los efectos lineales de las otras variables.

Una forma de evaluar este hecho es mediante la Medida de Adecuación de la Muestra KMO propuesta por Kaiser, Meyer y Olkin. Dicha medida viene dada por la ecuación KMO donde  $r_{ij}(p)$  es el coeficiente de correlación parcial entre las variables  $X_i$  y  $X_j$  eliminando la influencia del resto de las variables.

$$KMO = \frac{\sum_{i \neq j} \sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} \sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} \sum_{i \neq j} r_{ij}^2(p)}$$

$$KMO = \frac{(r_{\text{global}})^2}{(r_{\text{global}})^2 + (r_{\text{hombres}})^2} \qquad KMO = \frac{(r_{\text{global}})^2}{(r_{\text{global}})^2 + (r_{\text{mujeres}})^2}$$

$$KMO_{\text{hombres}} = 0,584$$

$$KMO_{\text{mujeres}} = 0,457$$

KMO es un índice que toma valores entre 0 y 1 y que se utiliza para comparar las magnitudes de los coeficientes de correlación observados con las magnitudes de los coeficientes de correlación parcial de forma que, cuanto más pequeño sea su valor, mayor es el valor de los coeficientes de correlación parciales  $r_{ij(p)}$  y, por lo tanto, menos deseable es realizar un Análisis Factorial. Kaise, Meyer y Olkin aconsejan que si  $KMO \geq 0,75$  la idea de realizar un análisis factorial es buena, si  $0,75 > KMO \geq 0,5$  la idea es aceptable y si  $KMO < 0,5$  es inaceptable.

Se observa que el valor del  $KMO = 0.584$  para hombres está en el límite de los valores recomendados por Kaiser, Meyer y Olkin poniendo de manifiesto que, en este caso muy probablemente, el proceso de reducción de datos no sea muy espectacular.

Se observa que el valor del  $KMO = 0.457$  para mujeres es inferior al límite de los valores recomendados por Kaiser, Meyer y Olkin (0,5) poniendo de manifiesto que, en este caso con toda seguridad, el proceso de reducción de datos no es importante.

# Regresión lineal múltiple

Factor 1:  
Variable predictora 1

CAC

$\beta_1 = 7.65$   $p=0,001$



Factor 3:  
Variable predictora 3

Edad

$\beta_3 = -0.54$   $p=0,42$



Factor 2:  
Variable predictora 2

Sexo

$\beta_2 = -30.26$   $p=0,01$



$R^2 = 0,806$  (80,6%)

Variable RESULTADO  
Cuantitativa continua

EUA

Aplicación del modelo de regresión múltiple a los datos de un estudio que usa como variables predictoras el cociente albúmina/creatinina en orina matinal (CAC), sexo y edad del individuo frente a la excreción urinaria de albúmina en orina 24 horas EUA (variable resultado).

## Análisis de regresión lineal múltiple

EUA: Excreción urinaria de albúmina en orina de 24 horas

CAC: Cociente albúmina/creatinina en una muestra de orina matinal

EUA: Variable dependiente o resultado

CAC: Variable Predictora 1

SEXO: Variable Predictora 2

EDAD: Variable Predictora 3

Modelo de Regresión Lineal Múltiple			
	Coefficiente $\beta$	IC95% de $\beta$	p
Constante	110,85	13,88 a 187,84	0,02
CAC	7,65	7,13 a 8,16	0,001
Edad	- 0,54	- 1,86 a 0,78	0,42
Sexo	- 30,26	- 53,93 a - 6,58	0,01

$$EUA (Y) = 110,85 + 7,65 CAC - 0,54 EDAD - 30,26 SEXO$$

$$EUA (Y) = 110,85 + 7,65 CAC - 30,26 SEXO$$

## Fases del análisis estadístico

1. ANÁLISIS DESCRIPTIVO

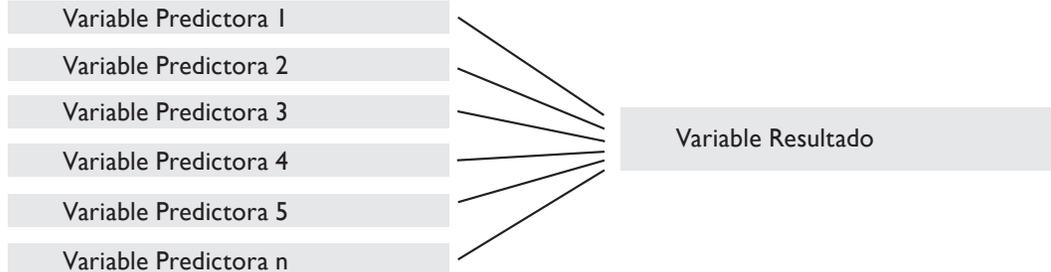


2. ANÁLISIS BIVARIANTE

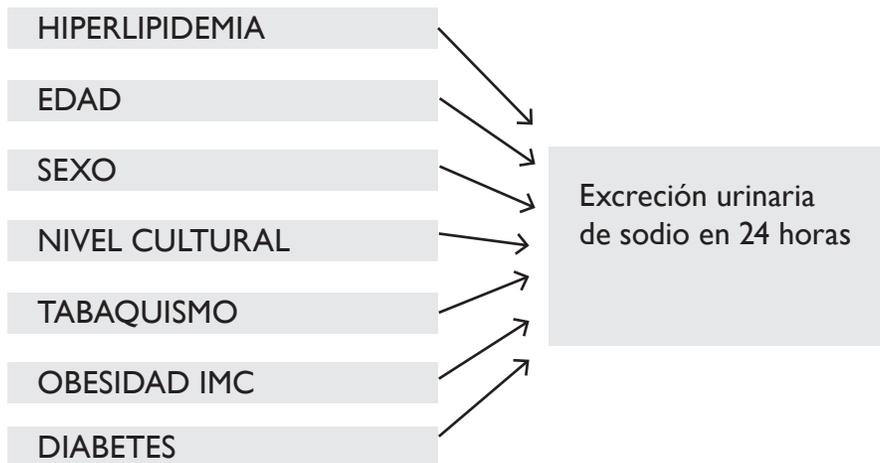


3. ANÁLISIS MULTIVARIANTE

¿Qué Variables PREDICTORAS se deben incluir en el Análisis de Regresión Lineal Múltiple?



Método de introducción de las variables	REGRESIÓN LINEAL MÚLTIPLE	REGRESIÓN LOGÍSTICA MULTIVARIANTE
Forward (hacia adelante)	Introducir progresivamente en el modelo las variables predictoras estadísticamente significativas, comenzando por la que tenga mayor significación	
Backward (hacia atrás)	Comienza por estimar un modelo que contiene todas las variables predictoras para ir eliminando posteriormente una a una, de forma sucesiva, aquellas que no sean significativas	
Stepwise	Es una combinación de los anteriores	



## ¿Qué variables predictoras se deben incluir en el análisis de regresión lineal múltiple?

Con significación estadística  
( $p < 0,05$ )

Edad

Sexo

Nivel cultural

Tabaquismo

Próxima a la significación  
estadística ( $0,05 < p < 0,1$ )

- IMC

- Hiperlipidemia

TABLA 2  
**Excreción urinaria de sodio en los distintos subgrupos de pacientes**

Variable	Valores (mmEq/24 h)	P
Edad		0,002
< 40 años	208,62 ± 80,27	
40-65 años	185,81 ± 86,01	
> 65 años	158,12 ± 65,84	
Sexo		0,000
Hombres	205,19 ± 86,52	
Mujeres	169,05 ± 76,40	
IMC		0,103
Normal	150,80 ± 79,00	
Sobrepeso	184,18 ± 80,85	
Obesidad	193,02 ± 91,43	
Nivel cultural		0,011
Leer y escribir	164,90 ± 71,73	
Educación primaria	201,22 ± 87,40	
Educación secundaria	193,85 ± 90,59	
Estudios superiores	181,59 ± 78,43	
Tabaquismo		0,000
Sí	207,92 ± 80,90	
No	171,76 ± 80,82	
Diabetes		0,634
Sí	190,42 ± 102,99	
No	183,05 ± 79,80	
Hiperlipidemia		0,100
LDL > 150	172,99 ± 81,31	
LDL < 150	189,07 ± 79,67	

IMC: índice de masa corporal; LDL: lipoproteínas de baja densidad.

## Análisis de regresión lineal múltiple

¿Cuál de las distintas variables predictoras con significación estadística ( $p < 0,05$ ) del análisis univariante NO aparecen en el modelo multivariante?

Nivel cultural

¿Cuál de las distintas variables predictoras próximas a la significación estadística ( $0,05 < p < 0,1$ ) del análisis univariante SI aparecen en el modelo multivariante?

Obesidad (IMC)

Resultados del análisis de Regresión Lineal Múltiple

	Coeficiente $\beta$	IC95% de $\beta$	p
Constante	96,67	19,2 a 174,15	0,015
Edad	- 1,17	- 1,89 a - 0,44	0,002
Sexo	28,71	4,65 a 52,78	0,020
Tabaquismo	29,22	3,83 a 54,6	0,024
IMC	4,15	2,00 a 6,29	0,000

La codificación de las variables dicotómicas se realizó sexo: mujer = 0, hombre = 1, tabaquismo: no fumador 0, fumador 1).  $IMC = \text{peso} / (\text{talla})^2$

En el análisis de regresión lineal MÚLTIPLE se incluyeron las variables predictoras con Significación estadística en el análisis de regresión lineal SIMPLE (edad, sexo, nivel cultural, tabaquismo;  $p < 0,05$ ) y aquellas próximas a la significación (IMC y hiperlipidemia;  $p = 0,1$ ), identificándose como verdaderas variables predictoras o independientes sólo la edad, sexo, IMC y el tabaco  $r = 0,388$  ( $p < 0,05$ ). El resto de las variables NO fueron incluidas en el modelo.

## Pregunta de investigación

¿Qué relación existe entre los niveles de colesterol en sangre y la tensión arterial?

Variable PREDICTORA  
Cuantitativa continua

Colesterol



Variable RESULTADO  
Cuantitativa continua

Tensión arterial

REGRESIÓN LINEAL SIMPLE

$$Y = \beta_0 + \beta_1 * X$$

Coefficiente de correlación

$$R^2 = 0,691 \quad (69,1\%)$$

## Modelo de regresión lineal simple para la variable resultado tensión arterial diastólica ajustado por la variable predictora colesterol

Variable	Coeficiente $\beta$	EE $\beta$	IC95% $\beta$	t	p
Constante	26,91	7,15	12,63 – 41,19	3,76	< 0,001
Colesterol	0,23	0,03	0,17 – 0,29	7,70	< 0,001
	Suma de cuadrados	g.l	Media Cuadrática	F	p
Regresión	4067,11	1	4067,11	59,29	< 0,001
Residual	4459,01	65	68,60		
Total	8526,12	66	29,22		

$$Y = \beta_0 + \beta_1 * X$$

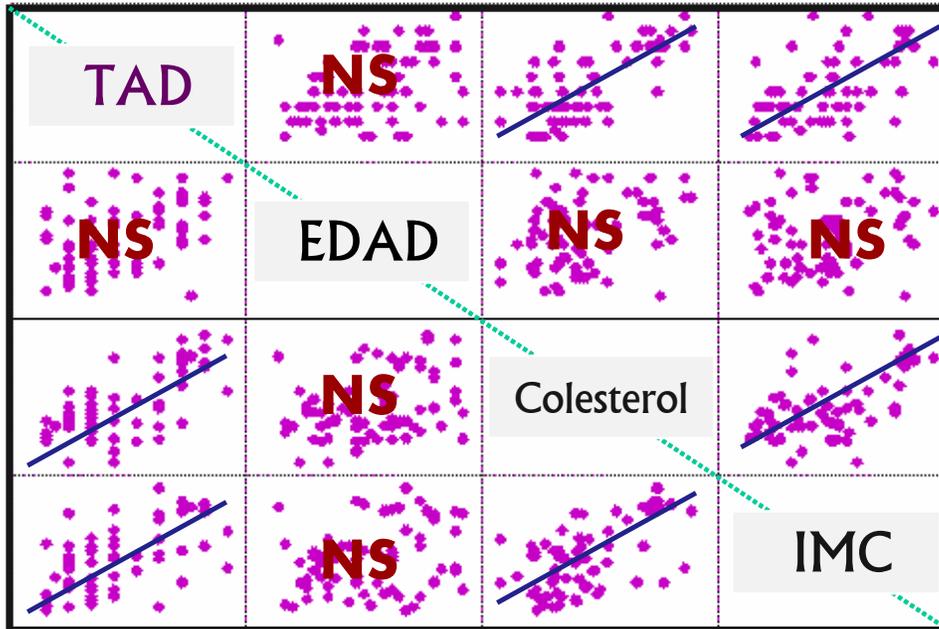
$$TAD = 26.91 + 0.23 * Colesterol$$

Es el modelo de regresión simple que resulta de tomar como única variable predictora el colesterol de un individuo. El valor del coeficiente de determinación  $R^2$  es 0.691 (69.1%).

¿Se puede mejorar el  $R^2$  incluyendo otra variable predictora?.

$R^2$  puede ser mejorado a 0.722 (72.2%) aplicando un modelo multivariante en el que se introduce el índice de masa corporal (IMC) como nueva variable predictora.

Correlación lineal de las variables predictoras edad, colesterol e índice de masa corporal con la variable resultado tensión arterial diastólica



# Regresión lineal múltiple

Variable  
Predictora 1

Causa o Factor 1: Cuantitativa



Variable RESULTADO

Interacción



Desenlace o efecto:  
Cuantitativa

Variable  
Predictora 2

Causa o Factor 2: Cuantitativa



$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2$$

# Regresión lineal múltiple

$$R^2 = 0,722 \quad (72,2\%)$$

Factor 1:  
Variable predictora 1

Colesterol

t= 5.26 p<0,001



Factor 2:  
Variable predictora 2

IMC

t= 2.45 p=0,017



Desenlace o efecto

Tensión arterial

Aplicación del modelo de regresión múltiple a los datos de un estudio que usa como variables predictoras de la tensión diastólica (variable resultado) el colesterol y el índice de masa corporal de un individuo.

## Modelo de regresión lineal simple para la variable resultado tensión arterial diastólica ajustado por las variables predictoras colesterol e índice de masa corporal

 $R^2 = 0,722 \quad (72,2\%)$ 

Variable	Coefficiente $\beta$	EE $\beta$	IC95% $\beta$	t	p
Constante	19,42	7,54	4,37 – 34,48	2,58	0,012
Colesterol	0,18	0,03	0,11 – 0,25	5,26	< 0,001
IMC	0,73	0,30	0,14 – 1,33	2,45	0,017
	Suma de cuadrados	g.l	Media Cuadrática	F	p
Regresión	4449,72	2	2224,86	34,93	< 0,001
Residual	4076,40	64	63,69		
Total	8526,12				

$$\text{TAD} = 19,42 + 0.18 * \text{Colesterol} + 0,73 \text{ IMC}$$

Los coeficientes de regresión  $\beta$  para ambas variables predictoras fueron 0.18 ( $\epsilon\epsilon\beta = 0.03$ ) y 0.73 ( $\epsilon\epsilon\beta = 0.30$ ) respectivamente, siendo ambos significativamente distintos de cero. Esto indica que por término medio la tensión arterial diastólica de un paciente se incrementa en 1.8 y 7.3 respectivamente por cada 10 unidades a mayores en su colesterol o índice de masa corporal. El valor del coeficiente de determinación  $R^2 = 0.72$  (72%) y la significación del contraste F global de la regresión ( $p < 0.001$ ) sugieren que gran parte de la variabilidad de la respuesta (TAD) viene explicada por el modelo ajustado.

# Regresión lineal múltiple

Factor 1:  
Variable predictora 1

Colesterol

t= 4.84 p<0,001



Factor 3:  
Variable predictora 3

Edad

t= 1.67 p=0.10



Factor 2:  
Variable predictora 2

IMC

t= 2.31 p=0,024



$R^2 = 0,52$  (52%)

Desenlace o efecto

Tensión arterial

Aplicación del modelo de regresión múltiple a los datos de un estudio que usa como variables predictoras de la tensión diastólica (variable resultado) el colesterol, índice de masa corporal y edad de un individuo.

## Modelo de regresión lineal simple para la variable resultado tensión arterial diastólica (TAD) ajustado por las variables predictoras colesterol, índice de masa corporal y edad $R^2 = 0,52$ (52%)

Variable	Coefficiente $\beta$	EE $\beta$	IC95% $\beta$	t	p
Constante	10,55	9,13	-7,70 a 28,81	1,15	0,252
Colesterol	0,17	0,03	0,1 – 0,24	4,84	< 0,001
IMC	0,68	0,30	0,09 – 1,28	2,31	0,024
Edad	0,24	0,14	- 0,05 a 0,53	1,67	0,100
	Suma de cuadrados	g.l	Media Cuadrática	F	p
Regresión	4622,52	3	1540,84	24,87	< 0,001
Residual	3903,60	63	61,96		
Total	8526,12	66			

Los coeficientes de regresión  $\beta$  para el colesterol fue 0.17 ( $\epsilon\epsilon\beta = 0.03$ ) y de 0.68 ( $\epsilon\epsilon\beta = 0.30$ ) para el IMC continuarán siendo distintos de cero ( $p < 0.05$ ). La introducción de la nueva variable predictora EDAD en el modelo, proporciona un cambio en el coeficiente de determinación  $R^2$  de tan sólo 0.028 que NO mejora de forma significativa ( $p = 0.059$ ). La entrada de la variable edad NO produce una mejora del ajuste del modelo. El coeficiente de regresión  $\beta$  de la EDAD (0.24) no es significativo ( $p = 0.100$ ).

## Modelo de regresión lineal simple para la variable resultado tensión arterial diastólica (TAD) ajustado por las variables predictoras colesterol, índice de masa corporal y edad

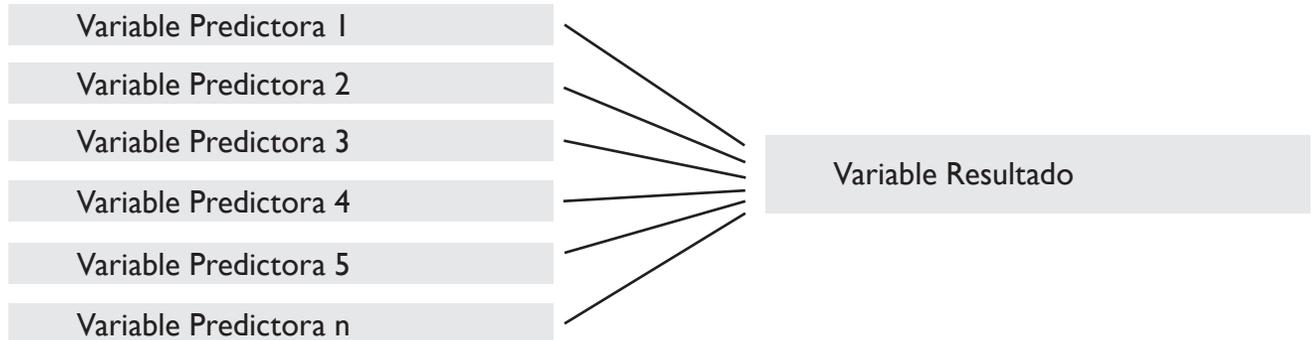
Variable	Coficiente $\beta$	EE $\beta$	IC95% $\beta$	t	p
Constante	10,55	9,13	-7,70 a 28,81	1,15	0,252
Colesterol	0,17	0,03	0,1 – 0,24	4,84	< 0,001
IMC	0,68	0,30	0,09 – 1,28	2,31	0,024
Edad	0,24	0,14	- 0,05 a 0,53	1,67	0,100
	Suma de cuadrados	g.l	Media Cuadrática	F	p
Regresión	4622,52	3	1540,84	24,87	< 0,001
Residual	3903,60	63	61,96		
Total	8526,12	66			

$$\frac{\beta_{\text{edad}}}{EE\beta_{\text{edad}}} = \frac{0,24}{0,14} = 1,71 < Z_{95\%} (1,96) \quad p = 0,100$$

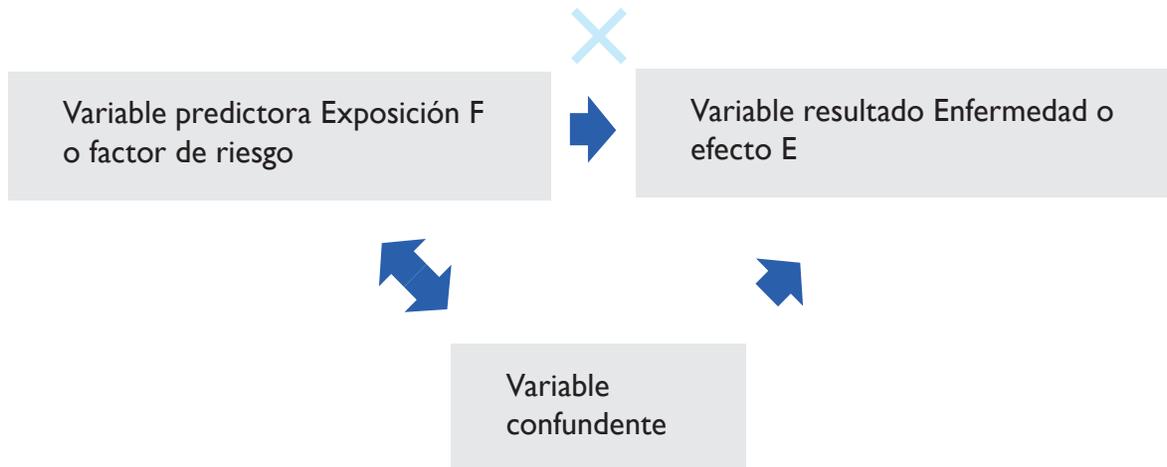
Método de introducción de las variables	REGRESIÓN LINEAL MÚLTIPLE	REGRESIÓN LOGÍSTICA MULTIVARIANTE
Forward (hacia adelante)	Introducir progresivamente en el modelo las variables predictoras estadísticamente significativas, comenzando por la que tenga mayor significación	
Backward (hacia atrás)	Comienza por estimar un modelo que contiene todas las variables predictoras para ir eliminando posteriormente una a una, de forma sucesiva, aquellas que no sean significativas	
Stepwise	Es una combinación de los anteriores	

## Análisis multivariante

Análisis multivariante para la detección de potenciales factores de confusión



## Sesgo de confusión



Es el único sesgo que se puede controlar (evitar) en la fase de análisis y no sólo en la fase de diseño.

Distorsión en las conclusiones del estudio, producida por la distribución desigual, en los grupos de comparación, de una tercera variable (confundente C).

## Portadores del Alelo €4

- Elevada prevalencia de enfermedad coronaria
- Mayor extensión de la arteriosclerosis
- Mayor tasa de IAM
- Mayor mortalidad de causa coronaria

Alelo €4



Enfermedad coronaria

### - ANÁLISIS UNIVARIANTE:

El 15,7% de los portadores del alelo €4 presentaban CI frente al 6,9% en NO €4 ( $p < 0,01$ )

ORc = 2,49 IC95%: 1,19-5,22

## Modelo de regresión logística

### Exposición

Alelo €4



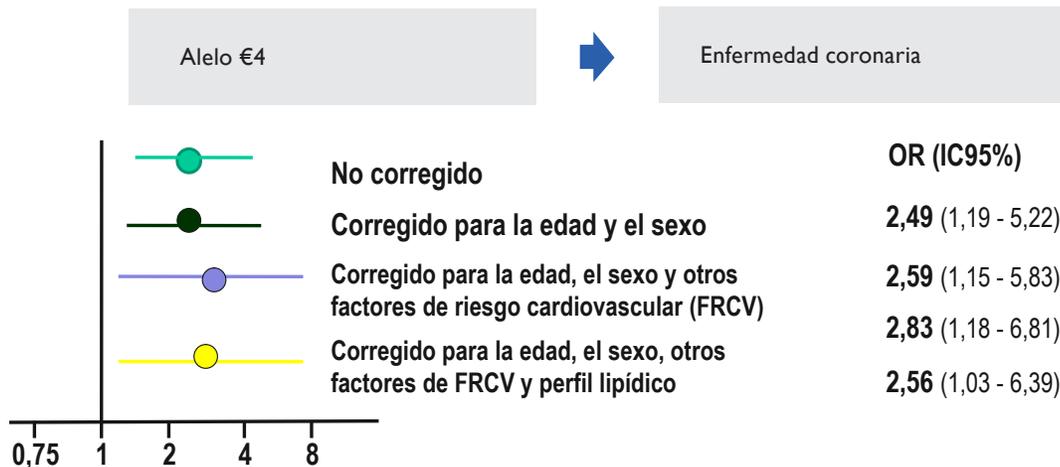
Enfermedad coronaria



### Potenciales FACTORES DE CONFUSIÓN

- Edad
- Sexo
- HTA
- Tabaquismo
- Obesidad (IMC)
- Lípidos y lipoproteínas

## Regresión logística multivariante para la detección de potenciales factores de confusión



Asociación entre la prevalencia de enfermedad coronaria y el tipo de apo E (€4 frente a no €4) expresada como OR e IC95% en una población de 386 pacientes con hipercolesterolemia. Se presenta la asociación simple (análisis univariante) (OR no corregido) y después de corregir, de forma escalonada, para distintas variables predictoras potencialmente confundentes como la edad y el sexo, para estas variables más la presencia de hipertensión arterial, tabaquismo e IMC y, por último, todos los anteriores más el perfil lipídico colesterol total, cHDL y triglicéridos mediante análisis multivariante.

## Reflexiones finales

Los distintos tipos de sesgos, la influencia del azar y la presencia de variables predictoras de CONFUSIÓN deben finalmente siempre, tenerse en cuenta, en todos los estudios como explicación posible de cualquier asociación estadística ya sea esta positiva, negativa o no existente.

Y como señalaba M. Susser en sus reflexiones sobre causalidad...  
“Cuando hay minas por todas partes uno no debe aventurarse sin un detector de minas (ANÁLISIS MULTIVARIANTE)”.

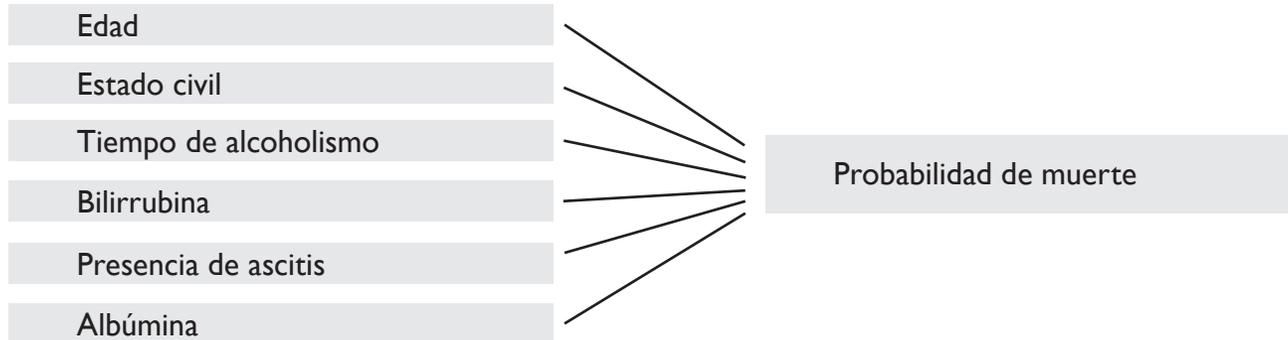
Susser M. Conceptos y estrategias en epidemiología. El pensamiento causal en ciencias de la salud. México: Biblioteca de la Salud; 1991.

	REGRESIÓN LINEAL MÚLTIPLE	REGRESIÓN LOGÍSTICA MULTIVARIANTE
Variable RESULTADO	Cuantitativa	Cualitativa (dicotómica)
Variables PREDICTORAS	Cualitativa Cuantitativa	Cualitativa Cuantitativa
ECUACIÓN	$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$	$\text{Log} \frac{P}{1 - P} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$

## Modelo de regresión múltiple de Cox

	REGRESIÓN MÚLTIPLE de COX
Variable RESULTADO	Cuantitativa Desenlace: supervivencia a tiempo t
Variables PREDICTORAS	Cualitativa Cuantitativa (factores pronóstico)
ECUACIÓN	$\ln(\lambda_i) = a + b_1x_1 + b_2x_2 + \dots + b_px_p$

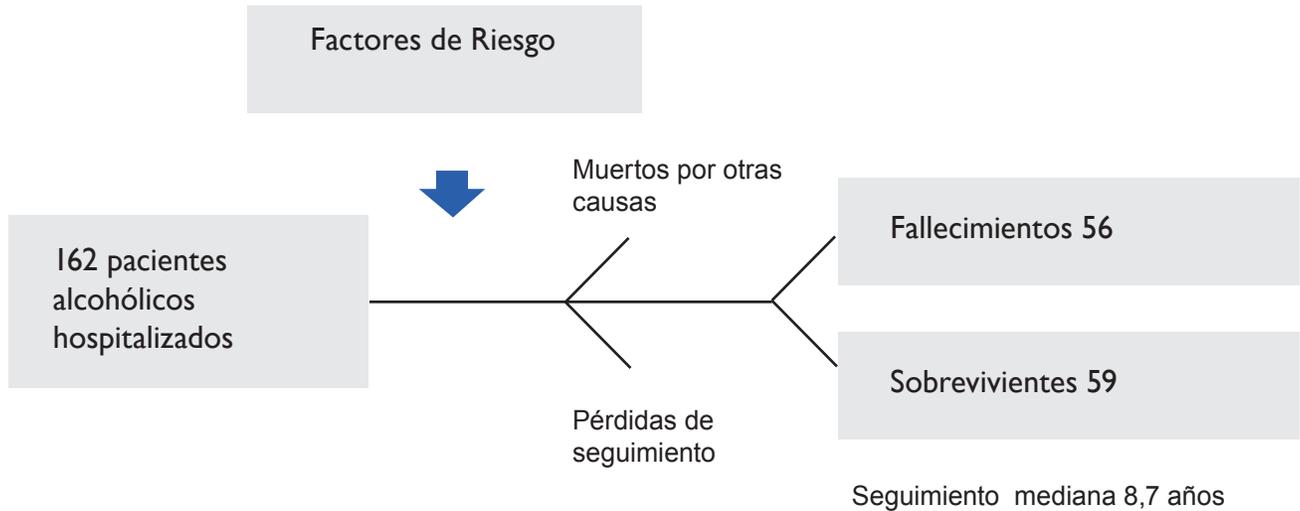
## Análisis multivariante



## Estudio de cohortes prospectivo de pacientes alcohólicos

VARIABLES PREDICTORAS	ANÁLISIS UNIVARIANTE (P)
Edad	$p < 0,05$
Estado civil	$p < 0,05$
Tiempo de alcoholismo	$p < 0,05$
Bilirrubina sérica	$p < 0,05$
Presencia de ascitis	$p < 0,05$
Albúmina sérica	$p < 0,05$

## Estudio de cohortes prospectivo de pacientes alcohólicos



## Estudio de cohortes prospectivo de pacientes alcohólicos



## Estudio de cohortes prospectivo de pacientes alcohólicos

Variables predictoras con significado pronóstico en pacientes alcohólicos  
Modelo de Regresión de Cox (método paso a paso)

Variable	RR	IC95% del RR	Significación
Edad	1,0763	1,0334-1,1209	0,0004
Ascitis	2,0940	1,0553-4,1551	0,0345
Bilirrubina	1,4220	1,2058-1,6769	0,0000
Estado civil	2,3922	1,1796-4,8515	0,0156

## Pruebas estadísticas de contraste de hipótesis más aplicables a cada situación

Variable Independiente o PREDICTORA	Variable Dependiente o RESULTADO	Pruebas de contraste de hipótesis	Observaciones y/o limitaciones del test
Cualitativa	Cualitativa	Ji cuadrado Prueba exacta de Fisher Test de McNemar Q de Cochran REGRESIÓN LOGÍSTICA	Si el tamaño muestral es grande Si el tamaño muestral es pequeño Dos grupos de muestras repetidas o apareadas Más de dos grupos de muestras repetidas o apareadas Varias variables predictoras cualitativas
Cualitativa	Cuantitativa	t de Student Análisis de la Varianza Test de Mann-Whitney Test de Wilcoxon Test de Kruskal-Wallis Test de Friedman	Compara medias entre dos grupos independientes Compara medias entre más de dos grupos independientes Dos grupos independientes No paramétricos Dos grupos medidas repetidas No paramétrico Más de dos grupos No paramétrico, medidas repetidas Más de dos grupos No paramétrico medidas repetidas
Cuantitativa	Cualitativa	Regresión logística simple Regresión logística multivariante	Una sola variable predictora cuantitativa Varias variable predictoras cuantitativa o cualitativas
Cuantitativa	Cuantitativa	Regresión lineal simple Correlación de Pearson Correlación de Sperman REGRESIÓN MÚLTIPLE	Paramétrico Paramétrico No paramétrico Varias variables predictoras cuantitativas
Cualitativa	Supervivencia	Kaplan-Meier Long-Rank Regresión de COX	Curvas de supervivencia Compara curvas de supervivencia Análisis Multivariante

# Capítulo 13

## Publicación Bioestadística: Bibliografía.

**Dr. Jacobo Díaz Portillo**  
Coordinador de Formación Continuada  
Hospital Universitario de Ceuta

## Bibliografía utilizada en el curso (I)

- Alvarez Cáceres R. *Estadística aplicada a las ciencias de la salud*. Madrid. Díaz de Santos 2007.
- Martínez-González MA, editor. *Bioestadística amigable*. 2ª ed. Madrid : Díaz de Santos, 2006.
- Martín Andrés A, Luna del Castillo JD. *Bioestadística para las ciencias de la salud*. 4ª ed. Madrid: NORMA; 1993.
- Milton JS, Tsokos JO. *Estadística para biología y ciencias de la salud*. Madrid: Interamericana McGraw Hill; 2001.
- Dawson-Saunders B, Trapp RG. *Bioestadística Médica*. 2ª ed. México: Editorial el Manual Moderno;1996.
- Pita Fernández, S. *Uso de la estadística y la epidemiología en atención primaria*. En: Gil VF, Merino J, Orozco D, Quirce F. Manual de metodología de trabajo en atención primaria. Universidad de Alicante. Madrid, Jarpyo Editores, S.A. 1997; 115-161.

## Bibliografía utilizada en el curso (2)

- Castillo Ron E. *Introducción a la estadística aplicada*. Santander: Ed. autor; 1993.
- Cobo E, Muñoz P, González JA. *Bioestadística para no estadísticos: bases para interpretar artículos científicos*. Ámsterdam : Elsevier Masson, 2007.
- Morell Ocaña M, Redondo Bautista M. *Metodología científica en ciencias de la salud*. Grupo editorial 33 (Málaga); 2002.
- Argimón Pallás JM, Jiménez Villa J. *Métodos de investigación aplicados a la atención primaria de salud*. 2ª ed. Barcelona: Mosby-Doyma; 1994.
- Sackett DL, Richardson WS, Rosenberg W, Hynes RB. *Evidence-based medicine: how to practice and teach EBM*. 2nd ed. London: Churchill-livingstone; 2000.
- Silva Ayçaguer LC. *Cultura estadística e investigación científica en el campo de la salud: una mirada crítica*. Madrid: Ediciones Díaz de Santos; 1997.

## Bibliografía utilizada en el curso (3)

- Bailar JC III. *Some uses of statistical thinking*. En: Bailar JC III, Mostoller F, editors. *Medical uses of statistics*. Boston: N Engl J Med Books, 1992; p. 25-57
- Bobenrieth Astete M.A . *Mitos y realidades en torno a la publicación científica*. *Med Clin (Barc)* 2000; 114: 339-341.
- Bobenrieth Astete M.A . *Lectura crítica de artículos originales en Salud. Medicina de Familia (And)* 2001; 2: 81-90.
- Antoja Ribó F. *The innocence of the storks*. *Rev Lab Clin* 2010; 3 (2): 51.
- Altman DG, Bland JM. *Absence of evidence is not evidence of absence*. *BMJ* 1995;311:485.
- Argimon JM. *La ausencia de significación estadística en un ensayo clínico no significa equivalencia terapéutica*. *Med Clin (Barc)* 2002; 118: 701-3
- Mora I, Ripoll R. *Cómo mejorar la calidad estadística de los artículos presentados a revistas biomédicas: lista de comprobación para los autores*. *Med Clin (Barc)*. 1999;113:138-49.

## Bibliografía utilizada en el curso (4)

- Salsburg DS. *The religion of statistics as practiced in medical journals*. Am Statistician 1985; 39: 220-223.
- Pocock SJ. *The role of statistics in medical research*. Br J Psychiat 1980; 137: 188-190.
- Chia KS. “Significant-itis”: an obsession with the P value. Scand J Work Environ Health. 1997; 23:152-4.
- Christenson P. *To p or not to p*. J Child Adolesc Psychiatr Nurs. 1995; 8:42.
- Neyman J, Pearson E. *On the problem of the most efficient tests of statistical hypothesis*. Philosophical trans of the Royal Society of London (A) 1933;231:289-337.
- Arís A. *Tómese una antes de acostarse*. 1ª edición. Barcelona: Planeta; 1998.

## Bibliografía utilizada en el curso (5)

- Martín Andrés A. *Bioestadística para ciencias de la salud*. 1ª ed. Madrid : Ediciones Norma-Capitel, 2004.
- Pérez López C. *Métodos estadísticos avanzados con SPSS*. Thomson, 2005.
- García A. *Métodos avanzados de estadística aplicada*. Madrid: UNED, 2001.
- Juez Martel P, Díez Vegas FJ. *Probabilidad y Estadística en Medicina*. Madrid: Díaz de Santos, 1997.
- Alvarez Cáceres R. *Estadística multivariante y no paramétrica con SPSS*. Madrid: Díaz de Santos, 1994.
- Armitage P, Berry G. *Estadística para la Investigación Biomédica*. Doyma, Barcelona, 1992.
- DL Sackett. *Epidemiología clínica*. Buenos Aires, Editorial Médica Panamericana, 1994.

## Bibliografía utilizada en el curso (6)

- Pita Fernández S, Pértega Díaz S. *Significancia estadística y relevancia clínica*. Cad Aten Primaria 2001; 8: 191-195.
- Escrig Sos J et al. *Intervalos de confianza: por qué usarlos*. Cir Esp. 2007;81(3):121-5.
- Camps García MT, Muñoz Rocan N, Calvo de Mora A, De Ramón Garrido E. *Los cocientes de probabilidad como medida de la utilidad de la determinación de anticuerpos antinucleares por inmunofluorescencia indirecta en la clínica. Aplicación de los resultados de la investigación clínica*. Med Clin (Barc) 1999;113: 476.
- Tibblea J et al. *Faecal calprotectin and faecal occult blood tests in the diagnosis of colorectal carcinoma and adenoma*. Gut 2001;49:402-408.

## Bibliografía utilizada en el curso (7)

- Seoanea T et al. *Estadística: Estadística Descriptiva y Estadística Inferencial*. SEMERGEN. 2007;33(9):466-71.
- Iraurgi I. *Significancia estadística y relevancia clínica*. Norte de Salud Mental 2009; 94 - 108.
- Rubio C. *Diseño estadístico de ensayos clínicos*. Med Clin (Barc) 1996; 107: 303-309.
- Plasencia A, Porta Serra M. *La calidad de la información clínica (II): significación estadística*. Med Clin (Barc) 1988; 90: 122-126.
- Carlos Manterola D, Viviana Pineda N. *El valor de “p” y la “significación estadística”. Aspectos generales y su valor en la práctica clínica. Interpretation of medical statistics*. Rev. Chilena de Cirugía 2008; 60: 86-89.

## Bibliografía utilizada en el curso (8)

- López de Ullibarri Galparsoro I, Pita Fernández S. *Medidas de concordancia: el índice Kappa*. Cad Aten Primaria 1999; 6: 169-171.
- Pita Fernández S, Pértegas Díaz S. *Relación entre variables cuantitativas*. Cad Aten Primaria 1997; 4: 141-144.
- Pita Fernández S, Pértegas Díaz S. *Pruebas diagnósticas*. Cad Aten Primaria 2003; 10: 120-124.
- Pita Fernández S, Pértegas Díaz S. *La fiabilidad de las mediciones clínicas: el análisis de concordancia para variables numéricas*. Cad Aten Primaria 2004.
- Pértegas Díaz S, Pita Fernández S. *Técnicas de regresión: Regresión lineal múltiple*. Cad Aten Primaria 2000; 7: 173-176.

## Bibliografía utilizada en el curso (9)

- Pita Fernández S, Pértegas Díaz S. *Técnicas de regresión: regresión lineal múltiple*. Cad Aten Primaria 2000; 7: 91-94.
- Pita Fernández S. *Correlación frente a causalidad JANO 1996; (1174): 59-60.*
- Pita Fernández S. *Determinación del tamaño muestral*. Cad Aten Primaria 1996; 3: 138-144.
- Pértegas Díaz S , Pita Fernández S. *La distribución normal*. Cad Aten Primaria 2001; 8: 268-274.
- Pértegas Díaz S , Pita Fernández S. *Métodos paramétricos para la comparación de dos medias. t de Student*. Cad Aten Primaria 2001; 8: 37-41.
- Passing H, Bablok W. *A new biometrical procedure for testing the equality of measurements from two different analytical methods*. J Clin Chem Biochem 1983; 21: 709-20.

## Bibliografía utilizada en el curso (10)

- Burgueño MJ, García-Bastos JL y González-Buitrago JM. *Las curvas ROC en la evaluación de las pruebas diagnósticas*. Med Clin (Barc) 1995; 104: 661-670.
- Hanley JA, McNeil BJ. *The meaning and use of the area under a Receiver Operating Characteristic (ROC) Curve*. Radiology 1982; 143: 29-36.
- López de Ullibarri Galparsoro I, Pita Fernández S. *Curvas ROC*. Cad Aten Primaria 1998; 5: 229-235.
- McNeil BJ, Keeler E, Adelstein SJ. *Primer on certain elements of medical decision making*. N Engl J Med 1975; 293: 211-215.
- Galen RS, Gambino SR. *Beyond normality: The predictive value and efficiency of medical diagnoses*. Nueva York: J Wiley and Sons, Inc., 1975.

## Bibliografía utilizada en el curso (II)

- Beck JR, Shultz EK. *The use of relative operating characteristic (ROC) curves in test performance evaluation*. Arch Pathol Lab Med 1986; 110:13-20.
- Swets JA. *Measuring the accuracy of diagnostic systems*. Science 1988; 240: 1.285-1.293.
- Zweig MH, Campbell G. *Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine*. Clin Chem 1993; 39: 561-577.
- Yupanqui Pacheco RM: *Introducción a la estadística bayesiana*. UNMSM. Facultad de Ciencias Matemáticas. EAP de Estadística, Lima, 2005 (disponible en: [http://sisbib.unmsm.edu.pe/bibvirtualdata/Tesis/Basic/yupanqui\\_pr/yupanqui\\_pr.pdf](http://sisbib.unmsm.edu.pe/bibvirtualdata/Tesis/Basic/yupanqui_pr/yupanqui_pr.pdf)).
- Altman DG, Moher D. *Elaboración de directrices para la publicación de investigación biomédica: proceso y fundamento científico*. Med Clin (Barc). 2005;125(Supl 1):8-13.

## Bibliografía utilizada en el curso (12)

- López, F. *Utilización adecuada de las pruebas estadísticas en los estudios de comparación de procedimientos*. Quim Clin 2000; 19: 237.
- Fleiss JL. *The design and analysis of clinical experiments*. Nueva York: John Wiley & Sons, Inc., 1986.
- Luis Prieto, Rosa Lamarca, Alfonso Casado. *La evaluación de la fiabilidad en las observaciones clínicas: el coeficiente de correlación intraclase*. Med Clin (Barc). 1998;110:142-5.
- (\*) Wakkers PJM, Hallendoorn HBA, Op de Weegh GJ, Heerspink W. *Applications of statistics in clinical chemistry. A critical evaluation of regression lines*. Clin Chim Acta 1975; 64: 173-84.
- Cornbleet PJ, Gochman N. *Incorrect least-squares regression coefficients in method-comparison analysis*. Clin Chem 1979; 25: 432-38.
- Zaydman L, Lasky FD, Van Nostrand C, Doptis PD. *Recommended linear regression models for comparison of method in clinical Chemistry*. Clin Chem 1991, 37: 960.

## Bibliografía utilizada en el curso (13)

- Frey E, Fuentes-Arderiu X, Queraltó JM. *Comparación estadística de métodos analíticos*. Educación continuada Química Clínica. 1988; 1: 69-79.
- López, F. *Utilización adecuada de las pruebas estadísticas en los estudios de comparación de procedimientos*. Quim Clin 2000; 19: 237.
- Westgard JO, Hunt MR. *Use and interpretation of common statistical test in method-comparison studies*. Clin Chem 1973; 19: 44-57.
- Bland JM, Altman DG. *Statistical methods for assesing agreement between two methods of clinical measurement*. Lancet 1986; 1: 307-310.
- Pita Fernández S, Pértegas Díaz S. *La fiabilidad de las mediciones clínicas: el análisis de concordancia para variables numéricas*. Cad Aten Primaria 2004.

## Bibliografía utilizada en el curso (14)

- Altman DG, Bland JM. *Statistical methods for assessing agreement between two methods of clinical measurement*. The Lancet 1986; 307-10.
- Altman DG, Bland JM. *Measurement in medicine: the analysis of method comparison studies*. The Statistician 1983; 32, 307-317.
- López, F. *Utilización adecuada de las pruebas estadísticas en los estudios de comparación de procedimientos*. Quim Clin 2000; 19: 237.
- Pértegas Díaz S , Pita Fernández S. *Determinación del tamaño muestral para calcular la significación del coeficiente de correlación lineal*. Cad Aten Primaria 2001; 9: 209-211.
- Jacobo Díaz Portillo .“*Influencia del grupo étnico en el patrón lipoproteico de una determinada población pediátrica de Ceuta*”, Tesis doctoral (1997). Facultad de Medicina de la Universidad de Cádiz.

## Bibliografía utilizada en el curso (15)

- Albert A, Harris EK. *Multivariate interpretation of clinical laboratory data*. New York: Marcel Dekker; 1987.
- Saez M. *El problema de las medidas repetidas. Análisis longitudinal en epidemiología*. Gac Sanit 2001;15(4):347-352.
- F. López Azorín. *La necesidad de mejores evaluaciones metodológicas y nuestra exigencia ante los criterios de aceptabilidad de los resultados*. Química Clínica 2003; 22 (6) 431-432.
- Sánchez-Cantalejo Ramírez E. *Regresión logística en salud pública. EASP 2000*.
- Silva Ayçaguer LC. *Excursión a la regresión logística en ciencias de la salud*. Madrid: Díaz de Santos, 1995.
- Hernández Aguado I, Porta Serra M, Miralles M, García Benavides F, Bolúmar F. *La cuantificación de la variabilidad en las observaciones clínicas*. Med Clin (Barc) 1990; 95: 424- 429.

## Bibliografía utilizada en el curso (16)

- Pértegas Díaz S, Pita Fernández S. *Representación gráfica en el análisis de datos*. Cad Aten Primaria 2001; 8: 112-117.
- López de Ullibarri Galparsoro I, Pita Fernández S. *Medidas de concordancia: el índice Kappa*. Cad Aten Primaria 1999; 6: 169-171.
- Prieto L, Lamarca R, Casado A. *La evaluación de la fiabilidad en las observaciones clínicas: el coeficiente de correlación intraclass*. Med Clin 1998; 110(4): 142-145.
- Argimon JM. *El intervalo de confianza: algo más que un valor de significación estadística*. Med Clin (Barc) 2002;118(10):382-4.
- Argimon JM. *La ausencia de significación estadística en un ensayo clínico no significa equivalencia terapéutica*. Med Clin (Barc) 2002;118(18):701-3.
- Ramalle-Gómara E, Andrés de Llano JM. *Utilización de métodos robustos en la estadística inferencial*. Aten Primaria 2003;32(3):177-82.
- Cobo E. *Análisis multivariante en investigación biomédica: criterios para la inclusión de variables*. Med Clin (Barc) 2002;119(6):230-237.

# Guía Práctica del Curso de Bioestadística Aplicada a las Ciencias de la Salud

**Dr. Jacobo Díaz Portillo**  
Coordinador de Formación Continuada  
Hospital Universitario de Ceuta



Edita: © Instituto Nacional de Gestión Sanitaria  
Subdirección General de Gestión Económica y Recursos Humanos  
Servicio de Recursos Documentales y Apoyo Institucional  
Alcalá, 56 28014 Madrid

El copyright y otros derechos de propiedad intelectual pertenecen al Instituto Nacional de Gestión Sanitaria.  
Se autoriza la reproducción total o parcial para uso no comercial, siempre que se cite explícitamente su procedencia.

Depósito Legal: M-33564-2011  
NIPO: 865-11-010-6  
Colección Editorial de Publicaciones del INGESA: 1.921

# Guía Práctica del Curso de Bioestadística Aplicada a las Ciencias de la Salud

Dr. Jacobo Díaz Portillo

Guía Práctica del Curso  
de Bioestadística Aplicada  
a las Ciencias de la Salud  
Dr. Jacobo Díaz Portillo



GOBIERNO  
DE ESPAÑA

MINISTERIO  
DE SANIDAD, POLÍTICA SOCIAL  
E IGUALDAD

