

**PROPUESTA DE UNA GUÍA PRACTICA ENFOCADA EN LA
UNIDAD DE ESTADÍSTICA DESCRIPTIVA PARA EL NIVEL
NM4, UTILIZANDO MICROSOFT EXCEL**

SEMINARIO PARA OPTAR AL GRADO DE
LICENCIADO EN EDUCACIÓN Y AL TÍTULO
DE PROFESOR DE EDUCACIÓN MEDIA EN
MATEMÁTICA E INFORMÁTICA EDUCATIVA

INTEGRANTES:
GRANDÓN VALDÉS, CLAUDIA TAMARA
JORQUERA UTRERAS, TONYA ANDREA

PROFESOR GUÍA:
TORRES BALCHEN, SERGIO LEIF

SANTIAGO-CHILE
2009

RESUMEN

Esta investigación de tipo cualitativa se sustenta bajo la disminuida integración de programas computacionales en los textos escolares a la hora de pasar una unidad o contenido específico, además de las vagas instrucciones para la utilización de estos, es así, como primeramente se selecciona la unidad de matemática a trabajar y el nivel a cual será enfocada la propuesta a crear, en este caso la unidad seleccionada fue Estadística Descriptiva para el nivel NM4, luego de esto se prosiguió con un análisis de tres libros de cuarto medio para conocer los tipos de secuencia didáctica planteados por cada uno de ellos, seguido del estudio de tres tipos de hoja de cálculo existentes con características adecuadas para el fin pedagógico.

Es así como la propuesta empieza a tomar cuerpo y claridad y surge instaurar un libro de contenidos de la unidad en cuestión además de incorporar ejemplos, actividades propuestas para el alumno en Microsoft Excel con su respectivo manual e indicaciones de trabajo, además de un solucionario para que así el estudiante pueda ir evaluando de manera autónoma su desempeño en las actividades propuestas en el libro.

Para poder llevar a cabo esta propuesta y poder observar resultados certeros, se realizó una encuesta a siete cursos de dos colegios ubicados en la comuna de Punte Alto, con la finalidad de ver resultados reales que avalen la calidad y éxito de la propuesta creada.

Índice

Introducción	Página 4
Capítulo I: Planteamiento del problema	
I.1. Problema.....	Página 8
I.2. Objetivo general.....	Página 8
I.3. Objetivos específicos.....	Página 8
Capítulo II: Marco Referencial	
II.1. Estadística descriptiva en el nivel NM4.....	Página 11
II.2. Estadística descriptiva en la PSU.....	Página 12
II.3. Secuencia didáctica.....	Página 15
II.4. Planilla Electrónica u hoja de calculo.....	Página 16
II.4.1. Definición.....	Página 16
II.4.2. Usos y aplicaciones.....	Página 16
II.4.3.Orígenes.....	Página 16
II.4.4. Hojas de calculo en el mercado.....	Página 17
Capítulo III: Marco Metodológico	Página 20
Capítulo IV: Presentación de Resultados	
IV.1. Análisis.....	Página 25
IV.1.1. Análisis de texto.....	Página 25
IV.1.2. Selección planilla de calculo.....	Página 29
IV.2. Construcción Guía Practica.....	Página 31
IV.2.1. Definición Secuencia Didáctica.....	Página 31
IV.2.2. Construcción Actividades.....	Página 35
IV.3. Implementación Actividad.....	Página 38
IV.3.1. Prueba Media Aritmética.....	Página 40
IV.3.2. Implementación Guía	Página 42
IV.3.2.1. Datos y Etapas Implementación Guía.....	Página 42
IV.3.2.2. Resultado Actividades.....	Página 44
IV.3.3. Cuestionario.....	Página 44
Capítulo V: Conclusiones	Página 49
Bibliografía	Página 51
Anexos	
Guía Práctica 4º Medio, Estadística Descriptiva.....	Página 53

Introducción

Este seminario tiene como objetivo crear una propuesta educativa con el fin de complementar aún más la enseñanza de la estadística descriptiva presente en los planes y programas del nivel NM4 en lo que a usos de tecnologías se refiere (utilización de planillas u hojas de cálculo para el manejo de medidas estadísticas). Donde los estudiantes puedan aplicar los contenidos aprendidos de manera diferente, trabajando actividades en hojas de cálculo (Microsoft Excel) con todas las indicaciones e instrucciones necesarias para llevarlas a cabo.

Este trabajo comprende de 5 capítulos: planteamiento del problema, marco referencial, marco metodológico, presentación de los resultados y conclusiones, además de un anexo, que es la guía práctica realizada.

En el planteamiento del problema, capítulo 1, se trata la problemática actual existente para la materia de estadística en el nivel NM4, esta problemática radica principalmente en la escasa información de instrucciones en los diversos libros para llevar a cabo la utilización del recurso tecnológico, en este caso la hoja de cálculo, se pretende que se aprendan contenidos de la materia de estadística descriptiva con tecnologías de la informática pero es muy poco lo que los mismo libros ofrecen como apoyo al estudiante para poder llevar a cabo la actividad. A raíz de este problema surge la necesidad de crear actividades con apoyo tecnológico, con instrucciones suficiente para que el estudiante las pueda llevar a cabo y además que ayuden a comprender la materia y sobre todo que lo ayuden a motivarse, es por esto, que un punto importante en esta propuesta es la familiaridad con la que se encontrarán en el libro de ejercicios desarrollados paso a paso con explicación en cada uno de sus procesos para que no quede duda alguna en cuanto a la manera de operar el programa Excel en la materia de estadística descriptiva para el nivel NM4. En esta etapa de la propuesta también se comprende los objetivos generales y específicos de esta investigación a realizar.

En el marco referencial, capítulo 2, se exponen principalmente 4 puntos importantes previos a la construcción de la investigación: Estadística descriptiva en el nivel NM4, Estadística descriptiva en la PSU, Secuencia Didáctica y Planilla electrónica u hoja de cálculo.

- Estadística descriptiva en el nivel NM4: inicia con la descripción de estadística descriptiva, este es un tipo de estadística que se ocupa de la recopilación y visualización de los datos recolectados, para su posterior

análisis. Esta parte de la estadística esta considerada en los planes y programas establecidos por el ministerio de educación y corresponde a la unidad 1 llamada "Probabilidad y Estadística". Así también se mencionarán los contenidos y aprendizajes esperados presente en los planes y programas.

- Estadística descriptiva en la PSU: se presentan ejercicios tipos de PSU y se hace un pequeño análisis de lo que abarca cada uno de estos ejercicios. Además se desglosa la estructura de la PSU área matemática para visualizar cuantas preguntas de cada contenido son evaluadas, de las 70 preguntas, 9 corresponden a "Probabilidad y Estadística".
- Secuencia Didáctica: en esta etapa queda plasmado el orden en que se presentan las actividades a través de las cuales se lleva a cabo el proceso de enseñanza-aprendizaje. Lo primordial es que se respete la sucesión de las actividades como un proceso en serie más que las actividades en si.
- Planilla electrónica u hoja de cálculo: aquí, se hace una descripción de lo que es una planilla u hoja de cálculo, así también se exponen sus orígenes, usos y aplicaciones: (*pequeñas bases de datos, informes, gráficos estadísticos, clasificaciones de datos, entre otros usos*).

Otro punto importante en este capítulo es "las hojas de cálculo en el mercado", aquí se hace referencia a algunas planillas u hojas de calculo mas conocidas como: Lotus 1-2-3, Calc, Microsoft Excel donde se mencionan en cada una de ellas orígenes, evolución en el mercado y distintas versiones para terminar en un cuadro comparativo entre las 3 planillas ya mencionadas.

En el marco metodológico, capítulo 3, etapa en la cual se hace referencia principalmente al tipo de investigación, se da una descripción de lo que es el enfoque cualitativo y como esta investigación se apega en gran manera a este tipo de enfoque, respetando las cualidades que debe poseer para enmarcarse dentro de una investigación cualitativa, algunas de estas son: recolección de datos sin medición numérica, los investigadores enfatizan tanto los procesos como lo resultados, la identificación del problema a investigar, identificación de los participantes, formulación de hipótesis, recolección de los datos (tipos de instrumentos de recolección), análisis de los datos , y conclusiones.

Presentación de los resultados, capítulo 4, en este capítulo dividido en 3 temas, análisis, construcción e implementación, en el apartado de análisis se estudiarán dos conceptos; análisis de textos y análisis de planillas de cálculo, en el segundo apartado construcción, se definen los diferentes pasos para poder llevar a cabo la propuesta y el orden que tiene cada contenido y actividad, en el tercer apartado implementación de las actividades, aquí se termina con la implementación de una de las actividades en siete cursos de cuarto año medio provenientes de dos colegios de Puente Alto.

Y finalmente se termina con una conclusión final de la investigación en donde se da paso a un análisis general de lo que fue la investigación en si.

Capítulo I

Planteamiento del Problema

I.1. Problema:

En la actualidad dentro de la educación, la estadística descriptiva ocupa un lugar fundamental en los planes y programas y a su vez en los libros entregados por el ministerio de educación para los estudiantes y profesores de cuarto año medio. El uso de la informática hoy en día esta al alcance de todos, por lo que se ha incorporado en el currículo el uso de esta. Si bien los textos escolares¹ de matemáticas contienen el material para ser trabajado en la unidad de estadística ("Educación matemática, cuarto año medio", editorial Santillana, año 2008), los esfuerzos por integrar el uso del recurso computacional son poco eficientes por la escasa información de instrucciones para llevar a cabo la utilización de este recurso. Es por esto que es necesario el uso de un software computacional, que facilite el tratamiento de datos y permita obtener una gran cantidad de información, acompañada de sus correspondientes gráficos, que facilite aun más el entendimiento de este.

A raíz de esta problemática surge la necesidad de diseñar estrategias metodológicas que utilicen una herramienta tecnológica (planilla de cálculo), que permita motivar al estudiante a comprender los contenidos de estadística descriptiva, analizar datos estadísticos e interpretar información obtenida de estos.

I.2. Objetivo General

Diseñar una secuencia didáctica -que incorpore el uso de una planilla electrónica- que colabore en el proceso de enseñanza-aprendizaje de la unidad estadística descriptiva presente en los planes y programas para el nivel NM4.

I.3. Objetivos Específicos

- Revisión de textos de estudio de matemáticas del nivel NM4, presentes en el mercado.
- Realizar un catastro de planillas electrónicas, factibles de ser utilizadas en la unidad de estadística descriptiva.

¹ Textos Matemáticas 4º medio: Mare-Nostrum (2003-2004), Zig-Zag (2005-2006), Santillana (2007-2008)

- Seleccionar planilla electrónica a utilizar en la secuencia didáctica.
- Diseñar secuencia didáctica para la unidad de estadística descriptiva presente en el nivel NM4.
- Testear las actividades de la secuencia didáctica en un grupo objetivo real (alumnos de NM4).

Capítulo II

Marco Referencial

En este capítulo se abarcan cuatro puntos importantes previos a la construcción de la investigación, los cuales son necesarios abordar para el desarrollo de la propuesta, estos son los siguientes, Estadística descriptiva en el nivel NM4, donde se presentan los contenidos y aprendizajes esperados presentes en los planes y programas de dicho nivel.

El segundo punto, Estadística descriptiva en la PSU, aquí se presentan ejercicios tipos PSU y se hace un pequeño análisis de lo que abarca cada uno de estos, el tercer punto, Secuencia Didáctica, el cual presenta la definición de dicho concepto, importantísimo factor a la hora de la construcción de la propuesta en proceso, y por último, se habla de la Planilla electrónica u hoja de cálculo, fundamental a la hora de la elección del programa a trabajar en dicha propuesta.

II.1. Estadística descriptiva en el nivel NM4.

La Estadística descriptiva es un tipo de estadística que se ocupa de la recopilación y visualización de los datos recolectados, para su posterior análisis.

Esta parte de la estadística está considerada en los planes y programas establecidos por el ministerio de educación y corresponde a la unidad 1 llamada "Probabilidad y Estadística" donde se encuentra de manera más detallada y específica en los contenidos y aprendizajes esperados para el nivel NM4.

Los contenidos y aprendizajes esperados presentes en los planes y programas, son los siguientes:

Contenidos:

1. Graficación e interpretación de datos estadísticos provenientes de diversos contextos. Crítica del uso de ciertos descriptores utilizados en distintas informaciones.

2. Selección de diversas formas de organizar, presentar y sintetizar un conjunto de datos. Ventajas y desventajas. Comentario histórico sobre los orígenes de la estadística.

3. Uso de planilla de cálculo para análisis estadístico y para construcción de tablas y gráficos.

4. Muestra al azar, considerando situaciones de la vida cotidiana: por ejemplo, ecología, salud pública, control de calidad, juegos de azar, etc. Inferencias a partir de distintos tipos de muestras.

Aprendizajes esperados:

Los alumnos y alumnas:

1. Conocen distintas maneras de organizar y presentar información incluyendo el cálculo de algunos indicadores estadísticos, la elaboración de tablas y gráficos utilizando planilla de cálculo o calculadora.

2. Reconocen la importancia de una muestra aleatoria simple para hacer inferencias sobre la población.

3. Conocen antecedentes históricos sobre la estadística y su relación con las probabilidades.

4. Comprenden y aprecian el papel de la estadística en la sociedad, conociendo algunos campos de aplicación.

Otro punto importante que consideran los planes y programa en la enseñanza de la estadística descriptiva es el uso de planillas de cálculos con el objetivo de obtener resultados de manera inmediata y además que los estudiantes puedan presentar datos mediante una variedad de gráficos (Dispersión, histograma, de barras, circular, etc.).

II.2 Estadística Descriptiva en la PSU

En la prueba de selección universitaria de matemáticas estipuladas por el DEMRE, se evalúan los contenidos dentro de un total de 70 preguntas, de las cuales se desglosan de la siguiente manera:

Números y Proporcionalidad 11 preguntas, Álgebra y Funciones 29 preguntas, Geometría 21 preguntas y Estadística y Probabilidad 9 preguntas, donde no se especifica cuantas son de estadística y cuantas de probabilidad.

Aquí se presentarán cuatro preguntas² tipo PSU enfocadas a estadística descriptiva, veamos las siguientes preguntas:

1.- En una muestra de alumnos de un colegio se tiene la siguiente distribución de edades:

Edad	Frecuencia
13	5
14	11
15	1
16	5
17	13

Tabla1: Pregunta uno, tipo PSU

La moda y la mediana de las edades de ese grupo son:

Moda - mediana

- A) 16 17
- B) 17 15
- C) 15 17
- D) 5 1
- E) 17 16

En esta pregunta se puede observar un ejercicio de variables discretas en donde se entrega una tabla de distribución y se pide calcular la moda y la mediana de estas edades, una pregunta con un mínimo grado de complejidad.

2.- El promedio (media aritmética) de los números 3; 2; 5; 5 y 6 es

- A) 4 B) 4,2 C) 5 D) 5,25 E) Ninguno de los anteriores.

En la segunda pregunta enfocada a datos no agrupados solo se pide calcular la media aritmética de un conjunto de datos, donde el resultado es directo y su dificultad mínima.

² Ensayo N° 8 el mercurio

3.- El gráfico de la figura representa la distribución de las notas obtenidas por 15 niños en una prueba. ¿Cuál(es) de las siguientes proposiciones es (son) verdadera(s)?

I) 9 niños obtuvieron notas mayores o iguales a 5.

II) La moda es la nota 5.

III) La quinta parte de los niños obtuvo nota inferior a 4.

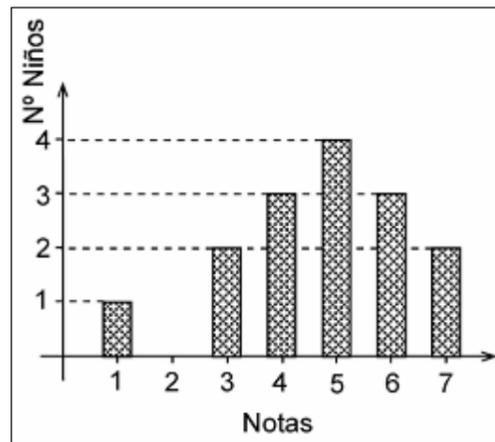


Figura 1: Gráfico pregunta tres

- A) Sólo I B) Sólo C) Sólo III D) Sólo I y III E) I, II y III

En la pregunta tres se muestra un gráfico que representa las notas obtenidas por una cierta cantidad de niños, y esta dirigida a la comprensión, análisis e interpretación de datos.

4.- La tabla adjunta muestra la distribución de los sueldos de 45 personas de una empresa. ¿Cuál(es) de las siguientes afirmaciones es (son) verdadera(s)?

Tramo	Nº de personas	Sueldo en pesos
A	3	5.000.000 - 7.000.000
B	2	2.000.000 - 3.000.000
C	5	800.000 - 1.200.000
D	15	500.000 - 700.000
E	13	300.000 - 400.000
F	7	150.000 - 250.000

Tabla 2: Pregunta cuatro, tipo PSU

- I) Hay exactamente 20 personas que ganan a lo menos \$400.000 de sueldo.
- II) La mediana de la distribución se encuentra en el tramo D.
- III) El total que se paga a las personas del tramo A es, a lo más, \$21.000.000.

A) Sólo I B) Sólo II C) Sólo I y II D) Sólo I y III E) Sólo II y III

Esta pregunta esta orientada a la parte de variables continuas, en donde se entrega una tabla de distribución de sueldos y se pide calcular mediana, además comprender la respuesta numérica, aquí se aprecia mayor complejidad al igual que en la pregunta tres a diferencia de las preguntas uno y dos.

II.3. Secuencia Didáctica

Las secuencias didácticas quedan establecidas por el orden en que se presentan las actividades a través de las cuales se lleva a cabo el proceso de enseñanza- aprendizaje. El énfasis entonces está en la sucesión de las actividades, y no en las actividades en sí.

La secuencia didáctica implicará entonces una sucesión deliberada (planificada) de actividades (es decir un orden), las que serán desarrolladas en un determinado período de tiempo (con un ritmo).

De acuerdo con Zabala Vidilla, las actividades de las secuencias didácticas deberían tener en cuenta los siguientes aspectos esenciales o propósitos generales:

- Indagar acerca del conocimiento previo de los alumnos y comprobar que su nivel sea adecuado al desarrollo de los nuevos conocimientos.
- Asegurarse que los contenidos sean significativos y funcionales y que representen un reto o desafío aceptable.
- Que promuevan la actividad mental y la construcción de nuevas relaciones conceptuales.
- Que estimulen la autoestima y el auto-concepto.
- De ser posible, que posibiliten la autonomía y la meta-cognición.

II.4. Planilla Electrónica u Hoja De Cálculo

II.4.1. Definición

Según Wikipedia³ *“Una hoja de cálculo es un programa que permite manipular datos numéricos y alfanuméricos dispuestos en forma de tablas (la cual es la unión de filas y columnas). Habitualmente es posible realizar cálculos complejos con fórmulas y funciones y dibujar distintos tipos de gráficas”.*

II.4.2. Usos y aplicaciones

Debido a la versatilidad de las hojas de cálculo modernas, estas se utilizan a veces para hacer pequeñas bases de datos, informes, gráficos estadísticos, clasificaciones de datos, entre otros usos. Las operaciones más frecuentes se basan en cálculos entre celdas, las cuales son referenciadas respectivamente mediante la letra de la columna y el número de la fila.

II.4.3. Orígenes

En 1961 se vislumbró el concepto de una hoja de cálculo electrónica en el artículo *Budgeting Models and System Simulation* de Richard Mattessich. Pardo y Landau intentaron patentar algunos de los algoritmos en 1970. La patente no fue concedida por la oficina de patentes por ser una invención puramente matemática. Pardo y Landau ganaron un caso en la corte estableciendo que "algo no deja de ser patentable solamente porque el punto de la novedad es un algoritmo". Este caso ayudó al comienzo de las patentes de software.

Dan Bricklin es el inventor universalmente aceptado de las hojas de cálculo. Bricklin contó la historia de un profesor de la universidad que hizo una tabla de cálculos en un tablero. Cuando el profesor encontró un error, tuvo que borrar y reescribir una gran cantidad de pasos de forma muy tediosa, impulsando a Bricklin a pensar que podría replicar el proceso en un computador, usando el paradigma tablero/hoja de cálculo para ver los resultados de las fórmulas que intervenían en el proceso.

Su idea se convirtió en VisiCalc, la primera hoja de cálculo, y la "aplicación fundamental" que hizo facilitar el trabajo a la hora de realizar cálculos con una gran cantidad de datos.

³ www.wikipedia.com

Posteriormente las hojas de cálculo se fueron desarrollando y adquiriendo cada vez mas importancia en la sociedad, siendo incluso fundamentales para las empresas, lo que masifico su uso, llegando a tener una gran variedad de hojas de calculo y a su vez una gran demanda en el mercado.

II.4.4. Hojas de cálculo en el mercado

En la actualidad en el mercado se cuenta con una gran gama de accesorios para computadores que hacen de nuestra vida más llevadera, facilitándonos el trabajo en diversos aspectos, uno de ellos como señalamos anteriormente son las hojas de cálculo, algunas de ellas se presentaran a continuación:

- **Lotus 1-2-3**

Es un clásico programa de planilla u hoja de cálculo creado o desarrollado por la empresa Lotus Development Corporation (que sería adquirida por IBM en 1996).

Lotus es el primer programa de hoja de cálculo en introducir rangos de celdas, macros para las planillas y celdas con nombres. Con Lotus 1-2-3 es más fácil la utilización de planillas de cálculo y agrega la posibilidad de hacer gráficos y bases de datos.

Lotus 1-2-3 es aún uno de los paquetes de aplicaciones más vendido de todos los tiempos.

La inmensa popularidad que logró alcanzar a mediados de la década de 1980 contribuyó significativamente a afianzar el éxito de los computadores dentro del ambiente corporativo y de oficina.

- **Calc**

Es una hoja de cálculo similar a Microsoft Excel, con un rango de características más o menos equivalente. Su tamaño es mucho menor y proporciona un número de características no presentes en Excel, incluyendo un sistema que automáticamente define series para representar gráficamente basado en la disposición de los datos del usuario.

Esta hoja de cálculo también puede abrir y guardar las hojas de cálculo en el formato de archivos de Microsoft Excel. El formato por defecto de OpenOffice.org 2.0 Calc se puede fijar para que sea el de Microsoft Excel, o el formato Open Document Format (ODF) de la organización Oasis. También apoya una amplia gama de otros formatos, tanto para abrir y guardar archivos.

- **Microsoft Excel**

Excel fue originalmente escrito para la Apple Macintosh en 1984 -1985. Excel es una de las primeras planillas de cálculo en utilizar una interfaz gráfica con menús desplegables y la capacidad de clickear utilizando un puntero. Anteriormente se consideraba, que la planilla de cálculo Excel era para mucha gente más fácil que usar que el comando de interfaz de línea de los productos de PC-DOS.

La planilla de calculo Microsoft Office Excel incluye una variedad de herramientas y formulas, además de su gran ventaja para facilitar la tarea de realizar cálculos tediosos.

Microsoft Office Excel, desarrollado y distribuido por Microsoft Corporation, es una aplicación de hojas de cálculo que se utiliza en variadas tareas, entre las que destacan financieras y contables.

Capítulo III

Marco Metodológico

La orientación de esta investigación es de tipo cualitativa, ya que cumple con características de este tipo de enfoque, además de haber utilizados instrumentos de recolección de tipo cualitativo y se utilizaron fuentes documentales como textos de estudio, para poder comprender este tipo de investigación es necesario ver el significado en su totalidad de este enfoque.

“El enfoque cualitativo, a veces referido como investigación naturalista, fenomenológica, interpretativa o etnográfica, es una especie de “paraguas” en el cual se incluye una variedad de concepciones, versiones, técnicas y estudios no cuantitativos (Grinnell, 1997)”⁴

En definitiva la investigación cualitativa *“es aquella donde se estudia la calidad de las actividades, relaciones, asuntos, medios, materiales o instrumentos en una determinada situación o problema. La misma procura por lograr una descripción holística, esto es, que intenta analizar exhaustivamente, con sumo detalle, un asunto o actividad en particular”⁵*.

El enfoque cualitativo utiliza la recolección de datos sin medición numérica para descubrir o afinar preguntas de investigación en el proceso de interpretación, pero para profundizar más en el significado de este enfoque en particular se debe indagar aún más.

Es por esto que para ahondar en este tipo de enfoque investigativo, veremos algunas de sus características principales, según Fraenkel y Wallen (1996) presenta cinco características básicas que describen las particularidades de este tipo de estudio.

- El ambiente natural y el contexto que se da el asunto o problema es la fuente directa y primaria, y la labor del investigador constituye ser el instrumento clave en la investigación.
- La recolección de los datos es mayormente verbal que cuantitativa.

⁴ Metodología de la Investigación, editorial Mc Graw Hill.

⁵ Dr. Lamberto Vera Vélez, UIPR, Ponce, P.R.

- Los investigadores enfatizan tanto los procesos como los resultados.
- El análisis de los datos se da más de modo inductivo.
- Se interesa mucho saber cómo los sujetos piensan y que significado poseen sus perspectivas en el asunto que se investiga.

Para llevar a cabo una investigación cualitativa se debe tener en cuenta el proceso investigativo que involucra a dicho enfoque, aunque no difiere mucho de los otros tipos de investigación hay algunas particularidades que deben ser consideradas según (Fraenkel y Wallen, 1996), estas son las siguientes:

La identificación del problema a investigar: No estricto a unas variables específicas, el mismo problema o asunto se reformula a medida que se lleva la investigación en sus inicios.

La identificación de los participantes: El cual generalmente es una muestra seleccionada, no aleatoria, ya que el investigador procura por una muestra que concierne más a los propósitos específicos de la investigación.

La formulación de hipótesis: Contrario a los estudios cuantitativos, las hipótesis no se formulan al inicio de la investigación, sino más bien que surgen a medida que se lleva a cabo la investigación. Las mismas pueden ser modificadas, o surgen nuevas o descartadas en el proceso.

La colección de los datos: Los cuales no se someten a análisis estadísticos (si algunos mínimos, tales como por cientos.)

El análisis de los datos: Es mayormente de síntesis e integración de la información que se obtiene de diversos instrumentos y medios de observación. Prepondera más un análisis descriptivo coherente que pretende lograr una interpretación minuciosa y detallada del asunto o problema de investigación.

Las Conclusiones: Estas se derivan o se infieren continuamente durante el proceso. Contrario a los estudios de índole cuantitativas que resultan al final de la investigación, en el estudio cualitativo se reformulan a medida que se vaya interpretando los datos en cuestión.

En el enfoque cualitativo también existen distintos tipos de investigaciones que son necesarias conocerlas para la contracción de la propuesta, es por esto que las nombraremos a continuación.

La primera de ella es la, **Observación participativa**, donde el investigador participa dentro de la situación o problema que se vaya a investigar.

También es aceptable que el investigador sea reconocido de antemano en una investigación, por ejemplo si se desea saber como se da el proceso de enseñanza en una sala de clase en una determinada materia y se quiere estar en la clase como participante y a la vez haciendo las observaciones e interpretaciones pertinentes al estudio.

La segunda es, **Observación no-participativa**, en donde el investigador observa y toma datos.

Y por ultimo la **Investigación etnográfica**, la cual combina tanto los métodos de observación participativa como las no participativas con el propósito de lograr una descripción e interpretación holística del asunto o problema a investigar.

En el enfoque cualitativo también existen tipos de métodos de recolección de información, sus usos son para analizar cualitativamente que quiere decir esto, “Como y por qué”, algunos ejemplos de estos tipos de instrumentos son, entrevistas libres y dirigidas, encuestas utilizando preguntas de respuestas abiertas, observaciones, interpretación de documentos, etc. Algunas de las ventajas de estos instrumentos es que permiten conocer las actitudes, creencias, motivos y comportamientos de una pequeña muestra de la población además de que establecen información de referencia que se puede utilizar para evaluar las conclusiones cualitativas como por ejemplo cambios en cuanto al conocimiento, actitudes, comportamientos, etc.

Es así como se construye este seminario, primero que todo en relación a las distintas bibliografías revisadas, entre ellas, tres libros de matemáticas para el nivel NM4⁶, en donde se realizó un análisis para conocer las diferentes secuencias didácticas utilizadas, luego se realizó una revisión de las planillas que serian aptas para relacionarlas con estadística descriptiva y así también factibles para los usuarios, en donde se compararon 3 hojas de cálculo, el siguiente paso, definir la

⁶ El análisis de los libros se encuentra en: Capítulo IV: Desarrollo, Análisis de Textos

secuencia didáctica presente en la propuesta además de construir las diversas actividades a trabajar.

Finalmente se da origen a la construcción de la guía tomando en cuenta los diversos análisis y recolección de información previos al diseño.

Realizada la propuesta, se prueba en dos colegios de Puente Alto, siete cursos de aproximadamente 40 a 45 alumnos, a la vez se realiza un cuestionario con preguntas cerradas y abiertas para conocer la apreciación de las actividades realizadas.

Capítulo IV

Presentación de Resultados

En este capítulo se explicará como se construyó la guía práctica de estadística descriptiva, sus diferentes pasos, el orden que tiene cada contenido y actividad, y por último la implementación de una de las actividades.

El capítulo se divide en tres apartados, estos son, Análisis, Construcción e Implementación, el primero se subdivide en dos, uno es el análisis de textos y el otro es el análisis de planillas de cálculo. El segundo apartado se desglosa en, secuencia didáctica, donde se define la secuencia a utilizar y construcción de actividades. Por último se presenta la implementación de una de las actividades que se construyó para conocer la efectividad de este.

IV.1. Análisis

En este subcapítulo se abarcarán dos puntos importantes en la elaboración de la propuesta en cuestión, en orden prioritario, el primer apartado será análisis de textos mediante los cuales se realizará la red de contenidos y la elección de actividades que se abordarán en la unidad de la propuesta.

El segundo apartado será la selección de hoja de cálculo, segundo paso importante en la elaboración de la propuesta.

IV.1.1. Análisis de textos

Para realizar esta propuesta, se observó la forma en que se aborda la unidad de estadística en los libros de 4^o medio de matemáticas y principalmente si estos se relacionan con el uso de la planilla de cálculo, tal como lo plantean los planes y programas de dicha unidad. Se realizó un análisis de 3 libros para la construcción de esta.

Los libros utilizados para este análisis son los siguientes:

	Matemática	Matemática	Matemática
	4º Medio	4º Medio	4º Medio
Edito-rial	Mare Nostrum	Zig – Zag	Santillana
Años	2003 – 2004	2005 – 2006	2007 – 2008
Autores	- Jorge Soto Andrade - Patricio González G.	- Roberto Hojman G. - Luís Huerta T. - Jorge Yutronic F.	- Alejandro Pedreros M. - Ángela Baeza P. - María Villena R. - Pablo Jonquera R. - Gabriel Moreno R.

Tabla 3: Libros análisis de texto

Se confeccionó una tabla de doble entrada, para verificar los contenidos que contiene cada libro. Se asignará lo siguiente en la tabla:

1: Esta presente

0: No esta presente

Contenidos estadística	Santillana	Mare Nostrum	Zig – Zag
Descriptiva			
Historia de la estadística	1	0	0
- Ejercicios	1	0	0
Conceptos básicos:			
- Población	1	1	1
- Muestra	1	0	0
- Variables cualitativas	1	0	1
- Variables cuantitativas	1	0	1
- Ejercicios	1	0	1
Frecuencia absoluta	1	1	1
Frecuencia relativa	1	1	1
- Ejemplos	1	1	1
Construcción tabla de frecuencias	1	0	1
Ejemplos	1	0	1

Ejercicios	1	0	1
Representaciones gráficas			
- Diagrama de tallo y hoja	1	1	1
- Histograma	1	1	1
- Gráfico circular	1	1	0
- Pictograma	1	0	0
- Gráfico de barras	1	1	0
- Gráfico de dispersión	1	0	0
- Ejemplos	1	1	0
- Ejercicios	1	1	0
Uso del computador	1	1	0
- Ejemplo	1	1	0
- Ejercicios	1	1	0
Medidas de tendencia central			
Media aritmética	1	1	1
Media aritmética ponderada	1	0	0
- Ejemplos	1	1	1
Mediana	1	1	1
- Ejemplos	1	1	1
Moda	1	1	1
- Ejemplo	0	1	1
Ejercicios	1	1	1
Medidas de dispersión			
- Rango	1	0	1
- Desviación media	1	0	1
- Varianza	0	0	1
- Desviación estándar o típica	1	1	1
Ejercicios	1	0	1
Medidas de localización			
- Cuartil	1	1	1
- Percentil	1	0	1
- Decil	1	0	1
Ejercicios	1	0	1
Diagrama de cajas	1	1	0
- Ejemplo	1	1	0

Tabla 4: Contenidos presentes en cada libro

Analizando los resultados de la tabla anterior se puede observar, que en solo cinco contenidos de los tres libros existe similitud en cuanto a los puntos abordados, como por ejemplo, en el contenido "Conceptos básicos", en los puntos, población, frecuencia absoluta y relativa, además de ejemplos, los tres libros concuerdan con esto.

En el contenido “Representaciones gráficas”, existe semejanza de los tres, en diagrama de tallo y hoja e Histograma, en el contenido “Medidas de tendencia central”, media aritmética, mediana, moda y ejemplos de cada uno de estos temas, en el contenido “Medidas de dispersión” existe similitud, solo en la Desviación estándar o típica y por último en el contenido “Medidas de localización” los libros concuerdan solo con el tema cuartiles, mientras que los otros temas están abarcados en un libro pero no en los otros.

Al observar el cuadro comparativo se puede concluir claramente que el libro que cumple con la mayor cantidad de características positivas es el libro Santillana entregado por el ministerio de educación.

A la vez realizamos ciertas observaciones de cada texto escolar, donde según el criterio del grupo de investigación, son importantes mencionar.

Textos escolares	Observaciones
Santillana	Cada gráfico se presenta con un ejemplo y su respectiva imagen, no se explica en que consiste cada gráfico.
	En el uso del computador se utiliza la planilla Excel, donde en forma general se explica como seleccionar los datos, y la construcción del gráfico de barras
	Presenta mediana y moda para solo datos no agrupados
Mare Nostrum	Explica en 4 pasos como introducir los datos en la planilla Excel, y la construcción del gráfico de barras.
	Se hace hincapié en la diferencia entre mediana y esperanza
	Se presenta la ecuación de equilibrio para explicar la desviación
Zig – Zag	El texto empieza con el uso de la planilla Excel, con pasos bastante claros para utilizar la planilla, se construye el gráfico de barras y circular, con todos los pasos necesarios.
	El orden de la materia es bastante confusa, en comparación, con el libro Santillana

Tabla 5: Observaciones libros análisis de textos

Se realizó este análisis, ya que consideramos que los textos escolares son fundamentales para el proceso de enseñanza-aprendizaje, tanto para el profesor como para el estudiante, este permite ahondar en los conceptos fundamentales de un subsector determinado. Si bien los libros tienen varias diferencias, todos deben cumplir con los requisitos exigidos por el ministerio de educación al abordar los objetivos fundamentales y contenidos mínimos obligatorios, en este caso, del subsector de matemáticas para el nivel NM4 y también los contenidos y actividades cumplen ciertas condiciones, que tanto los textos como la propuesta satisfacen.

Una de las principales, es abarcar todos los contenidos propuestos por los planes y programas del nivel NM4 que respectan a la unidad de estadística descriptiva además de incorporar diversas actividades para ser desarrolladas en Excel.

Estos contenidos a su vez fueron tratados rigurosamente, ya que al tener un error se puede confundir fácilmente el concepto, es por eso su rigurosidad.

En cada sección los contenidos son jerarquizados, se deben tener ciertos conocimientos para poder abordar el siguiente, y por último, cada contenido se relaciona con la cotidianidad, para ser más liviano el trato de este, explicando cada contenido a cabalidad.

Respecto a las actividades construidas por el grupo de investigación, estas fueron creadas con el principal motivo de incentivar y motivar al estudiante a aplicar los contenidos de estadística descriptiva aprendidos, otro punto importante es que para poder realizar estas actividades fue trascendental que cada actividad tuviera sus respectivas instrucciones para que el estudiante no se forme ningún tipo de duda. Por último las actividades tienen sus respectivas soluciones y están ordenadas según nivel de complejidad (de menor a mayor).

IV.1.2. Selección planilla de calculo

A continuación se presenta una tabla de comparación de tres tipos de hojas de cálculo; Microsoft Excel, Lotus 1-2-3 y Calc, la cual considera siete características que se evaluarán mediante un **si**, si presenta dicha característica o un **no**, si no presenta dicha característica, esta tabla de comparación será la que determinará que hoja de cálculo se utilizará para la confección de la propuesta que se quiere llevar a cabo. Observemos el siguiente cuadro:

Cuadro comparativo Microsoft Excel, Lotus 1-2-3 y Calc

	Microsoft Excel 	Lotus 1-2-3 	Calc 
Perfeccionamiento de profesores proyecto “Enlace”	Si	No	No
Mayor accesibilidad	Si	No	Si
Nuevas versiones	Si	Si	Si
Reconocimiento de usuarios	Si	No	No
Presente en la mayoría de los PCs de estudiantes Chilenos	Si	No	No
Creación de líneas de tendencia central y/o análisis de regresión polinómicas	Si	No	No
Asistente para funciones avanzadas	Si	No	No

Tabla 6: Análisis hojas de calculo

Realizado el análisis con estas tres hojas de calculo, se decidió por la utilización de la planilla Excel, ya que para el tipo de usuario al que va enfocada esta guía, es el programa de hoja de cálculo que mas se conoce o con el que la gran mayoría de las personas se encuentra mas familiarizado. Si bien es cierto, existe una gran cantidad de programas que hacen las veces de hojas de cálculo, como las ya mencionadas, Microsoft Office es la más indicada para colaborar con la enseñanza de la estadística descriptiva por la cantidad de características que reúne en desmedro de las otras dos, Lotus 1-2-3 y Calc.

Otro factor importante de la elección de esta hoja de cálculo, es que al implementar una hoja de calculo extraña para los alumnos se corre el riesgo de causar una reacción contraria, es decir una reacción de inseguridad al utilizarla.

También otro factor importantísimo a la hora de discriminar, es la interfaz grafica que ofrece Excel ya que es muy amigable para los usuarios, ofrece la opción de caracterizar las celdas, cambiar formato y color de letras, insertar formulas, realizar gráficos, estimar tendencias, añadir imágenes, etc.

Mucha gente conoce Excel pero solo lo utilizan para que los datos les queden alineados y desaprovechan las múltiples aplicaciones que se pueden obtener.

En conclusión, Excel es de las más completas herramientas en lo que a hojas de cálculo respecta, presenta cómoda interfaz gráfica, es el más utilizado por los usuarios al menos en el Sur de América, en Chile se utiliza Microsoft Office casi de manera estándar, sin dejar de lado lo más importante aún, que su utilización está considerada en los planes y programas estipulados por el ministerio de educación Chilena no de manera explícita, pero si se hace explícita con el perfeccionamiento de docentes a través del proyecto "Enlace" donde cada año se instruye a docentes con variadas aplicaciones de Office como Word, PowerPoint, Publisher, etc. y una de ellas también Microsoft Office, es este motivo de gran peso que lleva a cumplir con la finalidad de crear actividades de apoyo mediante Microsoft Excel.

IV.2. Construcción de Guía Práctica

Este segundo subcapítulo, construcción de guía práctica, se dividirá en dos apartados, el primero Definición de secuencia didáctica, donde se define el cuerpo de la guía y segundo, Construcción de actividades.

IV.2.1. Definición Secuencia Didáctica

Para realizar esta guía de estadística descriptiva, se planificó una secuencia didáctica que es la que se seguirá en cada sección. La secuencia construida es la siguiente:

1.- Exposición del concepto

Se formaliza el concepto, se expone de forma clara, para que el estudiante lo interiorice.

2.- Presentación de una situación o problema

Para interiorizar el contenido se plantea un problema a través de un ejemplo, donde se invita a buscar la posible solución.

3.- Búsqueda de solución

Se soluciona el ejemplo que se planteo, se explica paso a paso la forma de solucionar el problema, utilizando el concepto tratado.

4.- Ejercitación

Se refuerza el aprendizaje en el programa computacional Excel, dando las instrucciones necesarias para la utilización de este, y a la vez un video con la realización de cada uno de los ejercicios planteados, luego se plantean actividades para ejercitar los contenidos tratados, donde se hacen preguntas reflexivas para que los estudiantes obtengan un verdadero aprendizaje significativo, donde se tiene la opción de resolver las actividades bajo la herramienta computacional o tradicionalmente.

En la unidad hay dos excepciones, donde la secuencia no es la antes dicha, primero en el punto *1.- La estadística*⁷, donde se aborda la historia y los conceptos básicos, no se refuerza el contenido en Excel, por no tener relación para el uso de la planilla, y segundo en la sección *2.- Recopilación de datos, variables y frecuencias*⁸ se cambia el orden de la secuencia, ya que era imprescindible que a través de un problema se buscara la posible solución, la secuencia seguida es la siguiente:

1.- Presentación de una situación o problema

Para empezar los contenidos, se plantea un problema para que se vaya buscando una posible solución.

2.- Búsqueda de solución

Se soluciona el ejemplo que se planteo, se explica paso a paso la forma de solucionar el problema acercándose al concepto buscado.

3.- Exposición del concepto

Se formaliza el concepto, se expone de forma clara, para que el estudiante lo interiorice

⁷ Guía Propuesta: Pág. 56

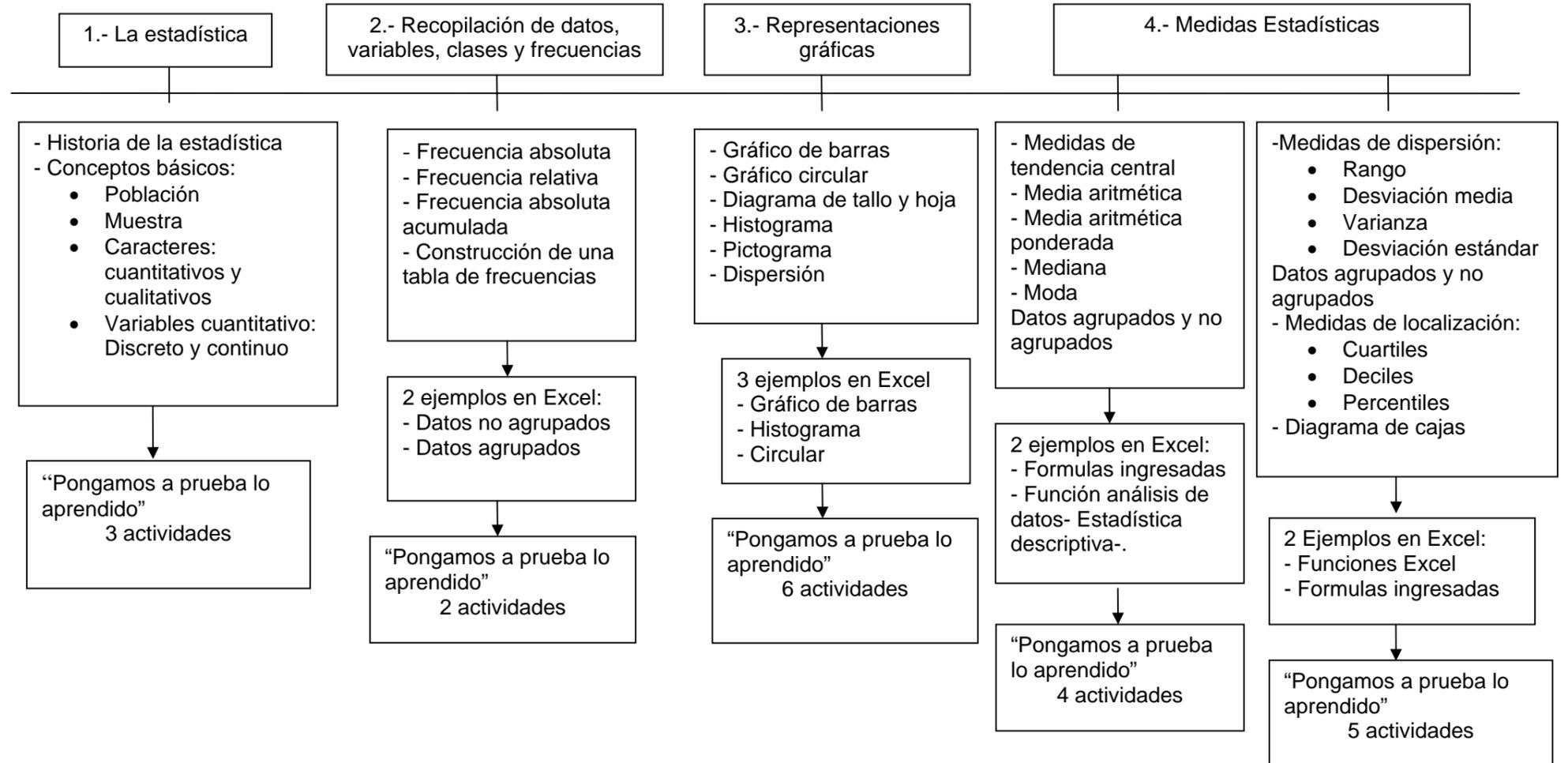
⁸ Guía Propuesta: Pág. 64

4.- Ejercitación

De la misma forma ya dicha anteriormente.

La forma de la unidad de estadística descriptiva propuesta se presenta en el siguiente cuadro resumen.

Unidad estadística descriptiva para el nivel NM4



IV.2.2. Construcción actividades

En este apartado se muestra de manera detallada la construcción de las actividades de la propuesta en cuestión, a continuación presentamos la secuencia de la construcción de la guía práctica.

La propuesta de esta guía práctica, se basa en todo lo que corresponde a estadística descriptiva para el nivel NM4, se organizó de acuerdo a los siguientes ejes:

- La historia de la estadística, donde se hace especial mención a Jhon Graunt, luego se infiere para que nos sirve la estadística., es importante conocer el origen para entender su función y utilidad. Continúa con *Conceptos básicos*: población, muestra, caracteres, modalidades, donde se muestran ejemplos de caracteres cuantitativos y cualitativos, también se muestran ejemplos de variables cuantitativos; discretos y continuos.

- A continuación con la sección “Pongamos a prueba lo aprendido”, se ejercita cada uno de los contenidos tratados con preguntas reflexivas, son 3 actividades, una de ellas, respecto a lo que es población y muestra, es la siguiente:

Interesa conocer la aceptación de un nuevo juguete dirigido a niños entre 6 y 10 años. ¿Cual crees tú que es la muestra más representativa?

- a. Ir a un parque de diversiones y realizar una pequeña encuesta a los niños presentes

- b. Conseguir el listado de 500 niños entre esas edades, que asisten a un colegio y mandar 200 encuestas.

- c. En una esquina del centro de la ciudad realizar una encuesta a niños que pasen por ahí.

- Continúa la siguiente sección: “**Recopilación de datos, variables, clases y frecuencias**”, se inicia con frecuencia absoluta (para ordenar los datos), luego se recuerda la sumatoria, para posteriormente

deducir la formula, sigue con frecuencia relativa, frecuencia absoluta acumulada y finaliza con construcción de una tabla de frecuencias.

- Se ejercitan los contenidos en el computador, con las instrucciones necesarias en Excel, donde cada frecuencia se va uniendo con preguntas, como lo siguiente: “¿Cuántas familias consumen menos de 5 latas al mes?. Para contestar a esta pregunta, podemos construir otra columna en nuestra tabla⁹”. Esta columna es la columna de frecuencias absolutas acumuladas, luego se propone una actividad con el uso de la planilla Excel, con su respectiva solución.

- En la sección “Pongamos a prueba lo aprendido” se presentan dos actividades, donde los estudiantes tienen las herramientas necesarias para resolverlas, con preguntas reflexivas.

- Luego la sección “**Representaciones gráficas**”, donde se abarcan los siguientes gráficos:

-diagrama de tallo y hoja (Se visualizan las diferencias de los datos, es básico para empezar)

-gráfico de barras	}	El primero se utiliza con las frecuencias absolutas y el circular con las frecuencias relativas, siguiendo el orden de las frecuencias ya vistas.
-gráfico circular		

-histograma	}	Son gráficos más complejos y a la vez los más frecuentes.
-pictograma		
-dispersión		

Se explica la utilidad de cada uno y con un ejemplo se va representando.

- Luego en Excel se explica como hacer el gráfico de barras, circular e histograma, Excel trae las funciones para estos, sin embargo en el pictograma se necesita creatividad, se puede utilizar Excel pero depende de lo que se quiera mostrar con el gráfico, y en el de dispersión, Excel tiene la

⁹ Guía Propuesta: Ejemplo Excel 1, Pág. 70

opción pero para dos variables, y el gráfico explicado es solo para una variable. Luego una propuesta de trabajo para ejercitar en Excel, con cada uno de los gráficos enseñados (de barras, circular, histograma)

- En la sección “Pongamos a prueba lo aprendido”, en cada actividad se reflexiona frente a que gráfico sería el correspondiente para cada caso, y se pregunta por conclusiones obtenidas visualizando el gráfico.

- Continúa la sección: “**Medidas estadísticas**”, donde las principales son las medidas de tendencia central y las medidas de dispersión.

- Medidas de tendencia central: Media aritmética, media aritmética ponderada, mediana, moda, tanto para datos agrupados y no agrupados. La principal medida de tendencia central es la media, al entender esta medida se da el paso para entender la mediana, donde se explica claramente la fórmula para datos agrupados, explicando cada componente de esta, y por último la moda.

- En la planilla Excel se ejercitan las medidas de tendencia central, de dos formas, una ingresando las respectivas fórmulas y segundo utilizando la función estadística descriptiva de la planilla.

- Luego “Pongamos a prueba lo aprendido”, donde se abarcan las medidas de tendencia central en su cabalidad.

- Medidas de dispersión: rango, desviación media, varianza, desviación estándar, para datos agrupados y no agrupados. Para entender la desviación se comienza con un ejemplo, donde dos grupos de datos tienen la misma media, entonces la pregunta es ¿Cuál de los dos grupos obtuvo un mejor rendimiento?, dando pie a las medidas de dispersión, para entender la desviación, se necesita entender el rango o recorrido y la varianza.

- Medidas de localización: cuartiles, deciles, percentiles. Solo se tratan estas medidas, ya que no se encontró necesario explicar otros como quintiles, los principales son los mencionados, y el entendimiento de este aclara las medidas de localización.

- En Excel se sigue ejercitando las medidas de dispersión y localización, se explica ingresando las formulas explicadas, y con la función de estadística descriptiva, se realizan conclusiones como: *“el Tercer año B presenta una mayor dispersión promedio que el Tercer año A, lo que nos llevaría a concluir que en general el Tercer año A tuvo un desempeño más homogéneo en el test que el Tercer año B”*. También se hacen preguntas para entender el significado, como: *“¿Qué significa que el valor Q1 sea igual a 4.9? Significa que aproximadamente el 25% de los estudiantes tuvieron notas inferiores a 4.9.”*

- Diagrama de cajas, Para la explicación de este tipo de diagrama, primero que todo se realiza una explicación de su utilización comparativa de diversos elementos de distribución, luego de esto siguiendo inmediatamente con un ejemplo de variables continuas presentando los datos en una tabla de intervalos, donde se calcula mediana y medidas de localización, terminando con el grafico del diagrama de las medidas calculadas.

- Luego “Pongamos a prueba lo aprendido”, donde se ejercitan las medidas de dispersión y localización.

IV.3. Implementación Actividad

Con el fin de conocer la efectividad de la guía práctica, se implementó una de las actividades a siete cuartos medios de dos colegios en la comuna de Puente Alto, a continuación se muestra una tabla con los datos de los colegios en que se realizaron las actividades:

Colegios / Datos	Comuna	Grupo Socioeconómico	Dependencia
La Consolidada	Puente Alto	Medio	Municipal
Saltairam	Puente Alto	Medio- Alto	Particular Subvencionado

Tabla 7: Datos colegios implementación actividad

Se decidió realizar la actividad respecto a la media aritmética, por lo tanto, primero se confecciono una pequeña prueba para medir el conocimiento de este, posteriormente se implemento la actividad.

Los datos de la fecha en que se realizó la prueba, las actividades, los cursos, los alumnos, se resumen en la siguiente tabla:

Colegios	La Consolidada				Saltairam		
	4º A	4º B	4º C	4º E	4º A	4º B	4º C
Datos/Cursos							
Cantidad de alumnos	42	41	44	44	39	42	41
Fecha aplicación Prueba "Media Aritmética"	29 de julio	30 de julio	29 de julio	29 de julio	29 de julio	30 de julio	30de julio
Cantidad de alumnos que responden Prueba	40	39	40	41	39	38	40
Fecha implementación actividad	30 de julio	31 de julio	30 de julio	30 de julio	30 de julio	31 de julio	31de julio
Cantidad de alumnos que realizan actividad	40	38	42	42	38	40	39

Tabla 8: Datos implementación actividad

Como se observa en la tabla la implementación se desglosa en tres, primero los resultados de la prueba "Media Aritmética", luego la implementación de la actividad que se subdivide en dos, etapas y resultados actividades, por último el cuestionario realizado para conocer las impresiones de los estudiantes respecto las actividades realizadas.

IV.3.1. Prueba Media Aritmética

La prueba realizada tiene el espacio respectivo para las respuestas y a la vez dos opciones, no conozco, no recuerdo, el fin de la prueba es conocer la situación de los estudiantes respecto lo que es la media aritmética, las preguntas realizadas son las siguientes:

Media Aritmética		
1.- ¿Que es la media aritmética?	No recuerdo	No conozco
2.- ¿Cuál es la media entre los siguientes datos: 2-4-7-4-9-10-3?	No recuerdo	No conozco
3.- ¿Conoces la media para datos agrupados y no agrupados?	No recuerdo	No conozco
4.- ¿Cuál es la formula para definir la media aritmética?	No recuerdo	No conozco
5.- ¿Sabes utilizar la planilla Excel?	No recuerdo	No conozco
6.- ¿Podrías utilizar la planilla Excel para calcular la media de un conjunto de datos? ¿como?	No recuerdo	No conozco

Tabla 9: Preguntas Prueba Media Aritmética

Para el análisis de cada una de las respuestas, se agregó una opción más "sin respuesta", donde en ella entran todos los estudiantes que no respondieron o respondieron más de una alternativa.

Análisis

En la *pregunta 1*, un 33% respondió no conocer la respuesta, y un 35% no recordar, mientras que un 12% de los estudiantes respondió satisfactoriamente y el mismo porcentaje no respondió correctamente, por último un 9% no respondió ninguna de las opciones.

Se puede inferir que un gran porcentaje no recuerda y no conoce, por lo que mas de la mitad no sabe lo que es la media aritmética.

En la *pregunta 2*, el 26% declara no conocer la respuesta, un 23% no recuerda, mientras que un 30% contestó correctamente, y un 13% logro un resultado incorrecto, por ultimo un 9% no contestó ninguna opción.

Se puede concluir que casi la mitad trato de responder, sin embargo solo un 30% logro contestar correctamente.

En la *pregunta 3*, un 76% declara no conocer la media para datos agrupados y no agrupados, un 15% dice no recordar y un 6% no responde, sin embargo 8 estudiantes lograrón acercarse a una respuesta satisfactoria, que corresponde a un 3%.

Un gran porcentaje no conoce la media para datos agrupados y no agrupados, por lo tanto, quizás no conocen los datos agrupados y no agrupados.

En la *pregunta 4*, un 30% no conoce la fórmula, un 31% no recuerda, mientras que un 14% responde correctamente, y un 16% incorrectamente, por ultimo un 9% no responde.

Una gran cantidad de alumnos no sabe la fórmula, sin embargo, en la pregunta 2 un 30% respondió correctamente.

En la *pregunta 5*, un 30% responde positivamente, un 30% declara no conocer, un 35% no recordar y un 5% no responde ninguna opción.

Mas de la mitad de los estudiantes no sabe utilizar Excel, más de 138 estudiantes no saben utilizar una planilla de cálculo.

En la *pregunta 6*, un 65% dice no conocer, un 14% no recuerda y un 6% no contesta, mientras que un 15% responde positivamente, pero no se explica claramente la forma de calcular la media.

Una gran cantidad de estudiantes, no sabe lo que es la media, por lo que no sabe calcular esta en una planilla de cálculo, mientras que el 15% que respondió positivamente, explica: "sumando cada celda, luego dividir por el total".

Se puede concluir que los estudiantes no manejan el concepto de la media aritmética, por lo tanto, la actividad que se implementara en los siete cursos es referida a esta.

A continuación se muestra un cuadro resumen con los datos mostrados anteriormente.

Preg./Datos	No conoce	No Recuerda	No responde	Respuesta satisfactoria o correcta	Respuesta Insatisfactoria o incorrecta
Pregunta 1	91	96	25	32	33
Pregunta 2	71	65	24	82	35
Pregunta 3	210	41	18	8	0
Pregunta 4	84	85	24	39	45
Pregunta 5	83	96	14	84	0
Pregunta 6	181	38	16	42	0

Tabla 10: Resultados prueba Media Aritmética

IV.3.2. Implementación Guía

En los siete cursos se implementó una parte de la guía práctica que contempla la media aritmética, en este momento se llevará a cabo la secuencia didáctica diseñada, con el fin de incorporar la planilla de cálculo. La implementación se desglosa en dos, Datos y Etapas Implementación Guía y los resultados de las actividades realizadas por los estudiantes.

IV.3.2.1. Datos y Etapas Implementación Guía

La actividad es la numero **4.1.1 Media Aritmética**, se trabaja el ejemplo 9, 10 y el ejemplo Excel 7, por ultimo el punto **4.1.5 Pongamos a prueba lo aprendido**, la actividad 1 y 3, con solo las preguntas correspondientes a la media aritmética.

Los aprendizajes previos, esperados y objetivo son los siguientes:

Aprendizajes Previos:

- Frecuencia Absoluta
- Frecuencia Relativa
- Marca de clase
- Tratamiento de datos agrupados y no agrupados

Aprendizajes esperados:

- Entender y analizar la media aritmética
- Utilización de planilla Excel:
 - uso de función: Estadística descriptiva

Objetivo: Conocer la factibilidad del uso de la planilla de calculo Excel y si apoya el conocimiento estadístico del estudiante

Tiempo: 90 minutos

Inicio:

Se le repartió a cada estudiante una copia impresa de la guía, luego se les dio las instrucciones correspondientes y el objetivo de la actividad, para luego proseguir con el desarrollo de esta.

Desarrollo:

Al empezar la actividad se observo que los alumnos estaban muy interesados en realizarla y llevar a cabo la guía, en el transcurso de la clase surgieron muy pocas dudas, pero estas totalmente enfocadas a detalles ínfimos, como por ejemplo: “¿Cómo arrastramos los datos hacia abajo? ¿Cómo vamos profesora?, etc”.

Final:

A los 60 minutos aproximadamente se prosiguió a retirar el desarrollo de las dos actividades, luego se les entrego un cuestionario para conocer sus impresiones.

IV.3.2.2. Resultados Actividades

Los resultados de las actividades realizadas por los estudiantes son los siguientes:

En la Actividad 1: 214 estudiantes respondieron correctamente la pregunta, 50 estudiantes incorrectas y 6 no respondieron la actividad

En la Actividad 2: Los estudiantes que lograron el resultado final (correcto) fueron 231, mientras que 32 contestaron incorrectamente, y 7 simplemente no contestaron.

A continuación se muestra una tabla resumen con los datos dichos anteriormente:

	Respuesta Correcta	Respuesta incorrecta	Sin Respuesta
Actividad 1	79%	19%	2%
Actividad 2	85.56%	11.85%	2.59%

Tabla 11: Resultados Actividades Media Aritmética

Se infiere de los datos, que los estudiantes lograron entender la guía, ya que los resultados correctos resaltan notablemente. Ahora con el fin de conocer las impresiones de la guía y conocer si las actividades las realizaron tradicionalmente o en el computador se hizo entrega del cuestionario que se muestra a continuación:

IV.3.3. Cuestionario

El fin del cuestionario es conocer si se logro el objetivo antes planteado, si se logro un aprendizaje significativo y a la vez sus impresiones respecto la guía practica. Las respuestas posibles es si, no, a veces.

Las preguntas y resultados del cuestionario realizado por los estudiantes, son los siguientes:

1 - ¿Se entendió la guía?

El 78.15% de los estudiantes respondió si, un 8.52% dice que no, un 12.59% a veces y un 0.74% no respondió la pregunta

2.- ¿Están claras las definiciones?

Un 57.41% declara que si, mientras que un 14.81% contesta no, un 26.30% responde a veces y un 1.48% no respondió la pregunta

3.- ¿Son suficientes las instrucciones para el uso de Excel?

El 74.44% dice que si son suficientes, un 9.63% declara que no, un 15.19% a veces y por ultimo un 0.74% no responde.

4.- ¿Fue eficiente el uso de Excel para el ejemplo planteado?

Declara que fue eficiente un 88.89% de los estudiantes, un 5.93% dice que no, un 5.19 declara que solo a veces.

5.- ¿Estas de acuerdo con utilizar la planilla Excel para el aprendizaje de la estadística?

Esta de acuerdo un 83.7%, un 10% dice no estarlo, un 5.19% no esta de acuerdo y un 1.11% no contesta.

6.- ¿Se reforzaron los aprendizajes previos?

Un 78.89% respondió que si, mientras que un 14.07% no, un 5.93% respondió a veces y un 1.11% no respondió.

7.- ¿Resolviste las actividades con el uso de la planilla?

El 54.07% declaro usar la planilla para resolver las actividades, un 44.44% dijo que no, y un 1.48% no respondió.

8.- ¿Utilizarías esta guía para aprender toda la unidad de estadística?

El 82.96% dice que si utilizaría la guía, un 10.37% dijo que no, un 5.93% a veces y un 0.74% no respondió.

9.- ¿Sabes lo que es la media aritmética?

Un 78.89% declara saber lo que es la media aritmética, el 14.44% dice no saber, un 5.19% responde a veces y no respondió un 1.48%

10.- ¿Cuál es la media entre los siguientes datos: 2-4-7-4-9-10-3?

Responde correctamente un 79.26%, un 14.07% respondió incorrectamente y un 6.67% no respondió.

Análisis

Se infiere de la *pregunta uno* que la guía esta bastante clara para el entendimiento de los estudiantes, ya que mas de la mitad contesto positivamente a la pregunta, sin embargo en la *pregunta dos*, solo un poco mas de la mitad dice que están claras las definiciones y un 26.3% declara que solo a veces, donde se infiere que la definición se presto para alguna confusión.

En la *pregunta tres y cuatro*, un gran porcentaje de los estudiantes declara que las instrucciones fueron suficientes para el uso de la planilla, y a la vez eficientes para el ejemplo planteado, lo que da a entender que los estudiantes encontraron que el uso de la planilla es eficiente para el cálculo de operaciones. La *pregunta cinco* confirma lo antes dicho, ya que una gran parte esta de acuerdo, sin embargo esto también puede ser por el entretenimiento de utilizar el computador y cambiar un poco lo que se hace tradicionalmente.

En la *pregunta seis* un gran porcentaje dice que se reforzaron los aprendizajes previos, por lo que se entiende que el uso de la planilla refuerza el aprendizaje ya obtenido.

En la *pregunta siete*, se observa que solo un poco mas de la mitad resolvió las actividades utilizando la planilla, y el resto lo hizo tradicionalmente, lo que es totalmente distinto a lo pensado por el grupo de investigación.

Cabe destacar que en la *pregunta ocho*, se refleja la aceptación de la guía por parte de los estudiantes.

En las *preguntas nueve y diez*, se logra el objetivo, que los estudiantes comprendan lo que es la media aritmética.

A continuación un cuadro resumen con lo dicho anteriormente:

Preg./ Resp.	Si	No	A veces	No responde
Pregunta 1	211	23	34	2
Pregunta 2	155	40	71	4
Pregunta 3	201	26	41	2
Pregunta 4	240	16	14	0
Pregunta 5	226	27	14	3
Pregunta 6	213	38	16	3
Pregunta 7	146	120	0	4
Pregunta 8	224	28	16	2
Pregunta 9	213	39	14	4
	Correcta	Incorrecta	No responde	
Pregunta 10	214	38	18	

Tabla 12: Resultados Cuestionario

Capítulo V

Conclusiones

Al confeccionar la guía práctica se indagó en la forma de poder abordar la materia y en cómo sería más fácil y entretenido para el estudiante, tomando en cuenta que estadística descriptiva se ve el último año de enseñanza media y los estudiantes se preparan para la prueba de selección universitaria, es por esto, que se realizó el análisis de los libros de cuarto medio de matemáticas, donde hay notables diferencias entre las editoriales, resaltando notablemente el libro Santillana, tanto los contenidos como las actividades.

El uso del computador es esencial hoy en día, es por esto que es de gran relevancia que este presente en la secuencia didáctica, y las instrucciones para el uso de este también.

El análisis ya mencionado, facilitó la definición de la secuencia didáctica, incorporando el uso de la planilla de cálculo, donde ninguno de los libros abarca a cabalidad.

La selección de la planilla de cálculo a utilizar se facilitó al comparar las planillas y notar claramente la factibilidad de la planilla Excel.

Una de las principales dificultades al realizar la guía, fue al momento de construir los ejemplos y actividades, ya que cada una debe llamar la atención de los estudiantes, por esto se trataron temas como la obesidad, cesantía, maltratos, etc.

Construida la propuesta, se hizo necesario obtener opiniones acerca de esta, y su real eficacia en la enseñanza de la estadística.

Para obtener conclusiones fehacientes, se implementó una temática de la guía y así conocer las impresiones de esta y resultados.

Para poder realizar la actividad, en este caso, actividad respecto a la media aritmética, se efectuó una pequeña prueba para saber cuáles eran los conocimientos respecto a esta, luego se implementó la guía práctica confeccionada por el grupo de investigación.

Los resultados obtenidos fueron bastante favorables, pero son solo siete cursos, en total eran 270 alumnos, para poder abarcar mas estudiantes se necesita más tiempo y para implementar más actividades también.

Son varios los factores que pueden influenciar al momento de probar una actividad, los estudiantes pueden haber adquirido nuevos conocimientos, aunque esta prueba se realizo de un día a otro, o las condiciones del día en que implementamos la actividad nos favoreció, sin embargo, los resultados son bastantes positivos y superaron las expectativas totalmente.

La guía practica propuesta va enfocada totalmente a los estudiantes, sin embargo, esta a libre opción de ser utilizada por profesores para sus clases de estadística descriptiva, cada ejemplo, cada contenido fue revisado meticulosamente, y también las instrucciones para el uso de la planilla Excel.

Se espera en un futuro llevar a cabo la guía practica propuesta, y como profesoras obtener una apreciación mas clara respecto a su efectividad.

El seminario fortalece los aprendizajes adquiridos durante la carrera, tanto como especialidad y sobretodo pedagógicos, ya que la forma de abordar cada temática, es un desafío pedagógico para que se logre verdaderamente los objetivos del proceso enseñanza-aprendizaje.

Bibliografía

- [1] MINEDUC (2001). *Programa de estudio, cuarto año medio: Educación Matemática*. Chile: MINEDUC.
- [2] Del Pino Guido; Iglesias Pilar y Aravena Ricardo. *Análisis Estadístico; Interpretando problemas de la vida cotidiana*. Chile: MINEDUC.
- [3] Carreño Ximena y Cruz Ximena. *Matemática IV medio*. Editorial: Arrayán.
- [4] Pedreros Alejandro, Baeza Ángela, Villena María, Jorquera Pablo y Moreno Gabriel. (2007-2008). *Educación Matemática, cuarto medio*. Editorial: Santillana
- [5] Hojman Roberto, Huerta Luís y Yutronic Jorge. (2005-2006). *Matemática, cuarto medio*. Editorial: Zig-Zag
- [6] Soto Jorge y González Patricio. (2003-2004). *Cuarto medio, Educación Matemática*. Editorial: Mare Nostrum
- [7] Hernández Roberto, Fernández Carlos y Baptista Pilar. (1995). *Metodología de la Investigación*. Editorial: Mc Graw Hill
- [8] <http://es.wikipedia.org>
Última Visita: 12 de Julio del año 2009
- [9] <http://www.sectormatematica.cl/psu.htm>
Última Visita: 12 de Julio del año 2009
- [10] <http://www.educarchile.cl>
Última Visita: 12 de Julio del año 2009

Anexos

4^o Educación Media

GUIA PARA EL ESTUDIANTE
Incluye CD video de trabajo en Excel

Educación Matemática

"Estadística Descriptiva"

Tonya Jorquera Utreras
Claudia Grandón Valdés

Edición especial para el estudiante

Índice

1.- La estadística	
1.1. Historia de la estadística	Página 56
1.2. Conceptos básicos	Página 57
Población	Página 57
Muestra	Página 58
Caracteres cualitativos y cuantitativas	Página 58
Variables discretas y continuas	Página 59
1.3. Pongamos a prueba lo aprendido	Página 60
1.4. Solucionario	Página 62
2.- Recopilación de datos, variables, clases y frecuencias	
Ejemplo 1	Página 64
Frecuencia Absoluta	Página 65
Frecuencia Relativa	Página 67
Frecuencia Absoluta Acumulada	Página 67
2.1. Construcción de Tablas de frecuencias	Página 68
Ejemplo 2	Página 68
Ejemplo Excel 1	Página 70
Ejemplo Excel 2	Página 81
2.2. Propuesta de Trabajo Excel	Página 83
2.3. Pongamos a Prueba lo Aprendido	Página 84
2.4. Solucionario	Página 85
3.- Representaciones Gráficas	
Gráfico de barras	Página 86
Ejemplo 3	Página 86
Gráfico Circular	Página 86
Ejemplo 4	Página 87
Diagrama de Tallo y Hoja	Página 87
Ejemplo 5	Página 88
Histograma	Página 88
Ejemplo 6	Página 89
Pictograma	Página 90
Ejemplo 7	Página 90
Gráfico de Dispersión	Página 91
Ejemplo 8	Página 91
Ejemplo Excel 3	Página 92
Ejemplo Excel 4	Página 97
Ejemplo Excel 5	Página 102
3.1. Propuesta de Trabajo Excel	Página 109
3.2. Pongamos a Prueba lo Aprendido	Página 112
3.3. Solucionario	Página 114
4.- Medidas Estadísticas	
4.1. Medidas de Tendencia Central	Página 117
4.1.1. Media Aritmética	Página 117
Ejemplo 9 y 10	Página 119
4.1.1.1. Media Aritmética Ponderada	Página 121
Ejemplo 11	Página 122
4.1.2. Mediana	Página 122
Ejemplo 12	Página 124
Ejemplo 13	Página 124
4.1.3. Moda	Página 127
Ejemplo Excel 6	Página 129
Ejemplo Excel 7	Página 133

4.1.4. Propuesta de Trabajo Excel	Página 136
4.1.5. Pongamos a Prueba lo Aprendido	Página 138
4.1.6. Solucionario	Página 140
4.2. Medidas de dispersión	Página 144
Ejemplo 15	Página 144
4.2.1. Rango o Recorrido	Página 145
4.2.2. Desviación Media	Página 145
Ejemplo 16	Página 147
4.2.3. Varianza	Página 147
Ejemplo 17	Página 148
4.2.4. Desviación Estándar	Página 149
4.3. Medidas de localización: Cuarteles, Percentiles y Deciles	Página 150
4.3.1. Cuarteles	Página 150
4.3.2. Deciles	Página 151
4.3.3. Percentil	Página 151
Ejemplo 18	Página 152
4.4. Diagrama de cajas o cajón con bigotes	Página 154
Ejemplo 19	Página 154
Ejemplo Excel 8	Página 156
Ejemplo Excel 9	Página 169
4.5. Propuesta de Trabajo Excel	Página 174
4.6 Pongamos a Prueba lo Aprendido	Página 175
4.7. Solucionario	Página 177

1.-La estadística

1.1. Historia de la estadística

Como bien sabemos los orígenes de la estadística se encuentran en tiempos remotos. En el antiguo Egipto, existía una administración pública que recopilaba regularmente datos sobre tierras cultivadas y producción. De la época Babilónica se han encontrado testimonios donde además de datos sobre producción y comercio, se encuentran información sobre impuestos. Los romanos lo aplicaron con el fin de hacer sus censos y empadronamientos, en todas las provincias romanas. En la biblia encontramos citas de censos entre los que se cuentan uno llevado a cabo por Moisés (Números, 1,2-1.46) y un empadronamiento en la época del nacimiento de Cristo (Lucas, 2.1). También fue usada por los chinos unos 2000 años antes de Cristo, así como también por los persas, asirios y griegos.

Sin embargo, la estadística moderna, metódica y científica se atribuye al inglés John Graunt (1620-1670) quien publicó los primeros resultados de sus estudios que analizaban las causas de muerte en Londres en el año 1662. Se considera también uno de los iniciadores de la estadística como ciencia a Godofredo Ancheren (1749). De aquí en adelante, muchos matemáticos contribuyeron a darle un fundamento teórico, transformándola finalmente en una rama de la matemática.

El desarrollo de la estadística no ha concluido; todos los días se crean diferentes métodos para el manejo de datos, su interpretación y la relación entre ellos. La computación, actualmente en desarrollo, ha hecho un gran aporte al procesamiento de la información.

No es fácil definir lo que se entiende por estadística. Podemos decir que es la ciencia que se ocupa de la recolección, clasificación y análisis de conjunto de datos procedentes de la observación o la experimentación.

Tanto por el método empleado como por el objetivo planteado, la estadística podemos clasificarla en descriptiva e inductiva.

La estadística inductiva se ocupa básicamente de estudiar y analizar, con espíritu científico todos los datos, tablas y gráficos obtenidos mediante la estadística descriptiva para llegar finalmente a la descripción de una población, a determinar relaciones entre una población y otra ideal.

La que nos preocupara por ahora será la estadística descriptiva, el objetivo principal de esta es la recopilación, clasificación y tabulación de datos. Si se

conocen algunos elementos (media, cuartel, varianza, etc.) que nos permiten completar la tabulación y utilizando un grafico adecuado podemos calcular mejor los resultados obtenidos de este trabajo cuantitativo.



John Graunt (1620-1674), un inglés que estudiaba los expedientes de los nacimientos y muertes descubrió que nacían más niños que niñas, pero también encontró que por estar los hombres más expuestos a accidentes ocupacionales, a enfermedades y la guerra, el número de hombres y mujeres en la edad de casarse era más o menos la misma.

Entonces... ¿Para que nos sirve la estadística?

La estadística nos facilita la interpretación de los resultados, hace un uso eficiente de la información, además proporciona un lenguaje común para comunicar los hallazgos científicos de diversas disciplinas.

1.2. Conceptos básicos

Población

Un conjunto de individuos o elementos que tienen ciertas propiedades comunes, por ejemplo:

- el personal de una empresa,
- la producción de una fábrica,
- el parque automotriz de la ciudad de Santiago en febrero de 2009,
- los accidentes ocurridos en Chile durante 2008.

Muestra

Cuando la población es muy numerosa, la observación de todos sus componentes se ve imposibilitada por el enorme costo, la gran cantidad de trabajo y el tiempo necesario para llevarla a cabo. Estos inconvenientes pueden ser superados mediante la selección de un conjunto representativo de la población, llamado **muestra**.

Observación

Es importante que la muestra sea representativa, por ejemplo cuando vas a la feria y compras papas ¿Crees que las que ubican adelante es un conjunto representativo de la calidad de todas las papas?

Debemos tener claro que los elementos de una población usualmente se llaman unidades estadística o individuos. Cada individuo de una población puede describirse según uno o varios **caracteres**.

Cada uno de los caracteres observados puede presentar dos o más modalidades, que son las diferentes situaciones posibles del carácter

Un carácter se dice **cualitativo** si sus diversas modalidades no son medibles, su valor será una característica del individuo. Por ejemplo:

1) Color de ojos: —————> CARACTER

- verdes
 - cafés
 - negros
 - pardos
 - azules
- } MODALIDADES

2) Profesión —————> CARACTER

- Profesor
 - Abogado
 - Medico
 - Ingeniero
 - Periodista
- } MODALIDADES

Un carácter se dice **cuantitativo** si sus diversas modalidades son medibles o numerables, es decir, si a cada una de sus modalidades se le asigna un número. Por ejemplo:

- 1) estatura
- 2) edad
- 3) peso
- 4) número de hijos
- 5) promedio de notas
- 6) numero de integrantes de una familia
- 7) el numero de piezas defectuosas en un lote de 100 piezas

Este número variable con la modalidad pero específico para cada una de ellas, se llama **variable estadística**. Las diferentes modalidades de un carácter cuantitativo son los diferentes valores posibles de la variable estadística.

Un dato cuantitativo se denomina **discreto** cuando solo puede tomar determinados valores, que son enteros positivos. Como vimos en los ejemplos anteriores estos serian

- 1) numero de hijos
- 2) numero de integrantes de una familia
- 3) el numero de piezas defectuosas en un lote de 100 piezas

Un dato cuantitativo se denomina **continuo** cuando puede tomar cualquier valor en un intervalo, ósea puede tomar por valor cualquier numero real. Siguiendo el ejemplo, estos serian

- 1) estatura
- 2) peso
- 3) promedio de notas
- 4) edad

1.3. PONGAMOS A PRUEBA LO APRENDIDO

Actividad 1

a.- ¿Para que tipo de estudios se utilizaba la estadística antiguamente?. Nombre 3 ejemplos

b.- Actualmente ¿en que has visto que se utilice la estadística?

c.- ¿Para que crees que sirve la estadística?

Actividad 2

a.- Para mejorar la gestión de los consultorios de salud en Santiago, interesa conocer la opinión de las enfermeras sobre diversos aspectos del funcionamiento de los hospitales. Se necesita como mínimo la opinión de 100 enfermeras. ¿Cual es la población y la muestra?

b.- Interesa conocer la aceptación de un nuevo juguete dirigido a niños entre 6 y 10 años. ¿Cual crees tú que es la muestra más representativa?

- i) Ir a un parque de diversiones y realizar una pequeña encuesta a los niños presentes
- ii) Conseguir el listado de 500 niños entre esas edades, que asisten a un colegio y mandar 200 encuestas.
- iii) En una esquina del centro de la ciudad realizar una encuesta a niños que pasen por ahí.

Actividad 3

a.- De los siguientes enunciados identificar el carácter y las modalidades posibles en cada uno.

- i) Se clasifica a los empleados de una empresa según el nivel de escolaridad
- ii) En el Colegio María Auxiliadora se clasifica a los estudiantes según: edad, sexo y promedio de notas.

b.- Determinar en cada caso, el tipo de carácter y si es discreto o continuo cuando corresponda.

- i) Las nacionalidades de los pasajeros en un vuelo determinado
- ii) La estatura de los jugadores de la selección chilena
- iii) Numero de hermanos por cada estudiante
- iv) La preferencia según equipo deportivo favorito de 50 personas
- v) Numero de empleados de una tienda comercial

1.4. SOLUCIONARIO

Actividad 1:

a)

- En el antiguo Egipto, existía una administración pública que recopilaba regularmente datos sobre tierras cultivadas y producción.
- En la época Babilónica se han encontrado testimonios donde se encuentra información sobre impuestos.
- Los romanos lo aplicaron con el fin de hacer sus censos y empadronamientos, en todas las provincias romanas.

b)

- Actualmente la estadística se utiliza por ejemplo para informar a los habitantes de un país sus preferencias por algún presidente en especial, o la tendencia de algún presidente.

c)

- Actualmente la estadística nos facilita la interpretación de los resultados, hace un uso eficiente de la información, además proporciona un lenguaje común para comunicar los hallazgos científicos

Actividad 2:

a)

- La población → Enfermeras.
- La muestra → 100 Enfermeras.

b)

- La muestra más representativa es conseguir el listado de 500 niños entre esas edades, que asisten a un colegio y mandar 200 encuestas.

Actividad 3:

a)

- Carácter → Empleados de una empresa - Modalidad → Nivel de escolaridad.
- Carácter → Estudiantes del colegio Maria Auxiliadora
- Modalidad → Edad, Sexo y Promedio de notas.

b)

- Carácter cualitativo discreto.
- Carácter cuantitativo continuo.
- Carácter cuantitativo discreto.
- Carácter cualitativo discreto.
- Carácter cuantitativo discreto.

2.- Recopilación de datos, variables, clases y frecuencias

Una vez que los datos han sido registrados, corresponde organizarlos y explorarlos para detectar posibles patrones y tendencias.

Veamos el siguiente ejemplo

Ejemplo 1

Se registraron las edades de 50 personas cesantes en la comuna de La Florida

45	33	31	33	41	41	40	45	31	45
41	45	40	31	31	40	40	28	45	41
40	31	31	33	41	41	33	45	31	41
41	45	40	31	31	40	40	28	41	31
41	41	40	31	31	40	33	41	33	31

Para poder observar detalladamente la cantidad de personas que corresponden a cada edad, se ordenan los datos de la siguiente manera:

EDAD	Numero de personas
28	2
31	13
33	6
40	10
41	12
45	7
	TOTAL 50

Se puede ver que hay 13 personas cesantes de 31 años, y solo 2 persona cesante de 28 años.

Piensa tú... ¿Que otras conclusiones podemos obtener de la tabla?

Entonces:

Frecuencia absoluta: Es el número de veces que se repite la modalidad (25, 26...45), o el número de individuos que pertenecen a la misma modalidad, la frecuencia absoluta se denota n_i

Recordar:

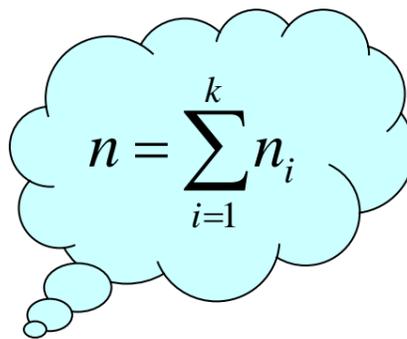
Para ahondar en la unidad de estadística descriptiva, es necesario conocer la notación de sumatoria ya que en lo que nos resta de materia es fundamental su uso.

Este símbolo \sum es la letra griega sigma mayúscula, que corresponde a la decimoctava letra del alfabeto griego, y a la letra **S** de nuestro alfabeto.

El uso de sumatoria tiene evidentes ventajas, puesto que permite escribir formulas de manera mas reducidas.

Si queremos la suma de $1 + 2 + 3 + 4 + 5 = \sum_{k=1}^{k=5} k$

Designaremos por n al número de individuos de la población (o muestra) y por n_i a la frecuencia absoluta, por lo tanto n será la sumatoria de todas las frecuencias absolutas.


$$n = \sum_{i=1}^k n_i$$

También con la frecuencia absoluta, podemos calcular otro tipo de frecuencia, llamada **frecuencia relativa**, que es el cociente entre la frecuencia absoluta y el número total de datos, esta puede ser interpretada mediante porcentaje multiplicando el resultado por 100, como se verá a continuación.

EDAD	n_i	FRECUENCIA RELATIVA (f_i)
28	2	$\frac{2}{50} = 0,04 \rightarrow 4 \%$
31	13	$\frac{13}{50} = 0,26 \rightarrow 26 \%$
33	10	$\frac{10}{50} = 0,2 \rightarrow 20 \%$
40	6	$\frac{6}{50} = 0,12 \rightarrow 12 \%$
41	12	$\frac{12}{50} = 0,24 \rightarrow 24 \%$
45	7	$\frac{7}{50} = 0,14 \rightarrow 14 \%$
	$n = 50$	$\frac{50}{50} = 1 \rightarrow 100 \%$

Se puede concluir que hay mas personas cesantes de 31 años, ya que el porcentaje mas alto, es de un 26%.

Piensa tú... ¿Que conclusiones podemos obtener de la tabla?

Observación:

Bueno nos podemos dar cuenta en la tabla anterior, que si sumamos los porcentajes nos dará el 100 % de la muestra, pero es importante tener en cuenta que existen casos especiales en donde por causa de las aproximaciones la suma de los porcentajes no será exactamente 100 %.

Entonces:

Frecuencia relativa: Es el número de veces que se repite la modalidad, referidos al total de individuos, en otras palabras, es la razón entre la frecuencia absoluta y el total de datos, que se puede expresar mediante el uso de porcentajes.

La frecuencia relativa se denota por f_i y la sumatoria de los f_i será igual a 1.

$$f_i = \frac{n_i}{n}$$
$$\sum_{i=1}^k f_i = 1$$

Siguiendo con el ejemplo anterior nos podemos preguntar, ¿cuántas personas cesantes tienen menos de 40 años?

Frecuencia Absoluta Acumulada, denotada por N_i , es la suma de los distintos valores de la frecuencia absoluta tomando como referencia un individuo dado. La última frecuencia absoluta acumulada es igual al n° de casos.

De acuerdo a la definición las frecuencias absolutas acumuladas suman frecuencias absolutas, por lo tanto:

$$\begin{aligned} N_1 &= n_1 \\ N_2 &= n_1 + n_2 = N_1 + n_2 \\ N_3 &= n_1 + n_2 + n_3 = N_2 + n_3 \\ N_4 &= n_1 + n_2 + n_3 + n_4 = N_3 + n_4 \\ &\vdots \\ &\vdots \\ N_k &= n_1 + n_2 + n_3 + \dots + n_k = n = N_{k-1} + n_k \end{aligned}$$

Así la frecuencia absoluta acumulada en la tabla de frecuencias es:

EDAD	n_i	N_i
28	2	2
31	13	15
33	10	25
40	6	31
41	12	43
45	7	50
	n = 50	

2.1. Construcción de tablas de frecuencias

Ejemplo 2:

En un estudio estadístico realizado por la revista medica de chile, en una escuela de natación, se obtuvieron las edades de 80 personas con obesidad:

14	27	14	9	18	14	10	15	21	23
5	10	11	10	24	12	11	16	17	14
10	12	13	16	15	26	14	12	25	12
17	9	26	16	8	14	12	14	22	16
6	19	12	22	6	10	16	26	18	14
8	15	14	11	15	6	16	13	16	23
6	26	27	26	6	10	16	14	28	14
25	28	14	28	5	5	28	13	16	13

Para formar nuestra tabla, primero se tabulara en **intervalos de clase**, de la forma $[x'_{j-1}, x'_j[$ donde x'_{j-1} es el límite inferior del intervalo y x'_j es el límite superior del intervalo, el número de intervalos (k) debe ser entre $5 \leq k \leq 20$, en este caso lo haremos de 6 intervalos, y el **rango** será la diferencia entre el menor y el mayor, $28 - 5 = 23$, luego para la **amplitud** de cada intervalo se dividirá el rango por el número de intervalos que se desea $23: 6 = 3,8 \approx 4$

El centro de cada uno de los intervalos se le denomina **marca de clase**.

$$x_i = \frac{x'_{j-1} + x'_j}{2} \quad i=1,2,3,\dots,k$$

Observación:

Esta nos servirá mas adelante para el cálculo de parámetros estadísticos.

EDADES [x'_{j-1} , x'_j [Marca de clase (x_i)	Frecuencia absoluta (n_i)	Frecuencia relativa (f_i)
5 – 9	7	10	10/80 = 0.125
9 – 13	11	16	16/80 = 0.2
13 – 17	15	30	30/80 = 0.375
17 – 21	19	5	5/80 = 0.0625
21 – 25	23	6	6/80 = 0.075
25 – 29	27	13	13/80 = 0.1625

Las personas con mayor obesidad fluctúan entre los 13 y 16 años.

Piensa tú... ¿Que conclusiones podemos obtener de la tabla?

Entonces:

Para construir una distribución de frecuencias basada en intervalos de clase se sigue el siguiente procedimiento:

- Fijar el número de intervalos “k” donde $5 \leq k \leq 20$

- Calcular la amplitud de los intervalos "a", donde $a = \frac{x_{max} - x_{min}}{k}$, donde x_{max} es la mayor de las observaciones y x_{min} es la menor de las observaciones. La amplitud debe ser un número sencillo.

- Formar los intervalos de clase, que en general son de la siguiente forma:

$$[x'_0, x'_1[, [x'_1, x'_2[, [x'_2, x'_3[, \dots, [x'_{k-1}, x'_k[$$

x'_0 Puede ser la menor de las observaciones. Los intervalos se eligen de manera que los puntos medios coincidan con los datos realmente observados.

- Calcular **las marcas de clase** x_i que son los puntos medios de los intervalos

Ejemplo Excel 1:

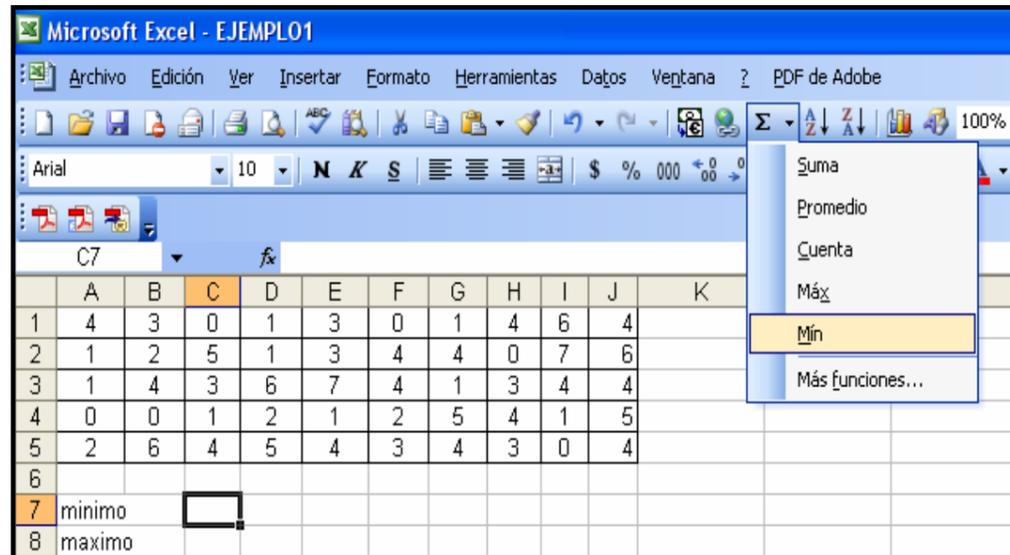
Una industria desea conocer el consumo familiar mensual de cierto producto, envasado en latas. Para tal objetivo, se ha encuestado a 50 familias, obteniéndose la siguiente información:

4	-	3	-	0	-	1	-	3	-	0	-	1	-	4	-	6	-	4
1	-	2	-	5	-	1	-	3	-	4	-	4	-	0	-	7	-	6
1	-	4	-	3	-	6	-	7	-	4	-	1	-	3	-	4	-	4
0	-	0	-	1	-	2	-	1	-	2	-	5	-	4	-	1	-	5
2	-	6	-	4	-	5	-	4	-	3	-	4	-	3	-	0	-	4

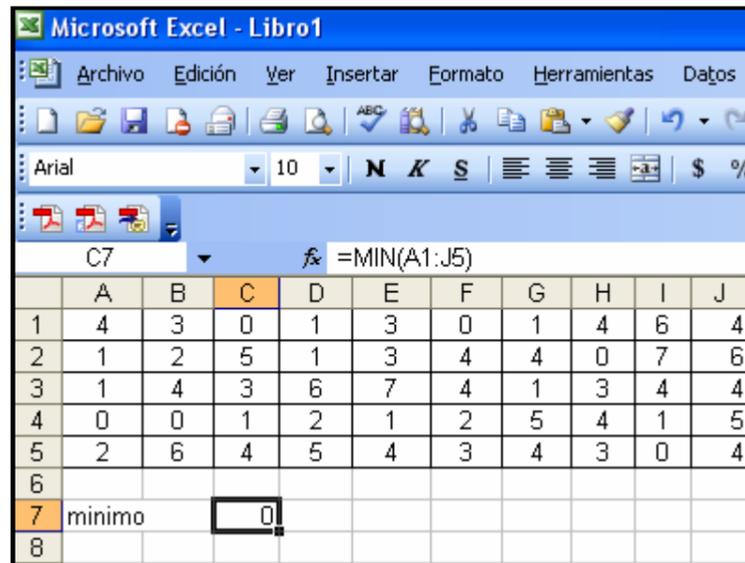
Solución: Supongamos que se desea construir una tabla de distribución de frecuencias que muestre de manera óptima el comportamiento de los datos. Para ello, es importante saber qué representa la información cuantitativa contenida en la tabla anterior. No representa el número de familias que consumen el producto en latas, sino que cada número representa la cantidad de latas consumida por una de las 50 familias.

Esta tabla de distribución de frecuencias podemos crearla con ayuda del programa Microsoft Excel. Veamos cómo hacerlo:

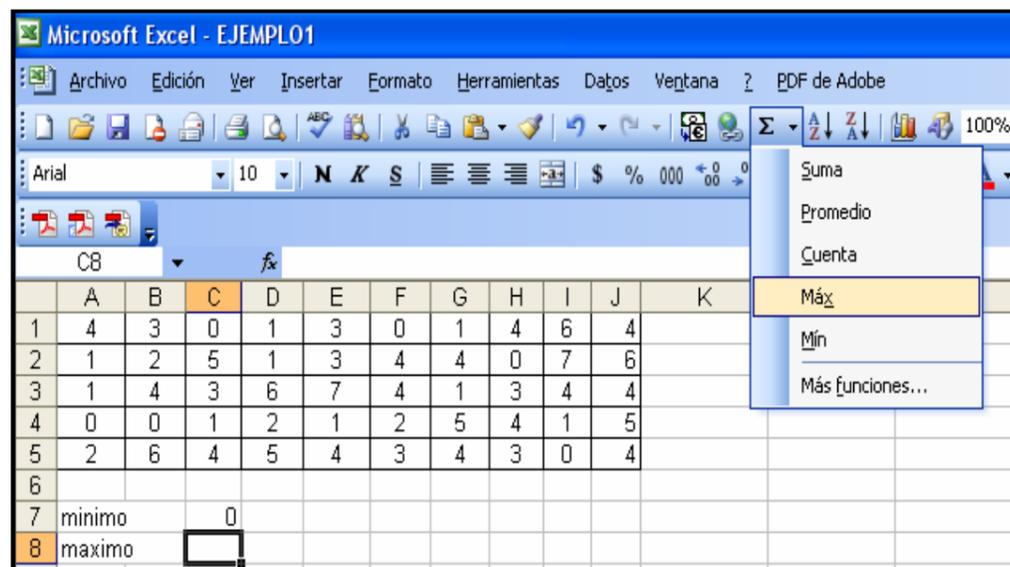
Notemos que la variable de interés “Número de latas consumidas por 50 familias” es de tipo discreta. Luego, para construir una tabla de distribución de frecuencias, podemos determinar, en primer lugar, los valores mínimo y máximo observados de la variable. En Excel, esto se puede hacer gracias a la función **MIN** y **MAX**, del siguiente modo: En la celda **A7** podemos escribir el texto “Mínimo” y nos situamos en la celda **C7**, hacemos clic en autosuma seleccionando **Mín**.



Entonces seleccionamos las celdas de **A1:J5**, seguido de la tecla **ENTER**. Aparecerá entonces en la celda **C7** el valor 0, que es el mínimo del conjunto de datos **A1:J5** (Ver siguiente figura).



Ahora, en la celda **A8** podemos escribir “Máximo”, y nos situamos en la celda **C7**, hacemos clic en autosuma seleccionando **Máx**.

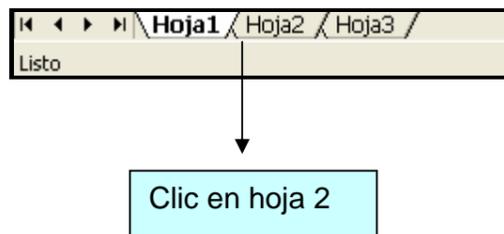


Entonces seleccionamos las celdas de **A1:J5**, seguido de la tecla **ENTER**. El resultado, 7, se ve en la siguiente pantalla:

The screenshot shows the Microsoft Excel interface with a data table and summary statistics. The formula bar shows $=MAX(A1:J5)$ in cell C8.

	A	B	C	D	E	F	G	H	I	J
1	4	3	0	1	3	0	1	4	6	4
2	1	2	5	1	3	4	4	0	7	6
3	1	4	3	6	7	4	1	3	4	4
4	0	0	1	2	1	2	5	4	1	5
5	2	6	4	5	4	3	4	3	0	4
6										
7	minimo		0							
8	maximo		7							
9										

Al conocer los valores mínimo y máximo, podemos construir una tabla de distribución de frecuencias en Excel. Podemos construir esta tabla en la misma hoja, pero para efectos de claridad, se hará en la hoja consecutiva (esto es, en la hoja 2. Ver figura). Hacemos clic entonces en **Hoja2**.



En la **Hoja2**, celda **A1**, podemos escribir "Número de latas", y en la casilla siguiente (**A2**) escribimos el menor valor observado de la variable, esto es, el valor 0. Luego, con el cursor situado sobre el valor 0, presionamos el botón principal del Mouse más la tecla Control.

La casilla **A2** saldrá con un signo +. Observar la figura

	A
1	Numero de latas
2	0 +
3	

Manteniendo presionado la tecla **Control**, arrastramos el Mouse hacia abajo. Notaremos que aparecen a la derecha del puntero los números 1,2,3,... a medida que descendemos con el Mouse. Si soltamos el botón principal del Mouse en la celda **A9**, aparecerán los ocho valores asumidos por la variable “número de latas” (de 0 a 7).

	A
1	Numero de latas
2	0
3	1
4	2
5	3
6	4
7	5
8	6
9	7
10	

En la celda **B1** escribimos “Frecuencia absoluta”. Ahora, seleccionamos las celdas comprendidas entre **B2** y **B9**. Hacemos clic en el botón **Insertar Función**, **Seleccionar una categoría: Estadísticas**, **Seleccionar una función: FRECUENCIA**. (Ver figura siguiente).

	A	B	C	D	E	F	G
1	Numero de latas	Frecuencia absoluta					
2		0	=				
3		1					
4		2					
5		3					
6		4					
7		5					
8		6					
9		7					
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							

Insertar función

Buscar una función:

Escriba una breve descripción de lo que desea hacer y, a continuación, haga clic en Ir

O seleccionar una categoría: Estadísticas

Seleccionar una función:

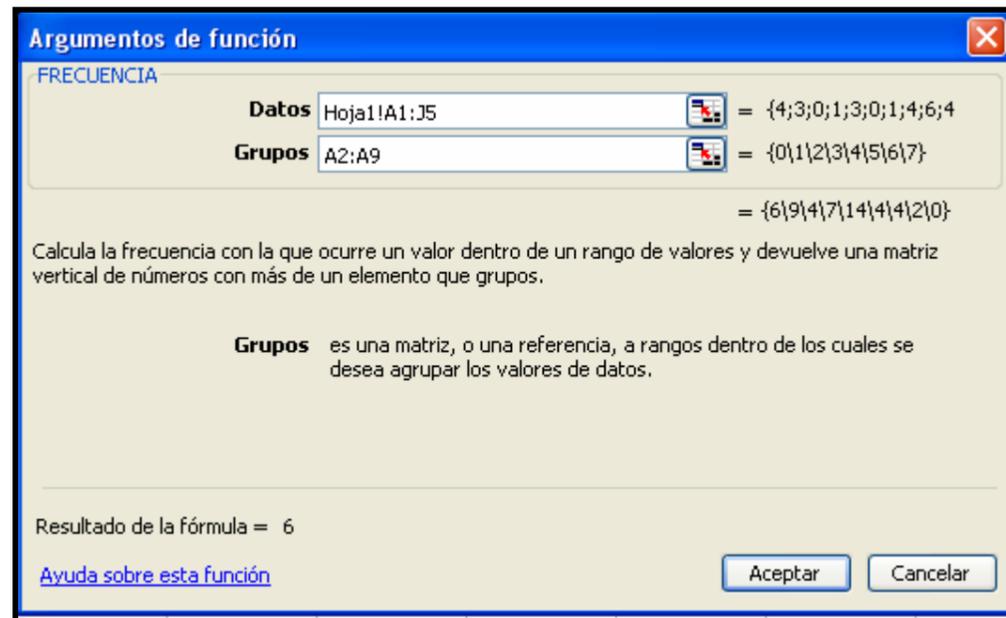
ERROR.TIPICO.XY
 ESTIMACION.LINEAL
 ESTIMACION.LOGARITMICA
 FISHER
FRECUENCIA
 GAMMA.LN
 INTERSECCION.EJE

FRECUENCIA(datos,grupos)
 Calcula la frecuencia con la que ocurre un valor dentro de un rango de valores y devuelve una matriz vertical de números con más de un elemento que grupos.

[Ayuda sobre esta función](#) Aceptar Cancelar

Al hacer clic en **ACEPTAR**, aparecerá el cuadro de diálogo **FRECUENCIA**. En la opción **Datos**, hacemos clic en **Hoja1** y seleccionamos el rango de datos que contiene a la muestra de 50 familias (en este caso, desde **A1** a **J5**).

Para la opción **Grupos**, seleccionamos el rango que contiene a las distintas categorías de la variable “Número de latas” (en este caso, dentro de **Hoja2**, seleccionamos desde **A2** hasta **A9**). Observar la siguiente figura:

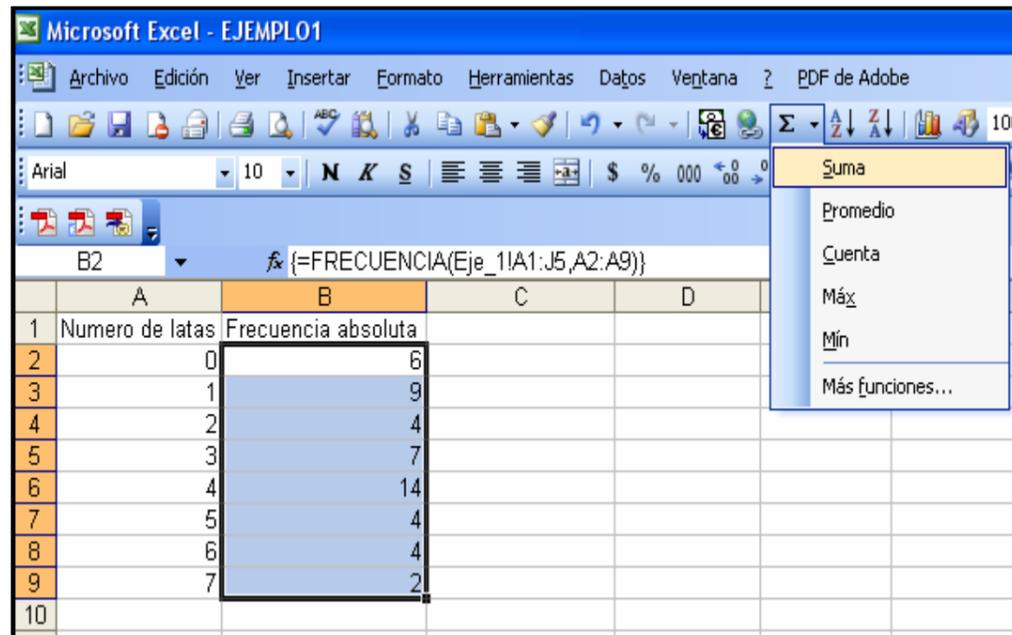


Ahora hemos llegado a un punto importante: **NO HACER CLIC EN ACEPTAR**. En vez de ello, presionamos al mismo tiempo las teclas **Control + Shift + Enter**. Esto producirá la siguiente salida:

	A	B
1	Numero de latas	Frecuencia absoluta
2	0	6
3	1	9
4	2	4
5	3	7
6	4	14
7	5	4
8	6	4
9	7	2

Resultado de presionar conjuntamente las teclas **CONTROL+SHIFT+ENTER**

En la celda **B10** podemos calcular la suma de las frecuencias absolutas calculadas anteriormente, seleccionando las celdas **B2:B9**, hacemos clic en **AUTOSUMA**, seleccionando **Suma**.



Así en la celda **B10** aparecerá el valor 50, que es el tamaño de la muestra.

	A	B
1	Numero de latas	Frecuencia absoluta
2	0	6
3	1	9
4	2	4
5	3	7
6	4	14
7	5	4
8	6	4
9	7	2
10		50

Valor de la suma total de las frecuencias absolutas

De la tabla anterior, ya es posible desprender que el valor 7, ubicado en la celda **B5**, indica que 7 familias de la muestra consumieron 3 latas al mes. Puede surgir en este momento la siguiente pregunta: ¿qué porcentaje de las familias encuestadas consumen cuatro latas? Podemos contestar esa pregunta, agregando una columna de frecuencias relativas a nuestra tabla anterior. En la celda **C1**, escribimos "Frecuencia relativa". Luego, en la celda **C2** escribimos la expresión $= B2/ \$B\10 (con signo peso, puesto que **B10** será una celda fija), que nos entrega la proporción de familias que no consumen latas (en este caso, 0.12).

	A	B	C
1	Numero de latas	Frecuencia absoluta	Frecuencia relativa
2	0	6	0.12
3	1	9	
4	2	4	
5	3	7	
6	4	14	
7	5	4	
8	6	4	
9	7	2	
10		50	

Para completar la columna de frecuencias relativas, arrastramos el Mouse desde la celda **C2** hasta la celda **C9** (sin soltar el botón), obteniendo las frecuencias relativas restantes:

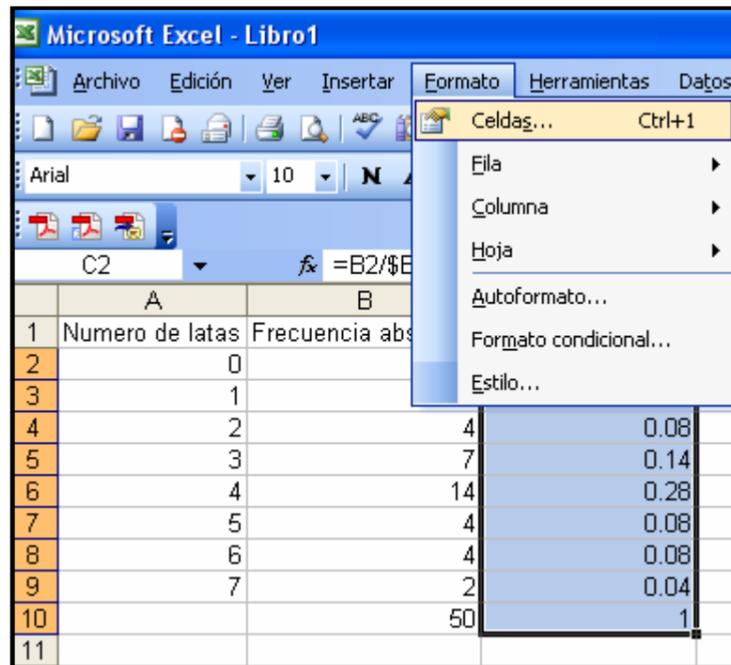
C2		fx =B2/\$B\$10		
	A	B	C	
1	Numero de latas	Frecuencia absoluta	Frecuencia relativa	
2	0	6	0.12	
3	1	9	0.18	
4	2	4	0.08	
5	3	7	0.14	
6	4	14	0.28	
7	5	4	0.08	
8	6	4	0.08	
9	7	2	0.04	
10		50		

Con la opción **AUTOSUMA** podemos calcular la suma de las frecuencias relativas, contenidas en la columna **C**, obteniendo lo siguiente:

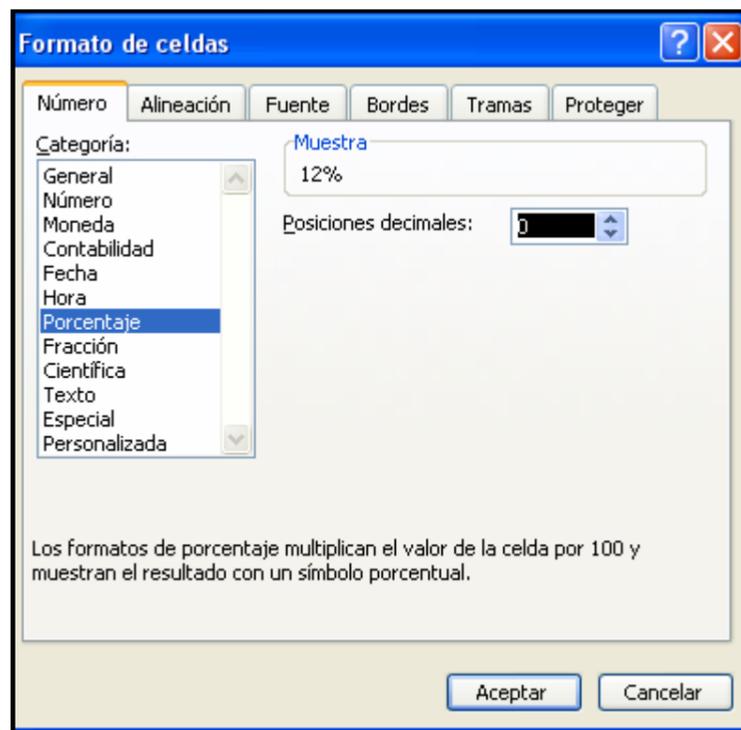
	A	B	C
1	Numero de latas	Frecuencia absoluta	Frecuencia relativa
2	0	6	0.12
3	1	9	0.18
4	2	4	0.08
5	3	7	0.14
6	4	14	0.28
7	5	4	0.08
8	6	4	0.08
9	7	2	0.04
10		50	1
11			

Como las frecuencias relativas representan porcentajes, podemos presentar en formato porcentual dichas frecuencias. Para ello, seleccionamos las celdas que

contienen a las frecuencias relativas (en este caso, desde **C2** hasta **C10**) y hacemos clic en **Formato – Celdas**.



Aparecerá entonces el cuadro de diálogo **Formato de Celdas**. En la opción **Número**, seleccionamos **Porcentaje**, y en **Posiciones Decimales**, elegimos el valor 0 (para mayor simplicidad).



A continuación se muestra el resultado de las instrucciones anteriores

	A	B	C
1	Numero de latas	Frecuencia absoluta	Frecuencia relativa
2	0	6	12%
3	1	9	18%
4	2	4	8%
5	3	7	14%
6	4	14	28%
7	5	4	8%
8	6	4	8%
9	7	2	4%
10		50	100%

Es posible desprender de la columna de frecuencias relativas, por ejemplo, que el 18% de las familias encuestadas consume una lata al mes. Otra pregunta de interés puede resultar: ¿Cuántas familias consumen menos de 5 latas al mes?. Para contestar a esta pregunta, podemos construir otra columna en nuestra tabla.

Esta columna es la **columna de frecuencias absolutas acumuladas**. En la celda **D1**, podemos escribir "frec. abs. acum." (abreviadamente, aunque no es estrictamente necesario abreviar). En **D2** (la primera frecuencia absoluta acumulada) escribimos la fórmula "**= B2**" (esto es, la referencia a la celda que contiene el primer valor de la frecuencia absoluta).

D2		fx =B2			
	A	B	C	D	
1	Numero de latas	Frecuencia absoluta	Frecuencia relativa	frec.abs. acu.	
2	0	6	12%	6	
3	1	9	18%		
4	2	4	8%		
5	3	7	14%		
6	4	14	28%		
7	5	4	8%		
8	6	4	8%		
9	7	2	4%		
10		50	100%		

Para determinar el valor de la segunda frecuencia acumulativa, escribimos en la celda **D3** la fórmula

$$= D2 + B3$$

El resultado se muestra a continuación:

	A	B	C	D
1	Numero de latas	Frecuencia absoluta	Frecuencia relativa	frec.abs. acu.
2	0	6	12%	6
3	1	9	18%	15
4	2	4	8%	
5	3	7	14%	
6	4	14	28%	
7	5	4	8%	
8	6	4	8%	
9	7	2	4%	
10		50	100%	

Ahora, sólo basta con mantener presionado el botón principal del Mouse y arrastrarlo hasta la celda **D9** (cuando aparezca el signo + sobre el cursor). De este modo tendremos completa la columna de frecuencias absolutas acumuladas.

	A	B	C	D
1	Numero de latas	Frecuencia absoluta	Frecuencia relativa	frec.abs. acu.
2	0	6	12%	6
3	1	9	18%	15
4	2	4	8%	19
5	3	7	14%	26
6	4	14	28%	40
7	5	4	8%	44
8	6	4	8%	48
9	7	2	4%	50
10		50	100%	

De la columna de frecuencias absolutas acumulativas podemos ver que 40 familias consumen menos de cinco latas al mes.

Ejemplo Excel 2:

La siguiente tabla corresponde a los sueldos que paga en la actualidad una empresa comercial (en miles de \$):

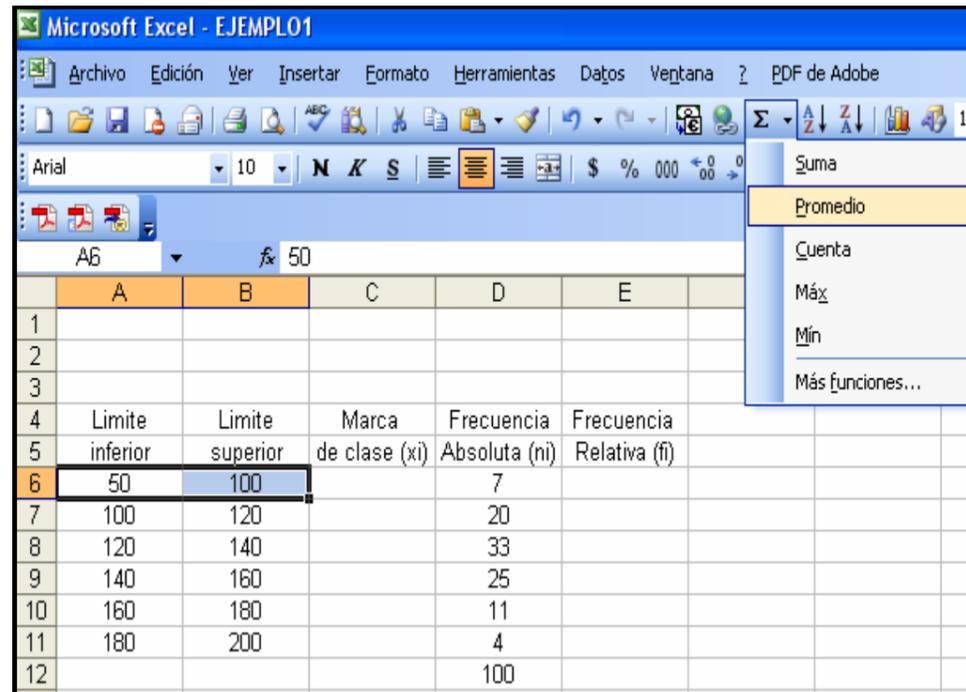
Sueldo	Número de empleados
50 – 100	7
100 – 120	20
120 – 140	33
140 – 160	25
160 – 180	11
180 – 200	4

Una vez cargado el programa Microsoft Excel, se procede a ingresar los datos a la planilla. En una hoja activa, ingresamos los valores correspondientes a los límites inferior y superior de los intervalos (idealmente, en columnas distintas. En este ejemplo, en las columnas **A** y **B**), y sus correspondientes frecuencias absolutas, en otra columna (en este caso columna **D**) y frecuencia relativa (Columna **E**):

	A	B	C	D	E
1					
2					
3					
4	Limite inferior	Limite superior	Marca de clase (xi)	Frecuencia Absoluta (ni)	Frecuencia Relativa (fi)
6	50	100		7	
7	100	120		20	
8	120	140		33	
9	140	160		25	
10	160	180		11	
11	180	200		4	
12				100	

Para determinar los valores de la columna “marca de clase” (que es el punto medio de cada intervalo) escribimos en la celda **C6** la expresión $= (A6 + B6) / 2$, seguido de la tecla **Enter**. Aparecerá en **C6** la primera marca de clase. Al arrastrar el contenido de la celda **C6** con el Mouse hasta la celda **C18**, completamos la columna de marcas de clase.

También podemos usar la función **PROMEDIO** para determinar la marca de clase, para cada intervalo, seleccionando la celda **A6** y **B6** simultáneamente, luego en la función autosuma hacemos clic en promedio y aparecerá la marca de clase, luego arrastramos hasta **C11**.



El resultado es el siguiente:

	A	B	C	D	E
1					
2					
3					
4	Limite inferior	Limite superior	Marca de clase (xi)	Frecuencia Absoluta (ni)	Frecuencia Relativa (fi)
5					
6	50	100	75	7	
7	100	120	110	20	
8	120	140	130	33	
9	140	160	150	25	
10	160	180	170	11	
11	180	200	190	4	
12				100	

Ahora completamos la tabla con las frecuencias relativas, recordando que se ingresa de la siguiente manera, en la celda **E6** escribimos la expresión **= D6/ \$D\$11**, luego arrastramos hasta la celda E11, obteniendo así lo siguiente

	A	B	C	D	E
1					
2					
3					
4	Limite inferior	Limite superior	Marca de clase (xi)	Frecuencia Absoluta (ni)	Frecuencia Relativa (fi)
6	50	100	75	7	7%
7	100	120	110	20	20%
8	120	140	130	33	33%
9	140	160	150	25	25%
10	160	180	170	11	11%
11	180	200	190	4	4%
12				100	

2.2. Propuesta de trabajo Excel

Los siguientes datos corresponden a la variable X ="número de hijos", medida en los cincuenta trabajadores de cierta fábrica:

1	2	3	2	4	4	2	1	0	2
3	2	4	5	2	1	0	1	2	4
2	4	4	2	0	2	3	5	3	2
0	2	3	3	2	2	3	2	4	5
0	2	2	1	4	1	1	3	3	1

Con apoyo de la planilla Excel, construir una tabla de distribución de frecuencias que resuma la información anterior. El resultado esperado de la actividad anterior se muestra en la siguiente imagen:

<i>número de hijos</i>	<i>frecuencia absoluta</i>	<i>frecuencia relativa</i>	<i>frecuencia absoluta acum.</i>
0	5	10,0%	5
1	8	16,0%	13
2	17	34,0%	30
3	9	18,0%	39
4	8	16,0%	47
5	3	6,0%	50
total	50		

2.3. PONGAMOS A PRUEBA LO APRENDIDO

Actividad 1

Se realizó una encuesta a 30 personas de acuerdo a la ciudad en que nacieron. Confeccionar la tabla de frecuencia, según los siguientes datos.

S: Santiago I: Iquique
C: Concepción T: Temuco

S	I	I	C	C	I
C	S	C	S	S	S
I	C	I	T	I	C
T	S	T	S	T	S
S	C	S	T	T	C

- 1.- ¿En qué ciudad nacieron más personas?
- 2.- ¿En qué ciudad nació la menor cantidad de personas, expresada en porcentajes?

Actividad 2

Confeccionar la tabla de distribución de frecuencias con 5 intervalos que contenga, intervalos, marca de clase, frecuencia absoluta y frecuencia relativa, según lo siguiente:

Se obtuvo el valor de la hora en miles de pesos en 50 trabajadores

13	9	22	37	21	7	20	18	35	21
16	33	10	31	29	11	10	26	24	2
24	20	0	22	14	22	5	19	17	12
12	28	21	8	3	27	23	11	3	16
34	12	30	5	19	39	25	4	9	24

- 1.- ¿Qué valor de la hora es más frecuente de los 50 trabajadores?
- 2.- ¿Qué valor de la hora es el menos frecuente?

2.4. SOLUCIONARIO

Actividad 1:

a)

CIUDAD DONDE NACIERON	Frecuencia Absoluta (n_i)	Frecuencia Relativa (f_i)
Santiago	10	0.33
Concepción	8	0.27
Iquique	6	0.2
Temuco	6	0.2
	30	

b) La ciudad de Santiago fue la ciudad que nacieron mas personas.

c) En la ciudad de Iquique y Temuco nacieron menos personas, un total de 20% en cada una.

Actividad 2:

a)

Miles de pesos	Marca de clase (x_i)	Frecuencia absoluta (n_i)	Frecuencia relativa (f_i)
0 – 9	4.5	9	0.18
9 – 17	13	13	0.26
17 – 25	21	16	0.32
25 – 33	29	7	0.14
33 – 41	53.5	5	0.1
		50	

b) El valor de la hora más frecuente de los 50 trabajadores esta entre los 17 mil y 25 mil pesos.

c) El valor de la hora menos frecuente esta entre los 33 mil y 41 mil pesos.

3.- Representaciones graficas

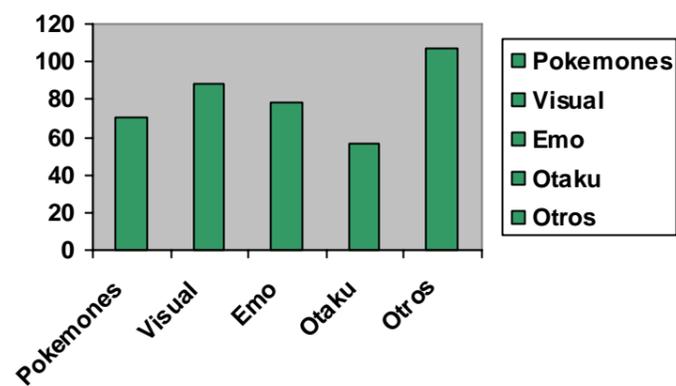
a) Grafico de barras

El grafico de barras es utilizado para comprender las frecuencias de las variables, por lo tanto la columna con mayor altura representa la mayor frecuencia.

Ejemplo 3:

En un colegio de Puente Alto se investigaron las diferentes tribus urbanas en los estudiantes de enseñanza media, los resultados arrojaron lo siguiente

Tendencias urbanas	Frecuencia absoluta (n_i)
Pokemones	70
Visual	88
Emo	78
Otaku	57
Otros	107
	n = 400



b) Grafico circular

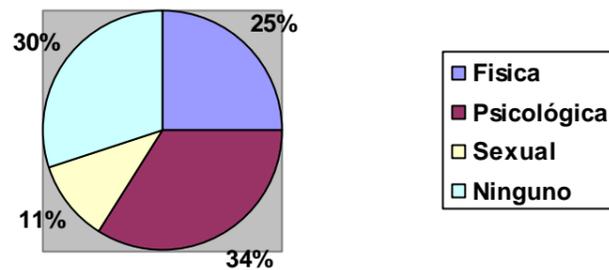
Si queremos poner de manifiesto la comparación entre las frecuencias con que aparece cada valor de la variable, es útil hacerlo mediante un grafico circular, este consiste en repartir los 360° del circulo proporcionalmente a las frecuencias de cada una de las clases en el que se ha dividido.

Ejemplo 4:

Un estudio realizado por el Sernam en marzo del 2007 dirigido a los tipos de maltratos realizados a las mujeres de la región metropolitana independientemente de su situación conyugal, arrojó los siguientes resultados:

Mujeres que han experimentado algún tipo de violencia

Tipo de violencia	Frecuencia relativa
Violencia física	25%
Violencia psicológica	34%
Violencia sexual	11%
No presenta ningún tipo de violencia	30%



c) Diagrama de Tallo y Hoja:

Existe otra alternativa para estudiar la variabilidad de los datos o bien para comparar dos grupos diferentes, esta consiste en estudiar los datos usando un diagrama de Tallo y hoja. Este tiene la ventaja de retener todo el detalle de los datos originales y al mismo tiempo permite ordenarlos rápidamente.

Ejemplo 5:

Los siguientes datos corresponden al ejemplo anterior, donde se especifica la edad de los varones y damas cesantes en la comuna de La Florida.

Varones:									
25	40	31	33	41	41	40	26	28	25
Damas:									
28	45	40	31	31	40	40	28	26	45

Para construir el diagrama de tallo y hoja, dividiremos las cifras en tallos y hojas, en este caso las decenas serán el tallo y las unidades serán las hojas, por lo tanto, el diagrama quedaría así:

Varones					Damas				
8	6	5	2	6	8	8			
	3	1	3	1	1				
1	1	0	0	4	0	0	0	5	5

Podemos observar que las edades donde hay mas cesantía es entre los 40-45 años, y con mayor frecuencia en los 40 años en el caso de las damas.

d) Histograma

Consiste en dibujar, en un sistema de coordenadas, rectángulos verticales, cuya base es la longitud del intervalo de clase y cuya altura representa la frecuencia correspondiente a cada clase.

Ejemplo 6:

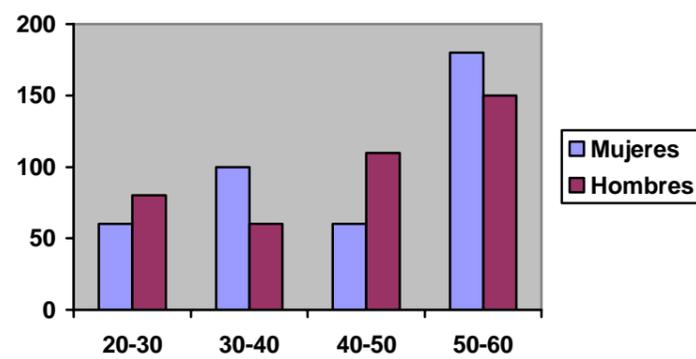
Un estudio realizado a la región metropolitana, arroja la cantidad de personas cesantes (hombres y mujeres) entre los 20 y 60 años.

Hombres

Edades	Frecuencia absoluta	Frecuencia relativa
20 – 30	80	20%
30 – 40	60	15%
40 – 50	110	27,5%
50 – 60	150	37,5%

Mujeres

Edades	Frecuencia absoluta	Frecuencia relativa
20 – 30	60	15%
30 – 40	100	25%
40 – 50	60	15%
50-60	180	45%



Observación:

También se utiliza para representar simultáneamente dos situaciones distintas de una misma variable.

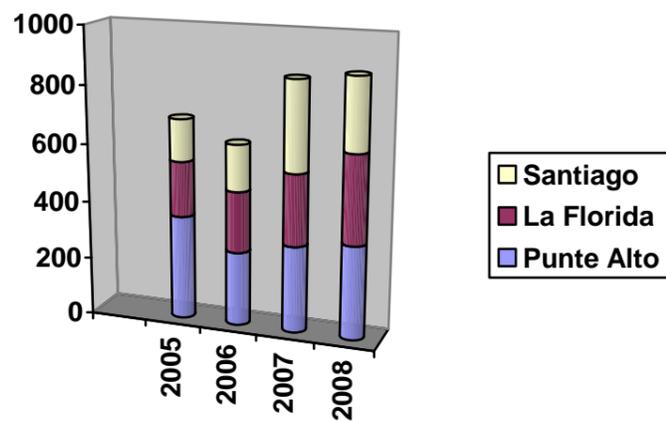
e) Pictograma

Son gráficos que están formados por figuras y tienen como ventaja que atraen al no experto, se usan frecuentemente cuando se quiere lograr el interés masivo del público, por ejemplo el daño que produce el tabaco, la obesidad infantil, etc.

Ejemplo 7:

El siguiente ejemplo refleja el consumo de cigarrillos en mujeres entre 15 y 25 años, realizado los años 2005-2006-2007-2008, en tres comunas de la región metropolitana.

Año	Comuna	Frecuencia absoluta
2005	Puente Alto	360
	La Florida	190
	Santiago	145
2006	Puente Alto	255
	La Florida	210
	Santiago	160
2007	Puente Alto	298
	La Florida	246
	Santiago	310
2008	Puente Alto	322
	La Florida	305
	Santiago	250



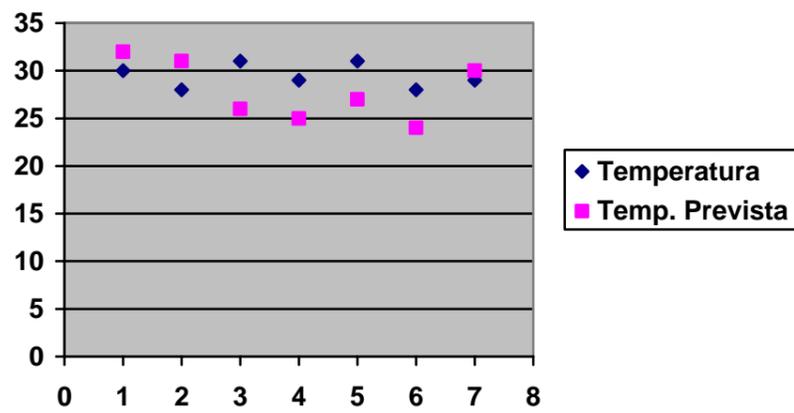
f) Grafico de dispersión

El grafico de dispersión se construye para comparar la homogeneidad o heterogeneidad de las variables.

Ejemplo 8:

Se hizo un estudio para comparar las temperaturas más altas que se pronostican y las reales en una semana

Días	Temperatura	Temperatura prevista
1	30°	32°
2	28°	31°
3	31°	26°
4	29°	25°
5	31°	27°
6	28°	24°
7	29°	30°



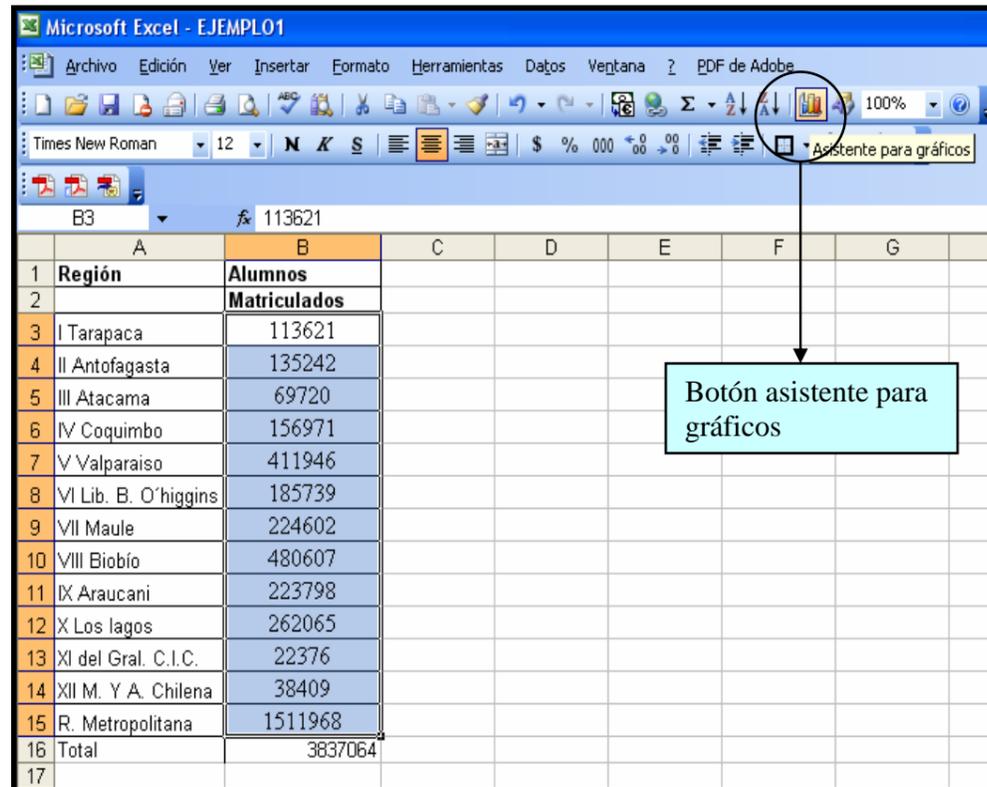
Ejemplo Excel 3

Alumnos Matriculados en la educación regular, en el país, según región, año 2008.

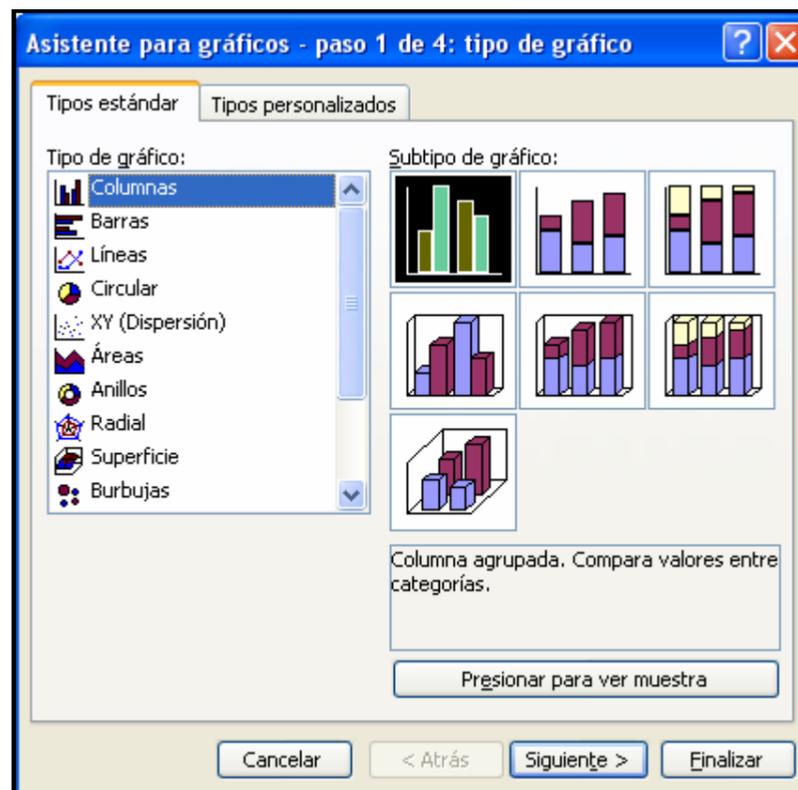
Región	Alumnos matriculados
I Tarapacá	113.621
II Antofagasta	135.242
III Atacama	69.720
IV Coquimbo	156.971
V Valparaíso	411.946
VI Lib. B. O'higgins	185.739
VII Maule	224.602
VIII Bío Bío	480.607
IX Araucanía	223.798
X Los Lagos	262.065
XI del Gral. C.I.C.	22.376
XII de M. Y A. Chilena	38.409
R. Metropolitana	1.511.968

A continuación aprenderemos a representar gráficamente, los datos de la planilla Excel.

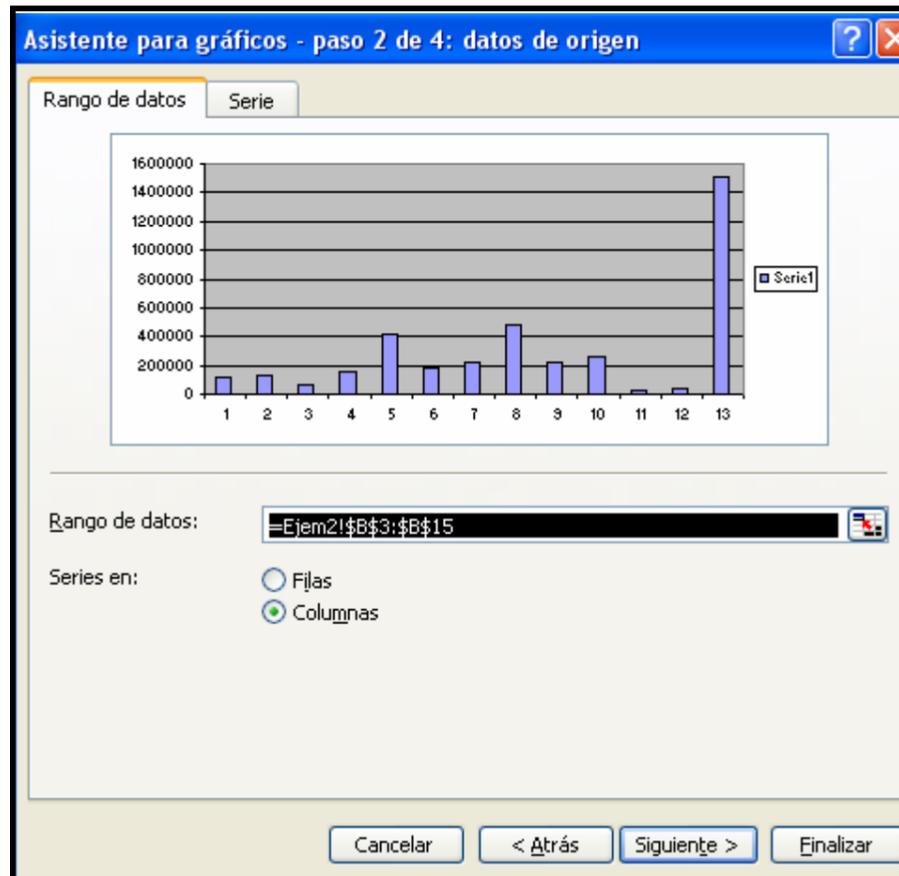
Como la variable aleatoria "Alumnos matriculados según región", es cualitativa, el gráfico que permite mostrar la distribución de frecuencias absolutas es el gráfico de barras (**columnas**). Para crearlo, seleccionamos el rango **B3:B15** (que contiene las frecuencias absolutas respectivas) y hacemos clic en el botón **Asistente para gráficos**. (Ver figura)



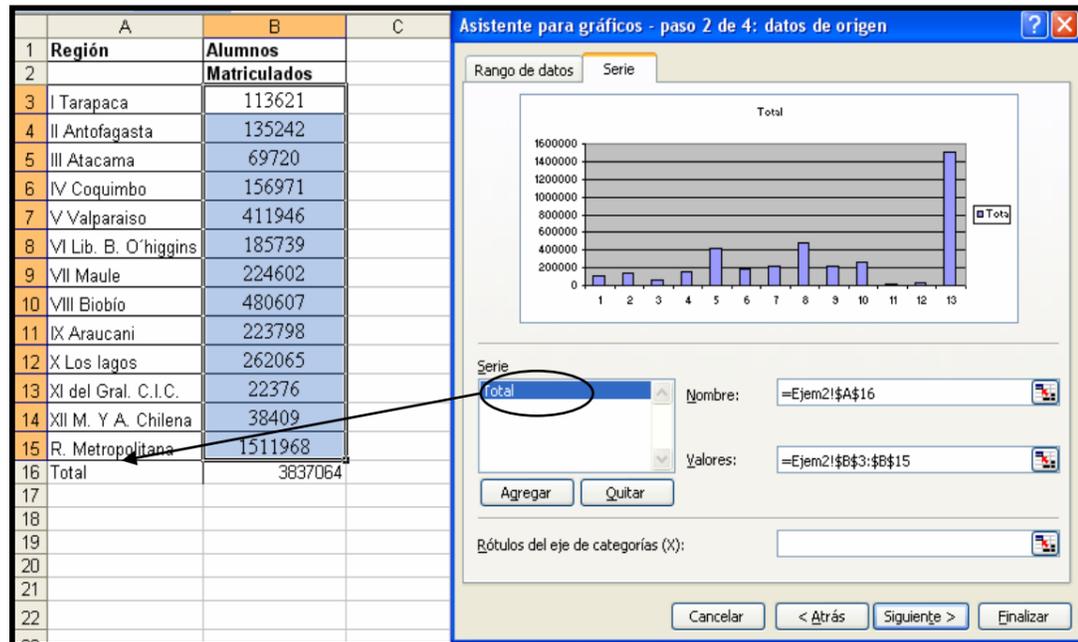
Al hacer clic sobre el botón Asistente para gráficos, aparecerá el cuadro de diálogo **Asistente para gráficos – paso 1 de 4** (Ver siguiente figura). Seleccionamos **Tipo estándar – Tipo de gráfico: Columnas – Siguiente:**



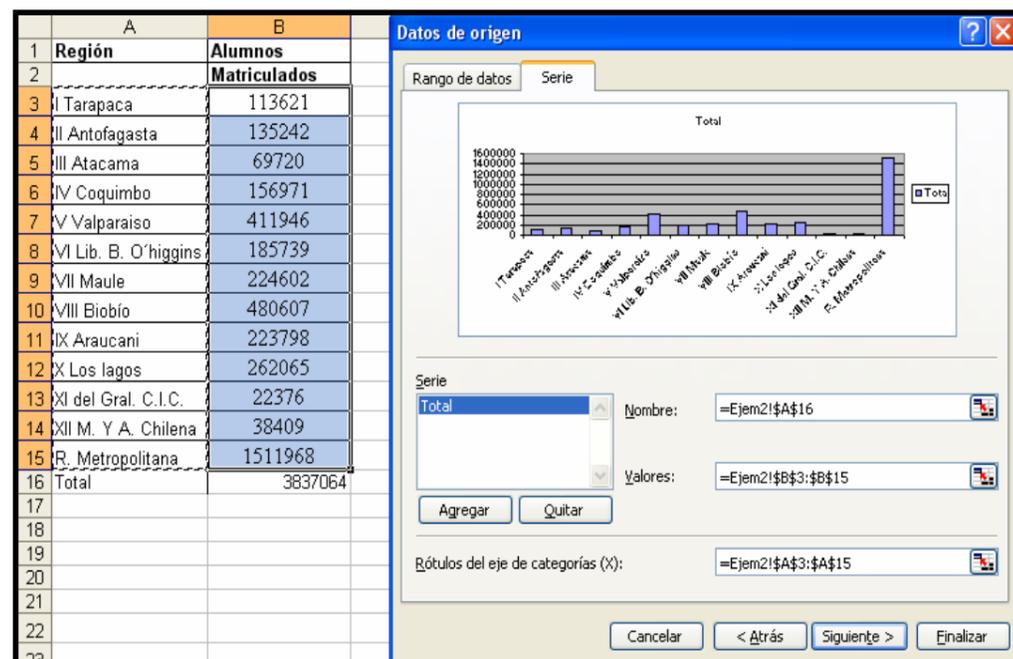
Al hacer clic en **Siguiente**, aparece el cuadro de diálogo **Asistente para gráficos – paso 2 de 4: datos de origen**. En **Rango de da datos**, **Series en**, escogemos **Columnas**, esto se aprecia en la siguiente pantalla



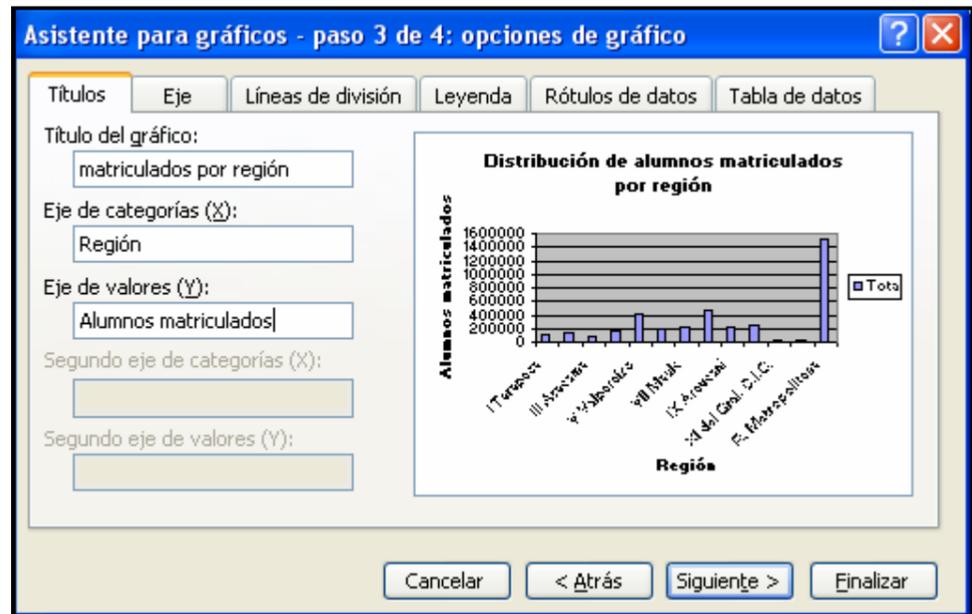
En este mismo cuadro de diálogo, hacemos clic en la aleta **Serie**, en **Serie** aparecerá **Serie 1**. Aquí escogemos serie 1, luego hacemos clic en **Nombre** y después seleccionamos **A16**, que es el total



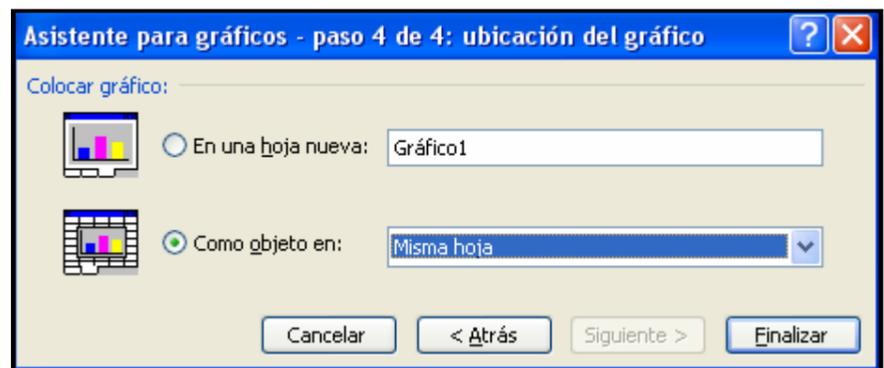
En **Rótulo del eje de categorías (X)** seleccionamos el rango que contiene a los nombres de las distintas regiones del país (en este ejemplo, el rango **A3:A15**)



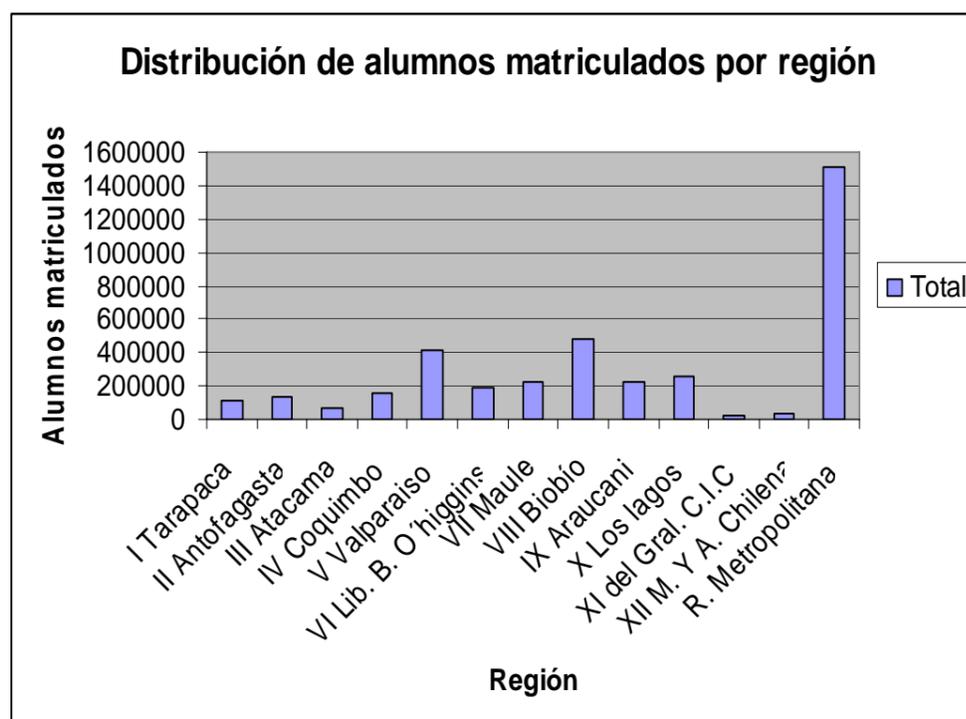
Al hacer clic en **Siguiente**, aparece la ventana **Asistente para gráficos: Paso 3 de 4**. En **Título del gráfico**, escribimos "Distribución de alumnos chilenos matriculados, según región". En **Eje de Categorías (X)**, escribimos "Región"



Haciendo clic en siguiente, aparecerá la opción **Ubicación del gráfico**. Dejamos el gráfico en la misma hoja de trabajo.



Al hacer clic en **Finalizar**, se obtiene el gráfico esperado:



Es importante destacar que todo gráfico estadístico debe llevar su correspondiente título, que explique en qué consiste la descripción del gráfico en cuestión.

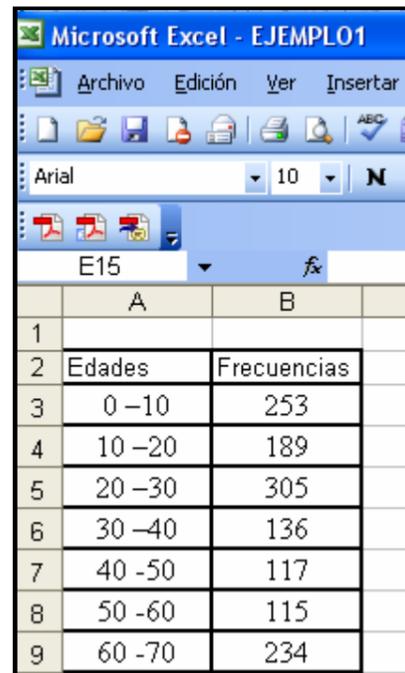
Ejemplo Excel 4

En la tabla siguiente se presenta edades, en años, de una muestra de 1000 personas que sufrieron accidentes en el hogar y que recurrieron a cierto centro asistencial.

EDAD (AÑOS)	Nº DE PERSONAS
0 –10	253
10 –20	189
20 –30	305
30 –40	136
40 -50	117
50 -60	115
60 -70	234

En este ejemplo, la variable “edades, en años”, es una variable continua, el gráfico que permite mostrar la distribución de frecuencias absolutas es el **histograma**.

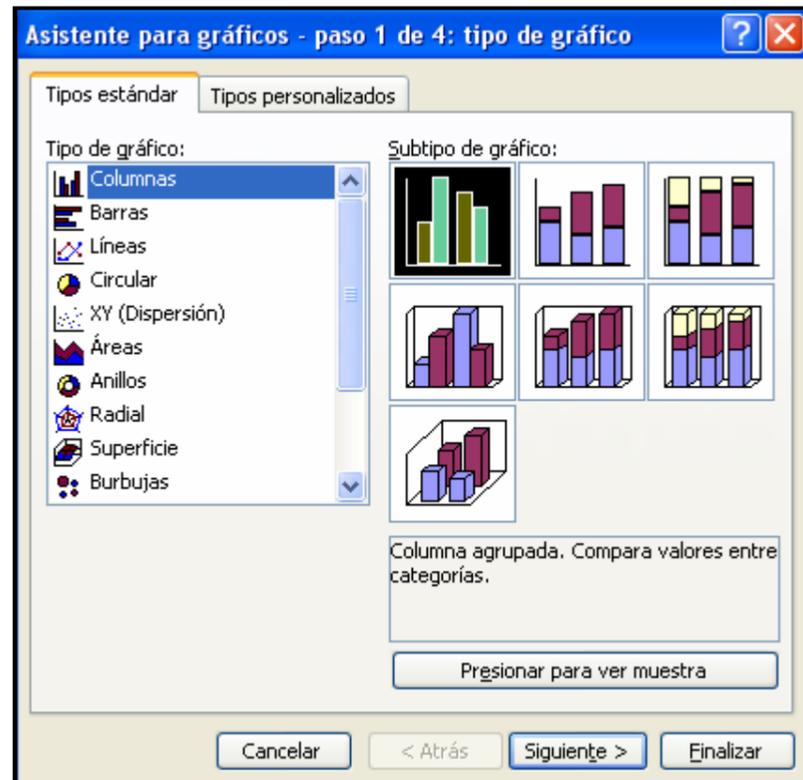
Ingresamos los datos en la planilla Excel, como se ve a continuación.



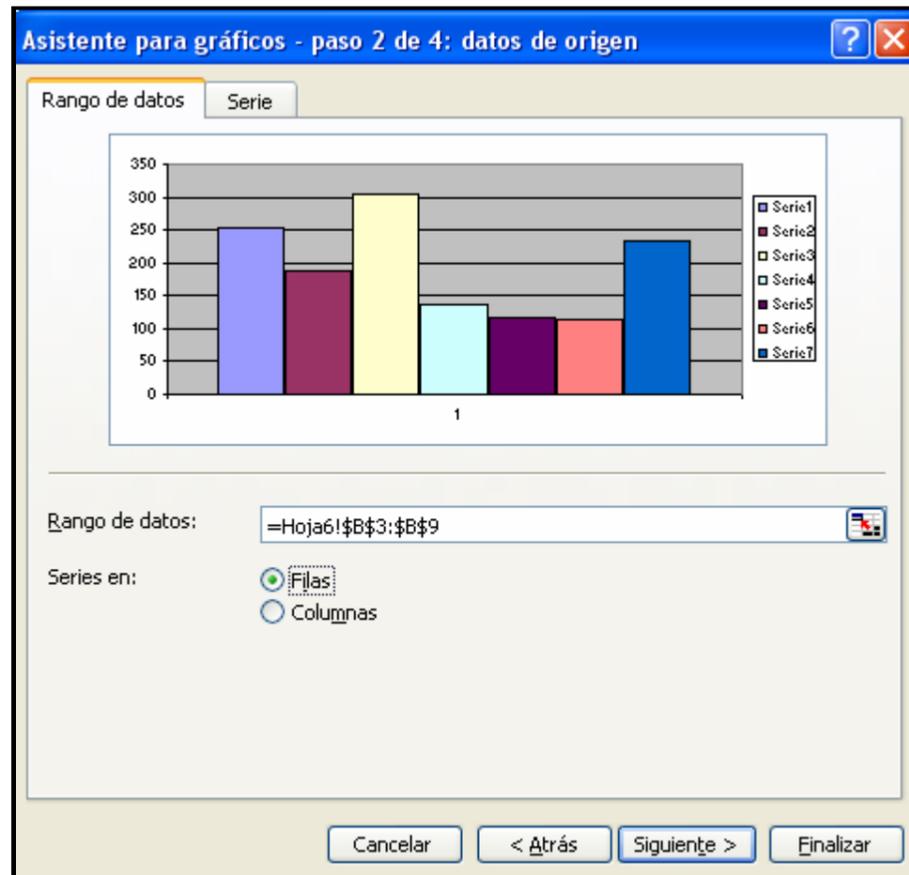
The screenshot shows the Microsoft Excel interface with a spreadsheet titled "EJEMPLO1". The spreadsheet has two columns: "Edades" (Ages) and "Frecuencias" (Frequencies). The data is as follows:

	A	B
1		
2	Edades	Frecuencias
3	0 -10	253
4	10 -20	189
5	20 -30	305
6	30 -40	136
7	40 -50	117
8	50 -60	115
9	60 -70	234

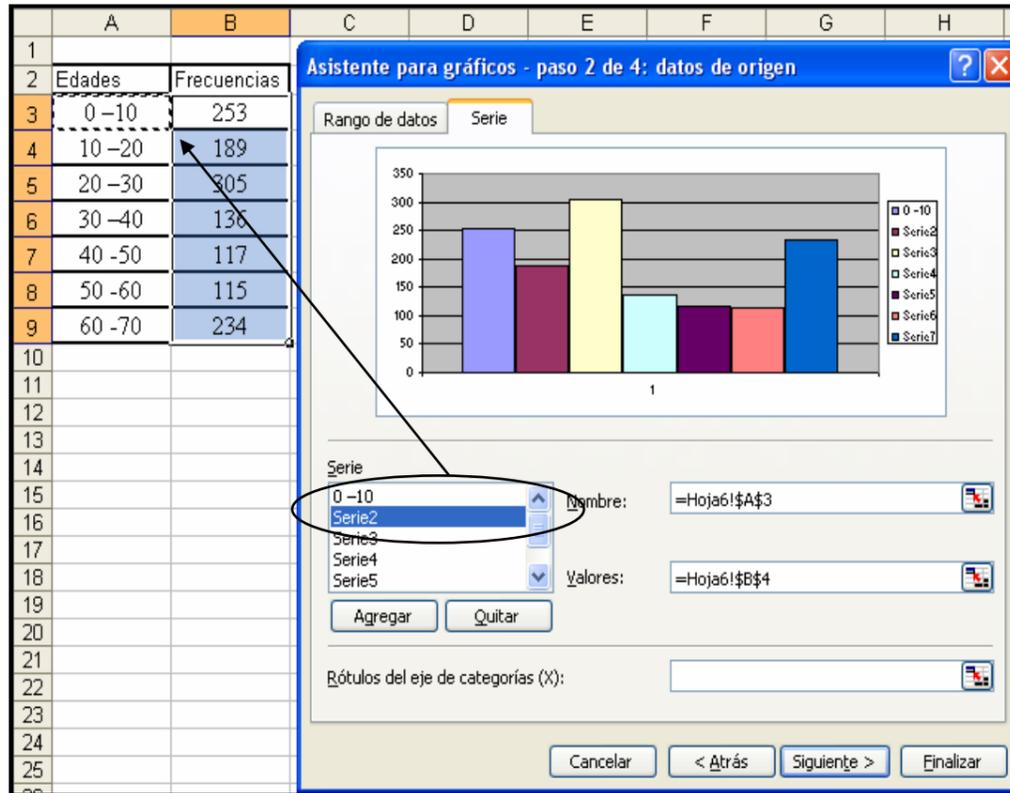
Para crearlo, seleccionamos el rango **B3:B9** (que contiene las frecuencias absolutas respectivas) y hacemos clic en el botón **Asistente para gráficos – paso 1 de 4**. Seleccionamos **Tipo estándar – Tipo de gráfico: Columnas – Siguiente:** (Ver figura)



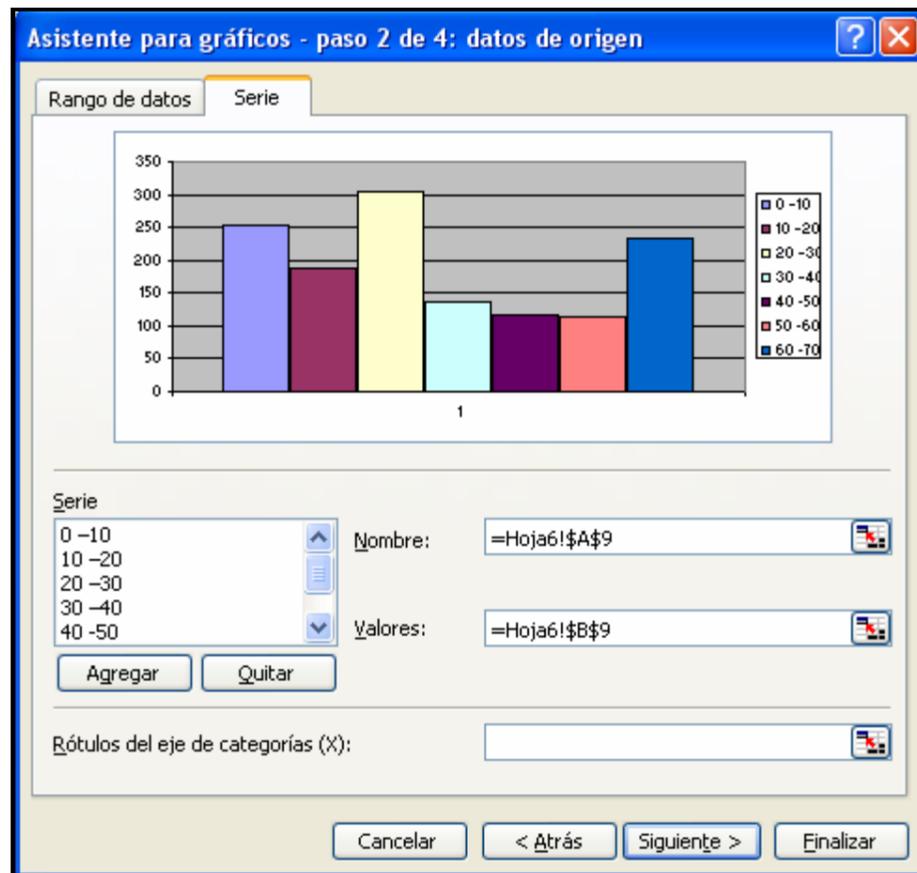
Al hacer clic en **Siguiete**, aparece el cuadro de diálogo **Asistente para gráficos-paso 2 de 4: datos de Origen**. En **Series en:** escogemos **Filas** (pues de esa forma están dispuestas las diversas frecuencias absolutas asociadas a los diferentes intervalos de clasificación).



En este mismo cuadro de diálogo, hacemos clic en la aleta **Serie**, donde aparecerá una lista de (**Serie1, serie2, serie3, etc.**). Luego de esto hacemos clic en serie1, llevando el cursor a la opción nombre y seleccionando nuestra primera categoría en este caso A3, y así sucesivamente, como se observa en la figura.

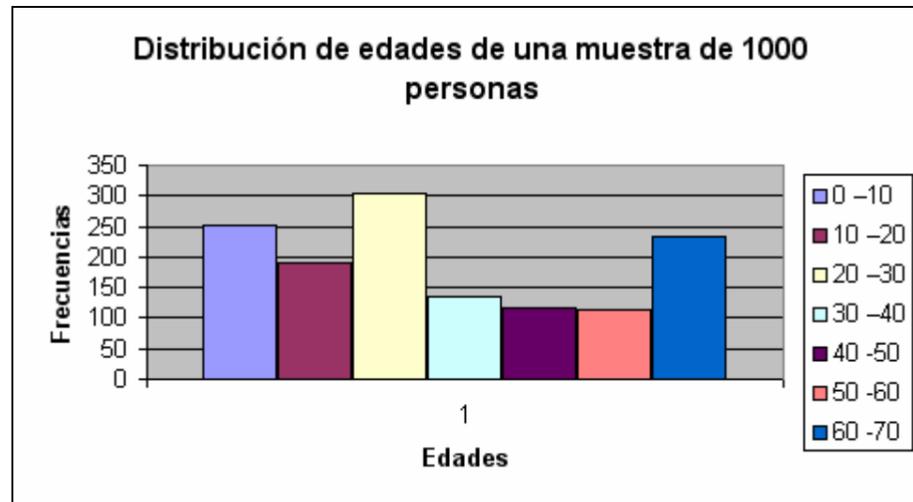


Entonces así escogemos los intervalos asociados a cada categoría. Esto se aprecia en la siguiente pantalla:



Al hacer clic en **Siguiente**, aparecerá el cuadro de diálogo **Asistente para gráficos** – **paso 3 de 4: Opciones de Gráfico**. Aquí es posible dar título al gráfico, dar los nombres de los ejes X e Y, entre otros atributos.

El grafico resultante es el siguiente:



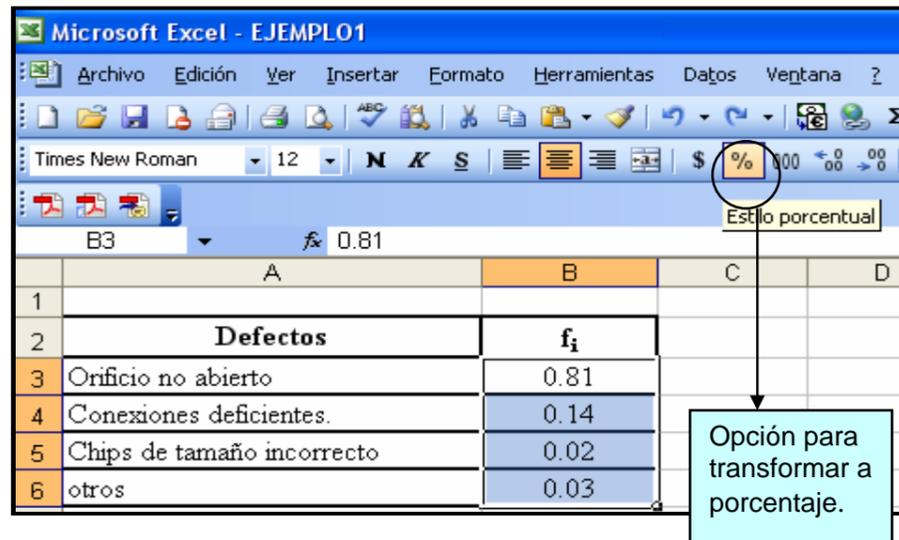
Ejemplo Excel 5

De 2000 chips de computadores inspeccionados por un fabricante. Se registraron los siguientes números de defectos:

Defectos	f_i
Orificio no abierto	0.81
Conexiones deficientes.	0.14
Chips de tamaño incorrecto	0.02
Otros	0.03

Construiremos el grafico circular, ya que tenemos la frecuencia relativa y este grafico es esencialmente para representar porcentajes.

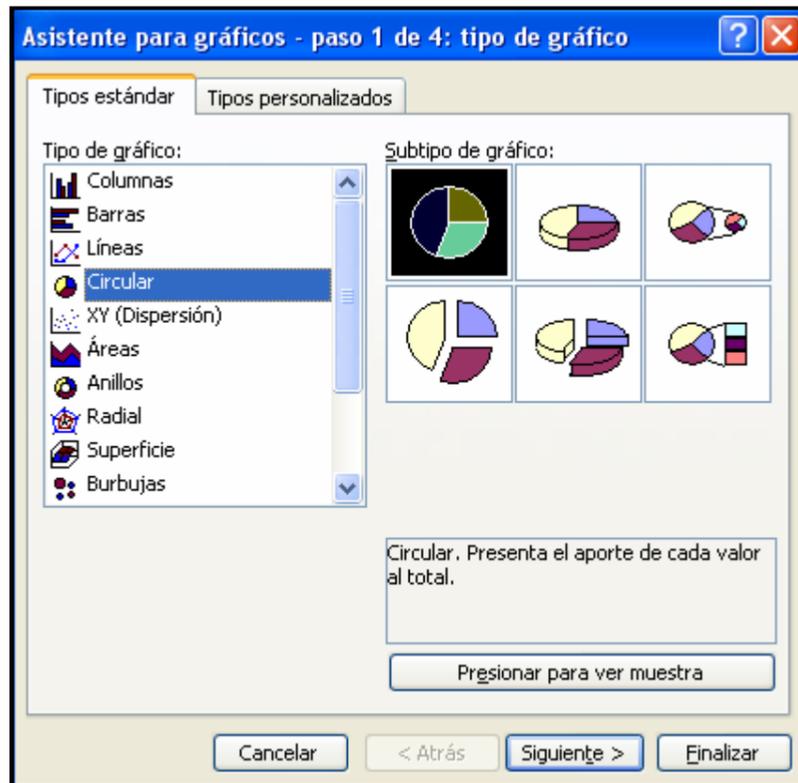
Para construir el grafico circular con porcentajes, vamos a convertir los decimales a porcentajes para esto seleccionamos de **B3:B6**, luego elegimos la opción estilo porcentual (%) de la siguiente manera:



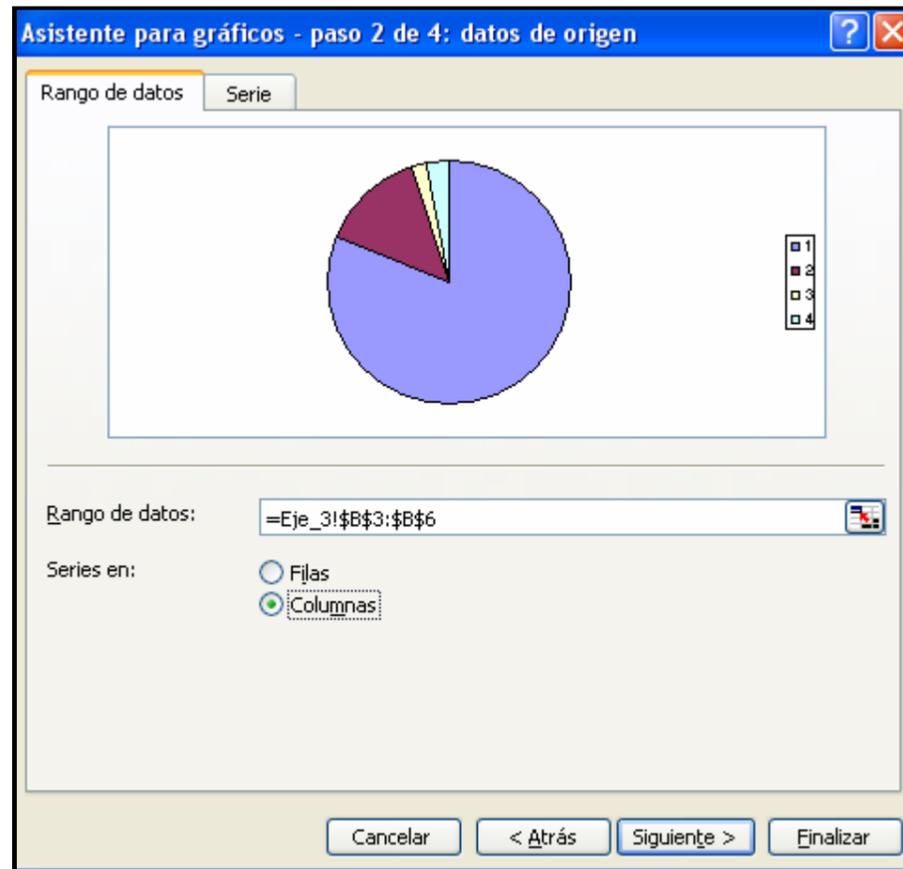
Obteniendo como resultado la siguiente tabla:

	A	B
1		
2	Defectos	f_i
3	Orificio no abierto	81%
4	Conexiones deficientes.	14%
5	Chips de tamaño incorrecto	2%
6	otros	3%

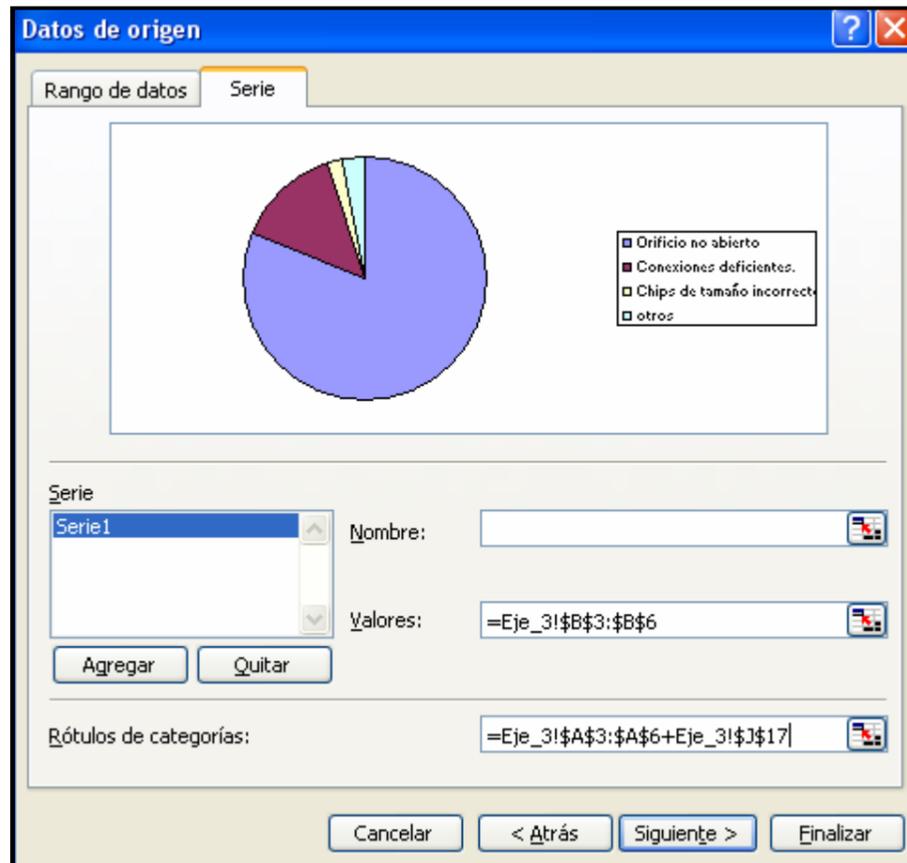
En el cuadro de diálogo **Asistente para gráficos – paso 1 de 4: tipo de gráfico**, seleccionamos **Tipo de Gráfico: Circular**



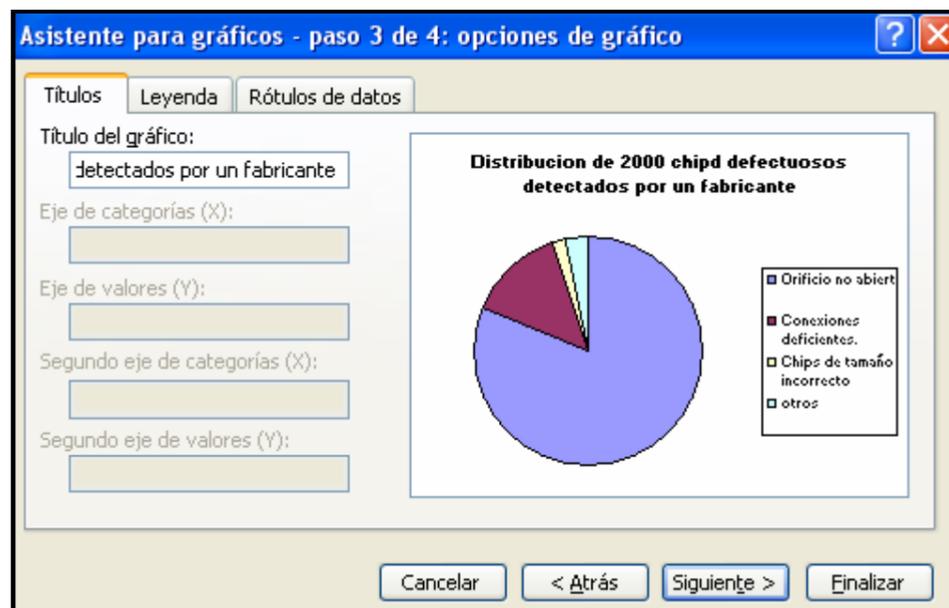
Al hacer clic en el botón **Siguiete**, aparecerá el cuadro de diálogo **Asistente para gráficos – paso 2 de 4**. Escogemos la opción **Series en: Columnas**.



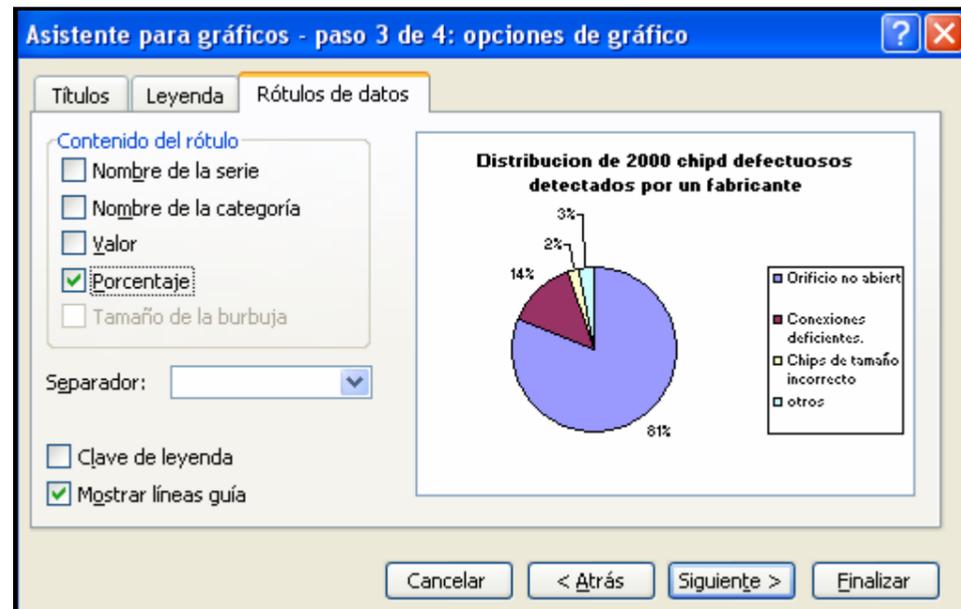
En la aleta **Serie**, en **Rótulo de categorías**, seleccionamos las celdas que contienen los nombres de las dos distintas clasificaciones (en este caso: orificio no abierto, conexiones deficientes, chips de tamaño incorrecto, otros, esto es, el rango **A3:A6**)



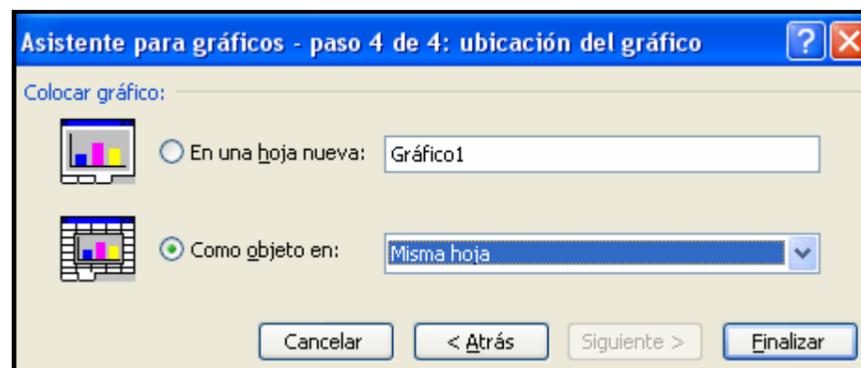
Al hacer clic en **Siguiete**, se mostrará el cuadro de dialogo **Asistente para gráficos: paso 3 de 4**. En la aleta **Títulos**, escribimos un título para el gráfico. En este caso escribimos: "Distribución de 2000 chips defectuosos detectados por un fabricante"



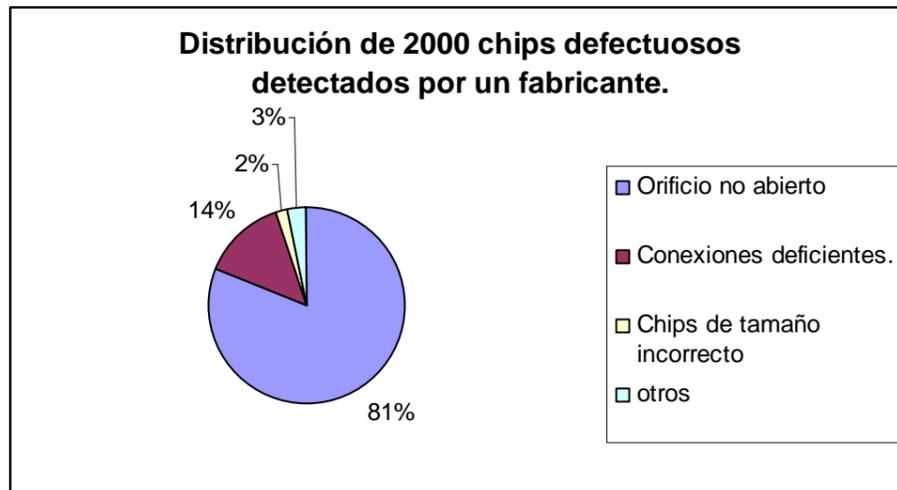
Para mostrar una medida cuantitativa de la distribución de Chips defectuosos, hacemos clic en **Rótulos de datos**, seleccionamos **Porcentaje**.



Al presionar el botón **Siguiete**, aparecerá el cuadro **Asistente para gráficos – paso 4 de 4: Ubicación del gráfico**. Aquí podemos escoger dónde se situará finalmente nuestro gráfico (ya sea, en una hoja nueva o bien, dentro de la misma hoja). En nuestro caso, ubicaremos el gráfico en la misma hoja.



Obteniendo así el siguiente grafico circular:



3.1 Propuesta de trabajo Excel

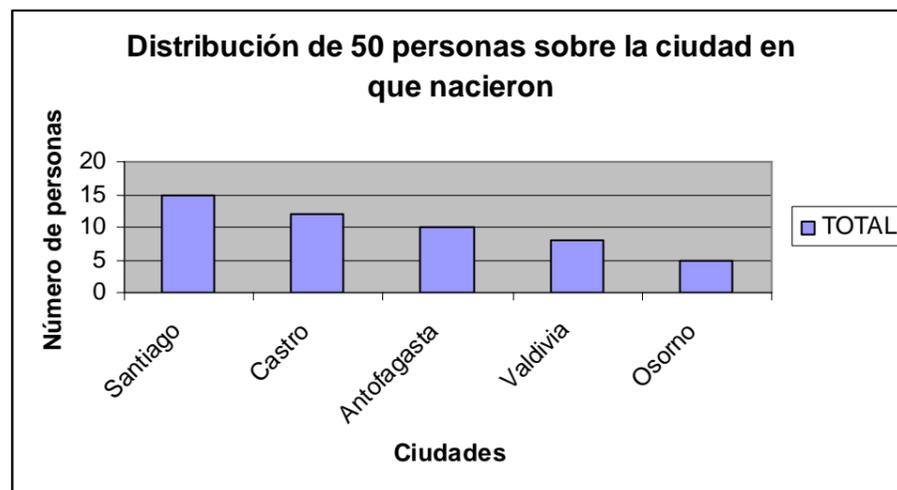
Actividad 1

Una encuesta realizada a 50 personas sobre la ciudad donde nacieron, entregó los siguientes datos.

CIUDAD DONDE NACIERON	NUMERO DE PERSONAS
Santiago	15
Castro	12
Antofagasta	10
Valdivia	8
Osorno	5
	50

Construir el grafico correspondiente según los datos.

El resultado esperado de la actividad anterior se muestra en la siguiente imagen:



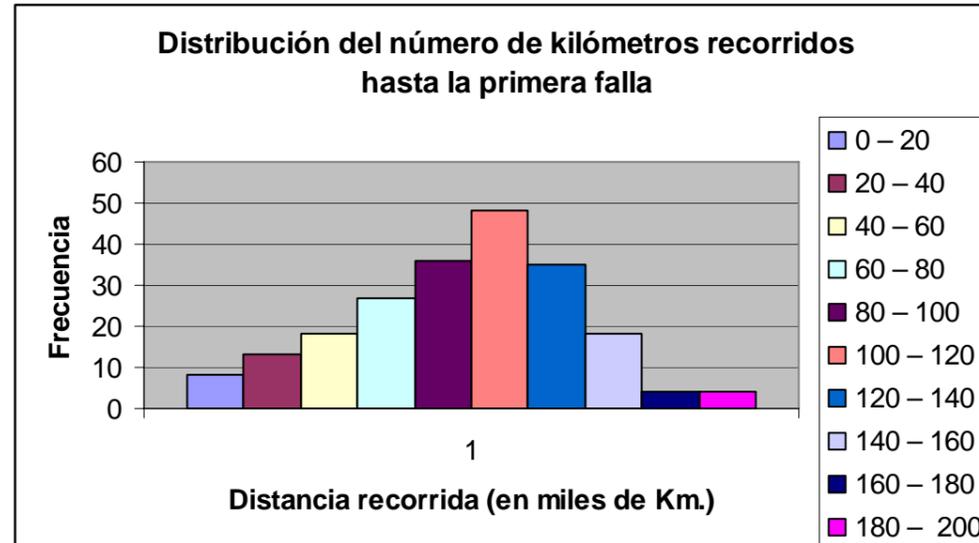
Actividad 2

La siguiente tabla presenta la distribución de frecuencias del número de kilómetros recorridos hasta que se produjo el primer fallo grave de motor en cierto modelo de vehículo:

Distancia recorrida (en miles de Km.)	Frecuencia
0 – 20	8
20 – 40	13
40 – 60	18
60 – 80	27
80 – 100	36
100 – 120	48
120 – 140	35
140 – 160	18
160 – 180	4
180 – 200	4

Siguiendo los pasos anteriores realizar el histograma respectivo.

El resultado esperado, es el siguiente:



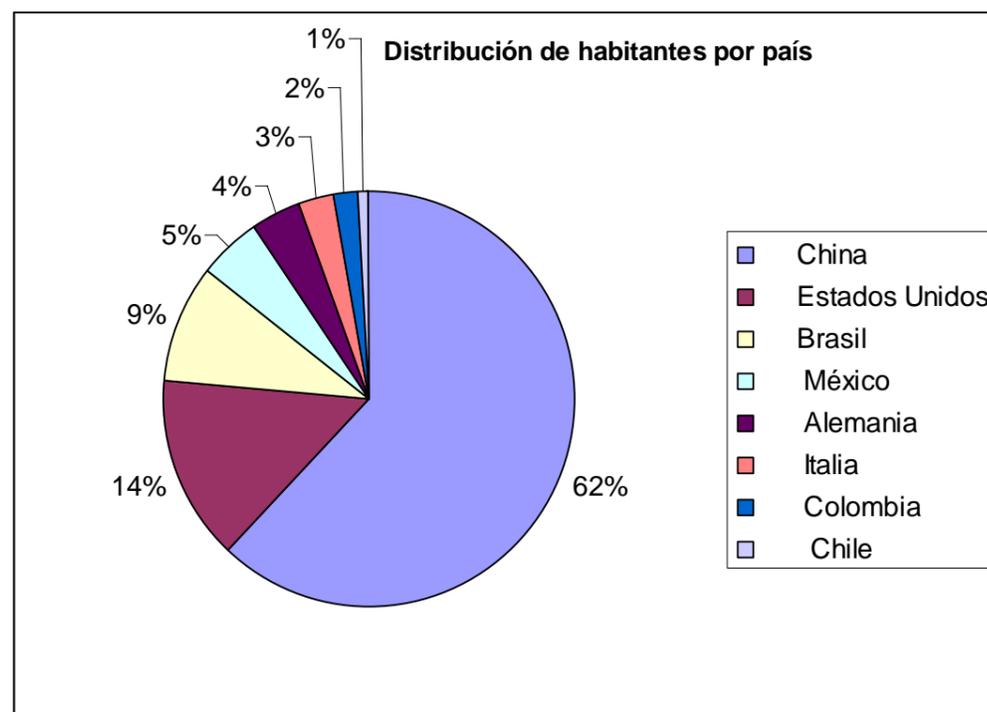
Actividad 3

La siguiente tabla indica los números de habitantes por países a mediados del año 2009.

País	Numero de habitantes
 China	1.338.613.000
 Estados Unidos	307.212.000
 Brasil	198.739.000
 México	111.212.000
 Alemania	82.230.000
 Italia	58.126.000
 Colombia	45.644.000
 Chile	16.602.000

Construir el grafico circular correspondiente a la información anterior

El resultado esperado para la actividad, se refleja en la siguiente imagen



3.2. PONGAMOS A PRUEBA LO APRENDIDO

Actividad 1

Los siguientes datos corresponden a la tasa bruta de natalidad y ,mortalidad ionfantil de algunos países de Latinoamérica.

Natalidad (niños nacidos vivos e un año, por cada mil habitantes)

21	47	29	27	23	33
28	29	35	33	18	28

Mortalidad (Número de muertes al año por cada mil habitantes, niños menores de un año)

26	51	63	40	17	63
56	43	42	109	22	23

Construya un diagrama de tallo y hoja para los datos anteriores

Actividad 2

La recaudación anual de impuestos de 40 contribuyentes está dada en la siguiente tabla (en miles de pesos)

<i>Impuesto</i>	<i>Frecuencia</i>
50 – 70	6
70 – 90	10
90 – 110	8
110 – 130	9
130 – 150	7

Construir el histograma correspondiente a la tabla de datos.

Actividad 3

La empresa constructora "Autopista Central" encargada de las nuevas vías de circulación del gran Santiago, ha notado que se ha incrementado el número de accidentes entre sus operarios en los últimos dos años. Con el objeto de estudiar un programa de prevención, han decidido tomar un año como muestra. La información recopilada relativa al número de accidentes semanalmente durante un año es la siguiente.

<i>N° de accidentes</i>	<i>N° DE SEMANAS</i>
3-5	10
5-7	12
7-9	8
9-11	13
11-13	9

Construir el pictograma, según los datos, de la forma más conveniente.

Actividad 4

La información que se presenta en la siguiente tabla corresponde al número de artículos vendidos en cierto día, por una muestra de vendedores de una gran tienda comercial.

<i>N° de artículos</i>	<i>N° de vendedores</i>
2	12
4	21
5	15
6	4
7	8

Construir el gráfico de dispersión correspondiente a los datos.

Actividad 5

Dada la siguiente información dibujar un gráfico circular porcentual y uno de barras.

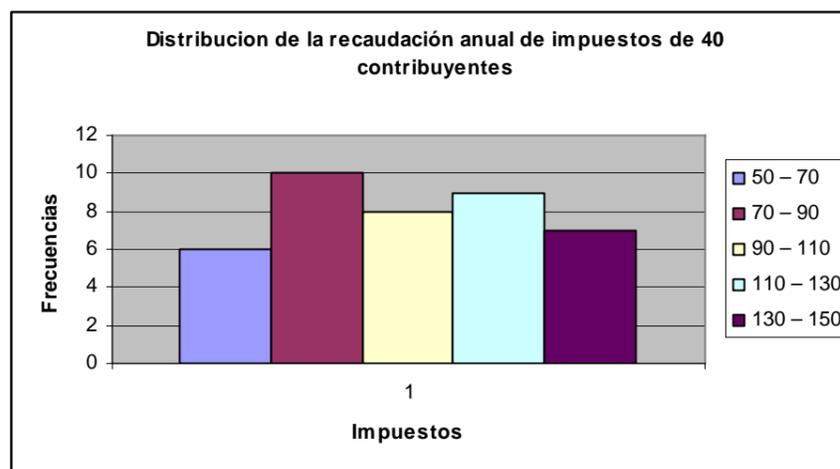
Edad	Frecuencia	Porcentaje	Grados
18	10	0.13	46.8
20	15	0.19	68.4
22	25	0.31	111.6
24	30	0.37	133.2

3.3. SOLUCIONARIO

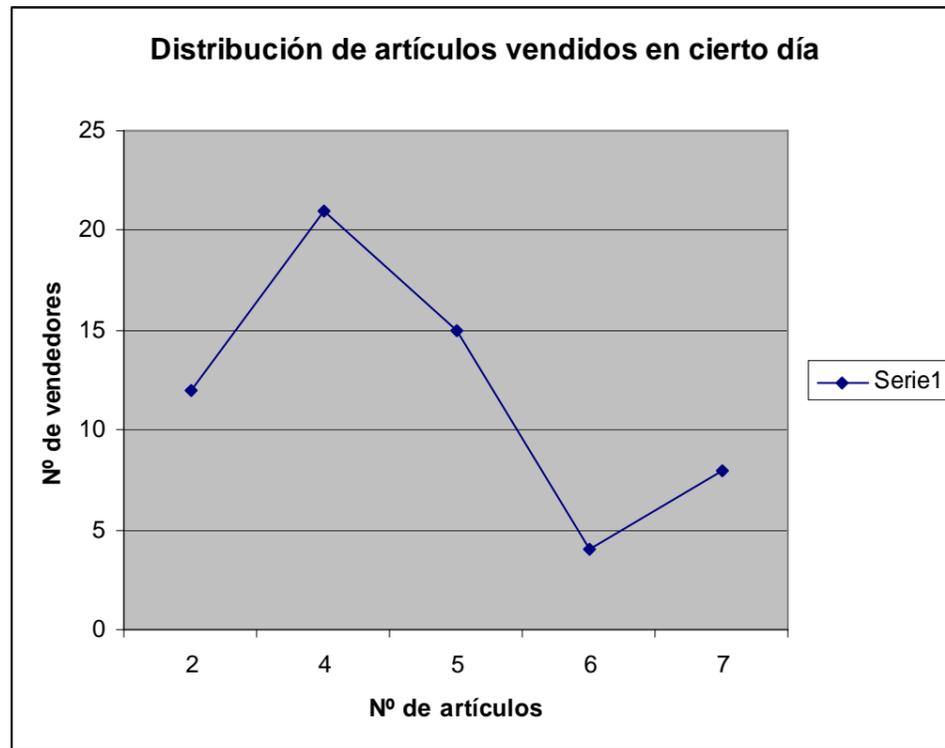
Actividad 1:

<u>Natalidad</u>		<u>Mortalidad</u>	
	8 1		7
9 9 8 8 7 3 1	2	2 3 6	
5 3 3 3			
7 4		0 2 3	
5		1 6	
6		3 3	
7			
8			
9			
10		9	

Actividad 2:

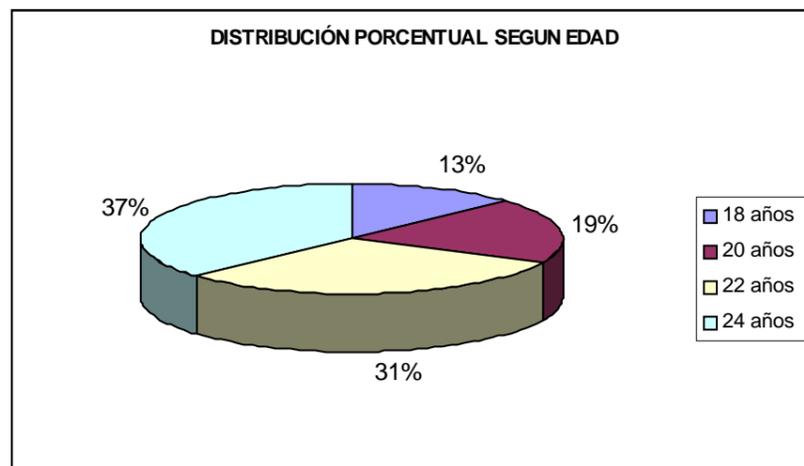


Actividad 4:

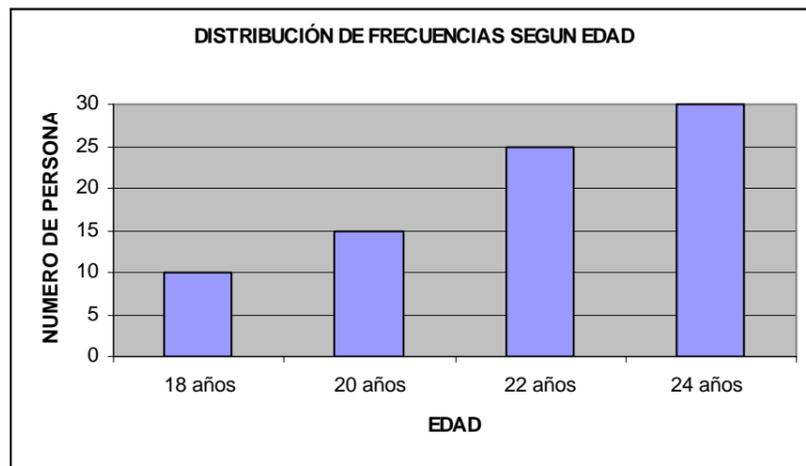


Actividades 5

a) Gráfico circular porcentual:



b) Gráfico de barras:



4.- MEDIDAS ESTADÍSTICAS

Después de confeccionar las tablas de distribución de frecuencias, es preciso calcular medidas estadísticas para analizar la información y comparar entre si dos o más muestras o poblaciones, ya que normalmente la inteligencia humana no es capaz de efectuar dichos análisis o comparaciones por imposibilidad material de retener los datos precisos para ello.

Las medidas estadísticas más usadas son las medidas de tendencia central y las medidas de variabilidad.

4.1. Medidas de tendencia central

Las medidas de tendencia central, de cuales nosotros estudiaremos la media, mediana y moda son valores alrededor los cuales se distribuye una población, tales valores tienden a situarse en el centro del recorrido de la variable y en torno a ellos se agrupan las observaciones.

4.1.1. Media aritmética

La media aritmética que denotaremos por \bar{x} , se calcula sumando todos los valores anotados por una variable y dividiéndolo por el numero de observaciones. Obviamente el número de observaciones es exactamente el tamaño de la población o muestra.

Media aritmética es la cantidad total de la variable distribuida en partes iguales entre cada observación, punto de equilibrio de las distancias de la variable.

Si $\{x_1, x_2, x_3, \dots, x_n\}$ es un conjunto de datos, entonces la media aritmética es:

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Si los datos están agrupados en una tabla de distribución de frecuencias, entonces la media aritmética es:

$$\bar{X} = \frac{x_1 n_1 + x_2 n_2 + x_3 n_3 + \dots + x_k n_k}{n} = \frac{\sum_{i=1}^k x_i n_i}{n}$$

Donde las x_i son los valores de la variable discreta o las marcas de clase de la variable continua y las n_i son las respectivas frecuencias absolutas.

La media aritmética también se puede calcular usando las frecuencias relativas

Teorema:

$$M(x) = \sum_{i=1}^k x_i f_i$$

Demostración.

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + \dots + x_k n_k}{n} = x_1 \frac{n_1}{n} + x_2 \frac{n_2}{n} + \dots + x_k \frac{n_k}{n}$$

$$= x_1 f_1 + x_2 f_2 + \dots + x_k f_k = \sum_{i=1}^k x_i f_i$$

Ejemplo 9

Los siguientes datos corresponden a una población de 8 individuos a los cuales se les aplico un test de inteligencia, 105- 103- 108- 110- 115- 118- 115- 115.

Calcular la media aritmética.

$$\bar{X} = \frac{105+103+108+110+115+118+115+115}{8} = 111.125$$

¿Qué podemos concluir con el resultado obtenido? El promedio de los 8 individuos en el test de inteligencia es de 111.125

Ejemplo 10

La siguiente tabla corresponde a los sueldos semanales que paga en la actualidad una empresa comercial (en miles de \$):

Sueldo	x_i	n_i	$x_i \cdot n_i$
80 – 100	90	7	525
100 – 120	110	20	2200
120 – 140	130	33	4290
140 – 160	150	25	3750
160 – 180	170	11	1870
180 – 200	190	4	760
		100	$\sum x_i \cdot n_i = 133.95$

$$\bar{X} = \frac{\sum x_i \cdot n_i}{n} = \frac{13395}{100} = 133.95$$

El promedio de sueldos semanales que pagan en la actualidad una empresa comercial es 133.95 miles de pesos.

Utilizando el teorema con la frecuencia relativa

Sueldo	x_i	n_i	f_i	$x_i \cdot f_i$
80 – 100	90	7	0.07	5.25
100 – 120	110	20	0.2	22
120 – 140	130	33	0.33	42.9
140 – 160	150	25	0.25	37.5
160 – 180	170	11	0.11	18.7
180 – 200	190	4	0.04	7.6
		100	$\sum f_i = 1$	$\sum x_i \cdot f_i = 13395$

PROPIEDADES DE LA MEDIA ARITMETICA

1. La media aritmética de una constante es la misma constante
2. La suma de las desviaciones de los valores de la variable respecto de la media aritmética, es cero. Es decir:

$$\sum_{i=1}^n (x_i - \bar{x}) = 0 \text{ para datos no agrupados}$$

$$\sum_{i=1}^k (x_i - \bar{x}) n_i = 0 \text{ para datos agrupados}$$

Demostración.

Se demostrará la propiedad para datos agrupados

$$\begin{aligned}\sum_{i=1}^k (x_i - \bar{x}) n_i &= \sum_{i=1}^k (x_i n_i - \bar{x} n_i) = \sum_{i=1}^k x_i n_i - \sum_{i=1}^k \bar{x} n_i \\ &= n \bar{x} - \bar{x} \sum n_i = n \bar{x} - \bar{x} n = 0\end{aligned}$$

3. La suma de las desviaciones cuadráticas de los valores de la variable respecto de la media aritmética, es mínima. Es decir,

$$\sum_{i=1}^n (x_i - \bar{x})^2 \text{ es mínima (para datos no agrupados)}$$

$$\sum_{i=1}^k (x_i - \bar{x})^2 n_i \text{ es mínima (para datos agrupados)}$$

4. Es una medida que toma en consideración todos los valores de la distribución. Por esta misma razón es muy sensible a la presentación de observaciones extremas o anómalas que hacen que la media se desplace hacia ellas. En consecuencia no es recomendable usar la media como medida de tendencia central en estos casos, pues la cantidad obtenida no es representativa del total de los datos.

a. Tiene la ventaja de que es única y siempre se puede calcular en variables medidas en a lo menos una escala de intervalos

b.- La media es el promedio más usado. A partir de ella se puede obtener el **valor total** (Valor suma de todas las observaciones)

$$\text{Total} = n \cdot \bar{x} = \sum_{i=1}^n x_i$$

4.1.1.1. Media aritmética ponderada

Existen variables cuyos valores poseen, por su naturaleza, distintas importancias dentro de un conjunto. En estos casos la media aritmética de la variable no se calcula de la forma en que fue definida, sino afectando cada valor de la variable por un coeficiente o peso o ponderación.

La media aritmética ponderada se calcula por:

$$\bar{x} = \frac{x_1 w_1 + x_2 w_2 + x_3 w_3 + \dots + x_p w_p}{w_1 + w_2 + w_3 + \dots + w_p} = \frac{\sum_{i=1}^p x_i w_i}{\sum_{i=1}^p w_i}$$

Ejemplo 11

Un estudiante universitario de pedagogía en matemáticas de la universidad Católica Silva Henríquez, desea saber con que nota se presenta al examen para aprobar el ramo probabilidad y estadística.

Las notas son las siguientes:

	Notas	Ponderación (%)
Prueba 1	3.8	25
Prueba 2	4.2	20
Talleres	5.6	15
Ayudantía	6.0	10
Trabajo final	5.8	30

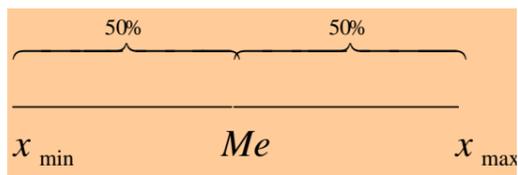
$$\bar{x} = \frac{3.8 \cdot 25 + 4.2 \cdot 20 + 5.6 \cdot 15 + 6.0 \cdot 10 + 5.8 \cdot 30}{25 + 20 + 15 + 10 + 30} = \frac{430.8}{100} = 4.308 \approx 4.3$$

Un estudiante de la universidad Católica Silva Henríquez se presenta al examen con nota 4.3 ponderada.

4.1.2. Mediana

La mediana de un conjunto de datos numéricos ordenados en forma creciente o decreciente, es el dato que se encuentra al centro de dicha ordenación (x_i), o la media aritmética de los datos centrales (en caso de que la muestra tenga un grupo de datos pares).

Mediana: Para un conjunto de datos ordenados, la mediana, denotada por **Me**, es aquel valor de la variable que supera a no más del 50% de los datos y es superado por no más del 50% restante.



Mediana para datos no agrupados

Su cálculo se realiza, una vez ordenados los datos, del siguiente modo:

- Si el número de datos es impar, el valor de la mediana será el que ocupa el lugar central, en la posición $\frac{n + 1}{2}$
- Si el número de observaciones es par, la mediana será el promedio de los dos valores centrales, en las posiciones $\frac{n}{2}$ y $\frac{n}{2} + 1$.

Por lo tanto la mediana para datos no agrupados es:

$$Me = \begin{cases} x_{\frac{n+1}{2}} & \text{si } n \text{ es impar} \\ \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} & \text{si } n \text{ es par} \end{cases}$$

Ejemplo 12

La mediana para el conjunto de datos 15, 13, 13, 14, 14, 17, 19, 18, 16 se calcula de la siguiente manera:

El conjunto ordenado es 13, 13, 14, 14, 15, 16, 17, 18, 19

$$n = 9 \text{ impar, se calcula } \frac{n+1}{2} = \frac{9+1}{2} = \frac{10}{2} = 5$$

La mediana es el dato que ocupa la quinta posición, luego $Me = x_5 = 15$

Ejemplo 13

La mediana para el conjunto de datos 33, 39, 38, 37, 36, 33, 35, 34, 36, 34, se calcula de la siguiente manera:

El conjunto de datos ordenados es 33, 33, 34, 34, 35, 36, 36, 37, 38, 39

$$n = 10 \text{ par, se calcula } \frac{n}{2} = \frac{10}{2} = 5 \text{ y } \frac{n}{2} + 1 = 5 + 1 = 6$$

la mediana es el promedio entre x_5 y x_6 , por lo tanto la mediana es

$$Me = \frac{x_5 + x_6}{2} = \frac{35 + 36}{2} = 35,5$$

Mediana para datos agrupados

La mediana es aquel valor de la variable cuya frecuencia absoluta acumulada es inmediatamente mayor a la mitad de las observaciones

Para calcular la mediana se usa la siguiente fórmula de interpolación.

$$Me = x'_{j-1} + \frac{\frac{n}{2} - N_{j-1}}{n_j} \cdot a_j$$

Donde:

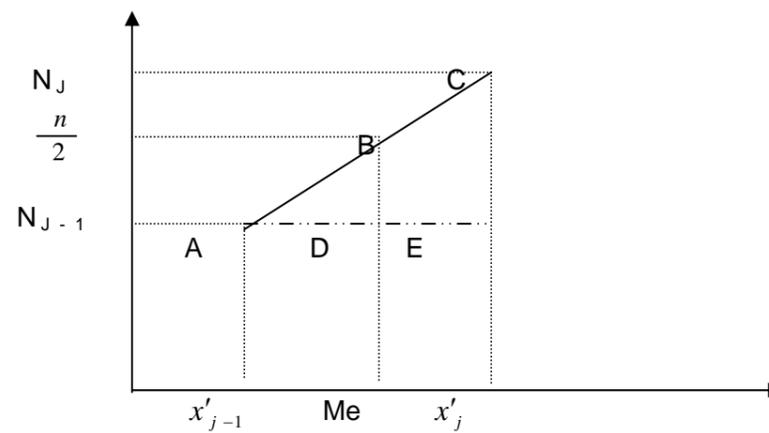
x'_{j-1} : Límite inferior del intervalo que contiene a la mediana

n_j : Frecuencia absoluta del intervalo que contiene a la mediana

a_j : Amplitud del intervalo que contiene a la mediana

N_{j-1} : Frecuencia absoluta acumulada del intervalo anterior al que contiene la mediana

Demostración:



Los triángulos ADB y AEC son triángulos rectángulos semejantes, luego

$$\frac{AD}{AE} = \frac{AB}{AC}$$

Como

$$\left. \begin{array}{l} AD = Me - x'_{j-1} \\ AE = x'_j - x'_{j-1} = a_j \\ AB = \frac{n}{2} - N_{j-1} \\ AC = N_j - N_{j-1} = n_j \end{array} \right\} \Rightarrow \frac{Me - x'_{j-1}}{a_j} = \frac{\frac{n}{2} - N_{j-1}}{n_j} \Rightarrow$$

$$Me = x'_{j-1} + \frac{\frac{n}{2} - N_{j-1}}{n_j} a_j$$

Siguiendo el **Ejemplo 10** anterior, la mediana sería la siguiente

Primero tenemos que calcular la mitad de las observaciones: $\frac{n}{2} = \frac{100}{2} = 50$

Ahora vemos el valor de la frecuencia absoluta acumulada que es inmediatamente mayor a 50, la que nos indicara en que intervalo esta contenida la mediana.

Sueldo	x_i	n_i	N_i
80 - 100	90	7	7
100 - 120	110	20	27
120 - 140	130	33	60
140 - 160	150	25	85
160 - 180	170	11	96
180 - 200	190	4	100
		100	

$$Me = x'_{j-1} + \frac{\frac{n}{2} - N_{j-1}}{n_j} \cdot a_j$$

$$Me = 120 + \frac{\frac{100}{2} - 27}{33} \cdot 20 = 133,94$$

El 50% de los sueldos que paga una empresa comercial es a lo más \$133,94.

4.1.3. Moda

En el lenguaje cotidiano, lo que está de moda es lo que más se lleva. En estadística la idea es la misma.

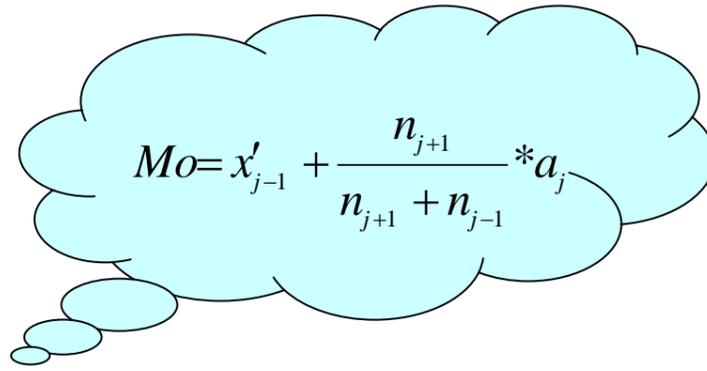
Moda es el valor con mayor frecuencia y la designamos por **Mo**. En un gráfico de frecuencias la moda corresponde al punto más alto, es decir, el valor de la variable que presenta mayor frecuencia.

La moda en un conjunto de valores no es necesariamente única. Tenemos distribuciones unimodales (una moda), bimodales (2 modas), etc. También puede ocurrir que una distribución no presenta moda.

Ejemplo 14

- 1.- En el conjunto {9,3,16,5,12,4,1,7} no existe moda
- 2.- En el conjunto {9,3,3,3,5,12,5,1,5,7,3} la moda es 3
- 3.- En el conjunto {9,8,8,8,5,1,7,5,3,5,6,1} las modas son 8 y 5

Si los valores para un dato se han agrupado en intervalos, entonces primero se determina el intervalo modal, aquel que posee la mayor frecuencia absoluta o relativa. En segundo lugar, para calcularla, se usa la fórmula de interpolación:



$$Mo = x'_{j-1} + \frac{n_{j+1}}{n_{j+1} + n_{j-1}} * a_j$$

Donde:

x'_{i-1} : límite inferior del intervalo modal (aquel que contiene a la moda)

a_i : amplitud del intervalo modal

n_{j+1} : frecuencia absoluta del intervalo posterior al modal

n_{j-1} : es la frecuencia absoluta del intervalo anterior al modal

Ahora calculamos la moda

Sueldo	x_i	n_i	N_i
80 – 100	90	7	7
100 – 120	110	20	27
120 – 140	130	33	60
140 – 160	150	25	85
160 – 180	170	11	96
180 – 200	190	4	100
		100	

En el intervalo donde hay mayor frecuencia es en el tercero, por lo tanto cada dato se ingresa según este intervalo.

$$Mo = 120 + \frac{25}{25 + 20} * 20 = 131.11$$

Por lo tanto la mayor cantidad de sueldo que paga una empresa comercial es de \$131.11.

Ejemplo Excel 6

En el Ejemplo Excel 2 realizado anteriormente, la tabla finalizó así:

	A	B	C	D	E
1					
2					
3					
4	Limite inferior	Limite superior	Marca de clase (xi)	Frecuencia Absoluta (ni)	Frecuencia Relativa (fi)
6	50	100	75	7	7%
7	100	120	110	20	20%
8	120	140	130	33	33%
9	140	160	150	25	25%
10	160	180	170	11	11%
11	180	200	190	4	4%
12				100	

Para calcular la media aritmética agregaremos una columna que multiplique la marca de clase por la frecuencia absoluta ($x_i \bullet n_i$).

	A	B	C	D	E	F
1						
2						
3						
4	Limite inferior	Limite superior	Marca de clase (xi)	Frecuencia Absoluta (ni)	Frecuencia Relativa (fi)	xi * ni
6	50	100	75	7	0.07	
7	100	120	110	20	0.20	
8	120	140	130	33	0.33	
9	140	160	150	25	0.25	
10	160	180	170	11	0.11	
11	180	200	190	4	0.04	
12				100		

Para calcular los valores de la columna $x_i \bullet n_i$ introducimos en **F6** la fórmula = **C6•D6**, y arrastramos el Mouse desde **F6** hasta **F11**. Resultando lo siguiente

	A	B	C	D	E	F
1						
2						
3						
4	Limite inferior	Limite superior	Marca de clase (xi)	Frecuencia Absoluta (ni)	Frecuencia Relativa (fi)	xi * ni
6	50	100	75	7	0.07	525
7	100	120	110	20	0.20	2200
8	120	140	130	33	0.33	4290
9	140	160	150	25	0.25	3750
10	160	180	170	11	0.11	1870
11	180	200	190	4	0.04	760
12				100		

La media de esta distribución se puede obtener dividiendo la suma (utilizando la función **SUMA** de **F6** A **F11**) de la columna $x_i \bullet n_i$ por la cantidad de datos (**D12**). En Excel, escribimos en una celda vacía la expresión = F12 /D12, obteniendo el siguiente valor:

	A	B	C	D	E	F
1						
2						
3						
4	Limite inferior	Limite superior	Marca de clase (xi)	Frecuencia Absoluta (ni)	Frecuencia Relativa (fi)	xi * ni
6	50	100	75	7	0.07	525
7	100	120	110	20	0.20	2200
8	120	140	130	33	0.33	4290
9	140	160	150	25	0.25	3750
10	160	180	170	11	0.11	1870
11	180	200	190	4	0.04	760
12				100		13395
13						
14			Media aritmetica		133.95	
15						

De la tabla de distribución de frecuencias anterior, podemos obtener fácilmente el valor de la **mediana**. Agregamos una columna para las frecuencias absolutas acumuladas, donde la primera frecuencia absoluta se mantiene (**G6 = D6**) y luego **G7 = G6+D7**, y entonces arrastramos de **G7** a **G11**, resultando lo siguiente.

	A	B	C	D	E	F	G
1							
2							
3							
4	Limite inferior	Limite superior	Marca de clase (xi)	Frecuencia Absoluta (ni)	Frecuencia Relativa (fi)	xi * ni	Frec. Abs. Acumulada
6	50	100	75	7	0.07	525	7
7	100	120	110	20	0.20	2200	27
8	120	140	130	33	0.33	4290	60
9	140	160	150	25	0.25	3750	85
10	160	180	170	11	0.11	1870	96
11	180	200	190	4	0.04	760	100
12				100		13395	
13							
14			Media aritmetica		133.95		
15							

Vemos que la frecuencia acumulada que supera a $100/2 = 50$ es 60. Luego la mediana está en el intervalo (120,140]. Para calcularla, usamos la expresión

$$Me = x'_{j-1} + \frac{\frac{n}{2} - N_{j-1}}{n_j} \cdot a_j$$

En este ejemplo particular, escribimos en la celda E15 la fórmula para determinar la mediana:

$$f_x = A8 + ((D12/2 - G7)/D8) * (B8 - A8)$$

Obteniendo el valor:

	A	B	C	D	E	F	G
1							
2							
3							
4	Limite inferior	Limite superior	Marca de clase (xi)	Frecuencia Absoluta (ni)	Frecuencia Relativa (fi)	xi * ni	Frec. Abs. Acumulada
6	50	100	75	7	0.07	525	7
7	100	120	110	20	0.20	2200	27
8	120	140	130	33	0.33	4290	60
9	140	160	150	25	0.25	3750	85
10	160	180	170	11	0.11	1870	96
11	180	200	190	4	0.04	760	100
12				100		13395	
13							
14			Media aritmetica		133.95		
15			Mediana		133.939394		

Quando los datos están tabulados como variable continua, es posible estimar el valor de la moda **solamente** cuando la distribución es **unimodal**. En este caso, el intervalo modal es (120,140], puesto que presenta la mayor frecuencia absoluta, igual a 33. Usando la expresión

$$Mo = x'_{j-1} + \frac{n_{j+1}}{n_{j+1} + n_{j-1}} * a_j$$

En este ejemplo particular, en la celda E16 vemos que la fórmula en Excel será

$$=A8+(D9/(D7+D9))*(B8-A8)$$

Cuyo valor es,

	A	B	C	D	E	F	G
1							
2							
3							
4	Limite inferior	Limite superior	Marca de clase (xi)	Frecuencia Absoluta (ni)	Frecuencia Relativa (fi)	xi * ni	Frec. Abs. Acumulada
6	50	100	75	7	0.07	525	7
7	100	120	110	20	0.20	2200	27
8	120	140	130	33	0.33	4290	60
9	140	160	150	25	0.25	3750	85
10	160	180	170	11	0.11	1870	96
11	180	200	190	4	0.04	760	100
12				100		13395	
13							
14			Media aritmetica		133.95		
15			Mediana		133.939394		
16			Moda		131.111111		

Es importante notar que el valor de la media, mediana y moda, se pueden determinar de manera exacta, pues Excel incorpora dentro de la herramienta **Análisis de Datos**, la opción **Estadística Descriptiva**. Para poder usar esta herramienta, debemos agrupar los datos originales en una columna (también se pueden agrupar en una fila, siempre y cuando sean menos de 256 observaciones). Como veremos a continuación en el siguiente ejemplo

Ejemplo Excel 7

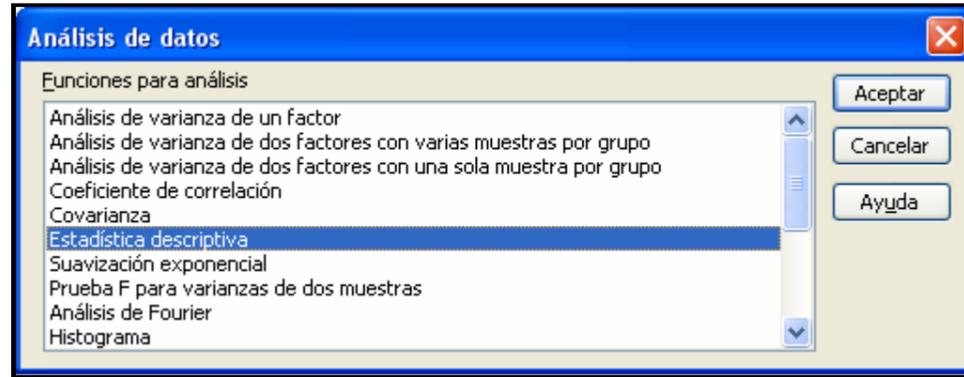
La siguiente información corresponde al número de piezas que los empleados de una fábrica logran armar en un tiempo de 12 horas de trabajo.

4	4	9	9	9	4	6	8	8	8	8	4	4	5	5	6
4	4	9	7	7	7	7	7	7	9	9	4	4	5	5	6
9	4	9	4	7	6	6	6	6	6	7	6	6	4	4	5

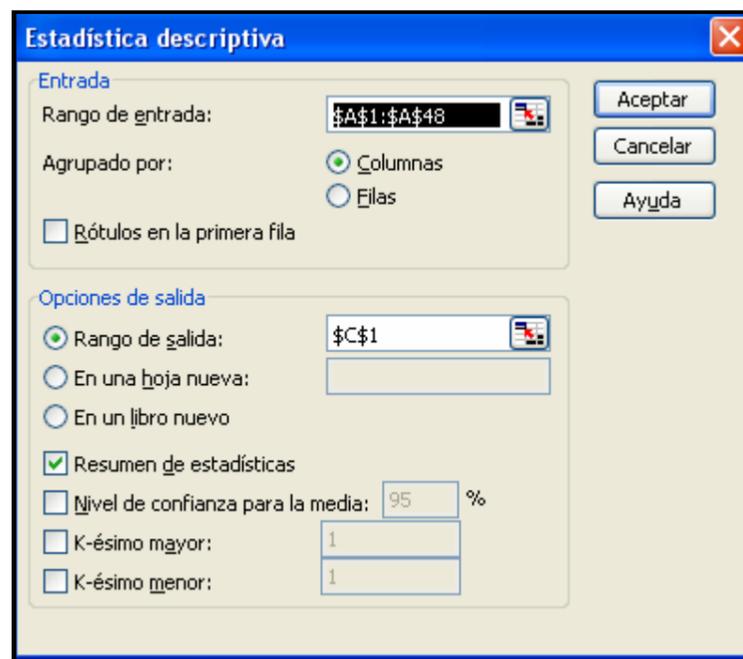
	A
1	4
2	4
3	9
4	4
5	4
6	4
7	9
8	9
9	9
10	9
11	7
12	4
13	9
14	7
15	7
16	4
17	7
18	6
19	6
20	7
21	6
22	8
23	7
24	6
25	8

Los numero de piezas fabricadas por 48 trabajadores están ingresadas en la Columna A. Por razones de espacio, sólo se muestran en la pantalla las primeras 24.

Hacemos clic en **Herramientas – Análisis de Datos – Estadística Descriptiva**.
Se muestra la siguiente salida:



Al hacer clic en **ACEPTAR**, aparecerá la ventana **Estadística Descriptiva**. En **Rango de Entrada**, seleccionamos los 48 números de piezas (en este ejemplo, el rango **A1:A48**). En **Agrupado por**, seleccionamos **Columnas**. En **Rango de Salida**, marcamos la celda **C1**, y finalmente marcamos la celda **Resumen de Estadísticas**. La siguiente salida contiene los pasos explicados:



Haciendo clic en **ACEPTAR**, se obtiene el siguiente resultado:

	A	B	C	D
1	4		Columna1	
2	4			
3	9		Media	6.1875
4	4		Error típico	0.25898097
5	4		Mediana	6
6	4		Moda	4
7	9		Desviación es	1.7942728
8	9		Varianza de l	3.21941489
9	9		Curtosis	-1.20755179
10	9		Coficiente de	0.23664181
11	7		Rango	5
12	4		Mínimo	4
13	9		Máximo	9
14	7		Suma	297
15	7		Cuenta	48
16	4			
17	7			
18	6			
19	6			
20	7			

La media exacta del número de piezas fabricadas por 48 trabajadores es 6,1875.

La mediana exacta es de 6 (resultado que se hace evidente **al ordenar de menor a mayor los datos**. Como hay una cantidad par de datos, la mediana se obtiene promediando las estaturas que ocupan las posiciones 24 y 25),

Observar la siguiente salida:

22	6
23	6
24	6
25	6
26	6
27	6
28	6
29	7

Promedio de los datos centrales

$$\frac{x_{24} + x_{25}}{2} = \frac{6 + 6}{2} = 6$$

Finalmente, se puede observar que esta distribución es **unimodal**, al ver que 4 es el número de piezas más frecuente dentro del conjunto de datos (13 trabajadores fabricaron 4 piezas en 12 horas).

4.1.4. Propuesta de trabajo Excel.

En un experimento la rama de psicología, pide a varios individuos que memoricen cierta secuencia de palabras. Los datos representan los tiempos en segundos, que necesitaron los participantes del experimento de la memorización

100	100	130	102	88	88	149	79	110	117
89	127	112	93	106	96	85	79	125	53
107	149	62	93	99	66	123	70	118	62
75	107	69	138	112	61	146	43	93	142
126	128	102	73	135	122	30	135	98	135
98	109	90	88	99	50	108	80	114	45
119	105	34	87	52	78	41	84	64	94
95	100	100	129	129	58	79	37	88	123
57	110	32	145	85	126	118	96	77	107
99	76	50	57	64	46	73	90	116	103

- Construya una tabla de frecuencia con 8 intervalos, en Excel, calcular media aritmética, mediana y moda.
- Ingresar los datos en una planilla Excel y mediante la opción estadística descriptiva, calcular media aritmética, mediana y moda.
- Comparar los resultados.

El resultado esperado para la actividad

15									
16	Limite	Limite	Marca de	Frecuencia	$x_i \cdot n_i$	Frec. Abs.			
17	Inferior	Superior	clase	Absoluta		Acumulada	amplitud		15
18	30	45	37.5	6	225	6	Media		93.9
19	45	60	52.5	9	472.5	15	Mediana		95.25
20	60	75	67.5	11	742.5	26	Moda		97.5
21	75	90	82.5	17	1402.5	43			
22	90	105	97.5	20	1950	63			
23	105	120	112.5	17	1912.5	80			
24	120	135	127.5	11	1402.5	91			
25	135	150	142.5	9	1282.5	100			
26				100	9390				

<i>Columna1</i>	
Media	93.51
Error típico	2.9261446
Mediana	96
Moda	100
Desviación es	29.261446
Varianza de l	856.232222
Curtosis	-0.60118242
Coefficiente de	-0.21266187
Rango	119
Mínimo	30
Máximo	149
Suma	9351
Cuenta	100

4.1.5. PONGAMOS A PRUEBA LO APRENDIDO

Actividad 1

Los siguientes datos son determinaciones de la corriente máxima anual de un río en metros cúbicos por segundo, 405- 355 – 419 – 267 – 370 – 391 – 612 – 383 – 434

Calcular e interpretar

- a) Media aritmética
- b) Mediana
- c) Moda

Actividad 2

Una publicidad sobre automóviles muestra la siguiente distribución de kilómetros recorridos por galón de gasolina, según información de algunos propietarios encuestados:

Kilómetros por galón	Número de propietarios
15 - 16	6
16 - 17	10
17 - 18	16
18 - 19	24
19 - 20	14

- a) ¿Qué recorrido mínimo registran el 9% de los autos más económicos?
- b) ¿Qué porcentaje de autos recorren entre 19 y 20 kilómetros por galón?
- c) Calcular mediana, moda y media aritmética e interpretarlas.

Actividad 3

El ingeniero a cargo del control de calidad del agua de una ciudad, es responsable del nivel de cloración del agua. Dicho nivel debe estar muy aproximado al que exige el Servicio de Salud correspondiente. Para vigilar el nivel de cloro, el ingeniero realiza un muestreo diariamente a algunos galones que contienen el agua que se desea analizar.

La tabla anexa exhibe las concentraciones de cloro correspondientes a 39 galones seleccionados como muestra de un día en partes por millón (p.p.m.)

16.2	15.7	16.4	15.4	16.4	15.8	16.0	15.2	15.7	16.6	15.8	16.2	15.9
15.9	15.6	15.8	16.1	15.9	16.0	15.6	16.3	16.8	15.9	16.3	16.9	15.6
16.0	16.8	16.0	16.3	16.4	15.9	15.8	15.7	16.4	16.9	16.7	16.0	15.9

- i) Organice los datos en una tabla de 5 intervalos, luego interprete lo siguiente.
- ii) ¿Cual es la concentración de cloro con mayor frecuencia?
- iii) ¿Cuál es la concentración de cloro promedio correspondiente a los 39 galones?
- iv) ¿Cuál es la mediana según los datos?

Actividad 4

Complete la siguiente tabla de distribución de frecuencias, sabiendo que la amplitud de los intervalos es constante y que la media aritmética es 4,94

INTERVALOS	Y_i	n_i	f_i	N_i	$Y_i n_i$
1 -					
-		20			
5 -					180
-				90	
-		10			

Calcular: mediana y moda.

4.1.6. SOLUCIONARIO

Actividad 1:

a) Media Aritmética

$$\bar{X} = \frac{405 + 355 + 419 + 267 + 370 + 391 + 612 + 383 + 434}{9} = 404$$

El promedio de corrientes máximas de un río en metros cúbicos por segundo es de 404.

b) Mediana

Ordenamos de menor a mayor los datos:

$$267 - 355 - 370 - 383 - 391 - 405 - 419 - 434 - 612$$

Luego: Cantidad de datos impar por lo tanto.

$$\bullet \quad X_{\frac{n+1}{2}} = X_{\frac{9+1}{2}} = X_5 = 391$$

El 50% de las corrientes máximas son a lo más 391.

c) Moda

$$267 - 355 - 370 - 383 - 391 - 405 - 419 - 434 - 612$$

Nos podemos dar cuenta que no existe ningún dato que se repite por lo tanto no existe moda.

Actividad 2:

Kilómetros por galón	Número de propietarios	(xi)	fi	Ni	xi-ni
15 - 16	6	15.5	0.09	6	93
16 - 17	10	16.5	0.14	16	165
17 - 18	16	17.5	0.23	32	280
18 - 19	24	18.5	0.34	56	444
19 - 20	14	19.5	0.2	70	273
	70		1		1255

a) El recorrido mínimo que registran el 9% de los autos más económicos es 15 Kilómetros por galón.

b) El 20% de los autos recorren entre 19 y 20 Kilómetros por galón.

c)

• **Mediana**

Amplitud= 16 – 15 = 1

$\frac{n}{2} = \frac{70}{2} = 35$ Por lo tanto nos fijamos en el cuarto intervalo, Luego:

$$Me = 18 + \frac{35 + 32}{24} \cdot 1 = 20.79$$

El 50% de los autos recorre a lo mas 20.79 Kilómetros por galón.

• **Moda**

$$Mo = 18 + \frac{14}{14 + 16} \cdot 1 = 18.47$$

La mayor cantidad de autos recorre 18.47 Kilómetros por galón.

• **Media Aritmética**

$$\bar{X} = \frac{1255}{70} = 17.93$$

El promedio de kilómetros que recorren los automóviles es de 17.93 Kilómetros por galón.

Actividad 3:

a)

Tabla de Distribución De Concentración De Cloro En Un Galón De Agua

$LI_i - LS_i$	x_i	n_i	$x_i \cdot n_i$	N_i
15,15 – 15,51	15,33	2	30,66	2
15,51 – 15,87	15,69	10	156,9	12
15,87 – 16,23	16,05	14	224,7	26
16,23 – 16,59	16,41	7	114,87	33
16,59 – 16,95	16,77	6	100,62	39
Totales		39	627,75	

b)

La concentración de cloro con mayor frecuencia es 14 y se encuentra entre el intervalo 15.87 y 16.23 de concentración de cloro.

c)

- **Media Aritmética**

$$\bar{X} = \frac{627.75}{39} = 16.096 \quad \text{Por lo tanto la concentración de cloro promedio}$$

correspondiente a los 39 galones es de 16.096 de cloro.

d)

- **Mediana**

$$\text{Amplitud} = 15.51 - 15.15 = 0.36$$

$$\frac{n}{2} = \frac{39}{2} = 19.5 \quad \text{Por lo tanto nos fijamos en el tercer intervalo, Luego:}$$

$$Me = 15.87 + \frac{19.5 - 12}{14} \cdot 0.36 = 16.68$$

El 50% de los galones presenta a lo mas un 16.68 de concentración de cloro.

Actividad 4:

INTERVALOS	Y_i	n_i	f_i	N_i	$Y_i n_i$
1 - 3	2	31	0.31	31	62
3 - 5	4	20	0.2	51	80
5 - 7	6	30	0.3	81	180
7 - 9	8	9	0.09	90	72
9 - 10	10	10	0.1	100	100

a)

$$\frac{n}{2} = \frac{100}{2} = 50 \text{ Por lo tanto nos fijamos en el segundo intervalo,}$$

Luego:

• **Mediana**

$$Me = 3 + \frac{50 - 31}{20} \cdot 2 = 11.1$$

• **Moda**

$$Mo = 1 + \frac{20}{20 + 0} \cdot 2 = 3$$

4.2. Medidas de dispersión

Ejemplo 15

Consideremos los siguientes conjuntos de valores referidos a las edades de los componentes de dos equipos de fútbol.

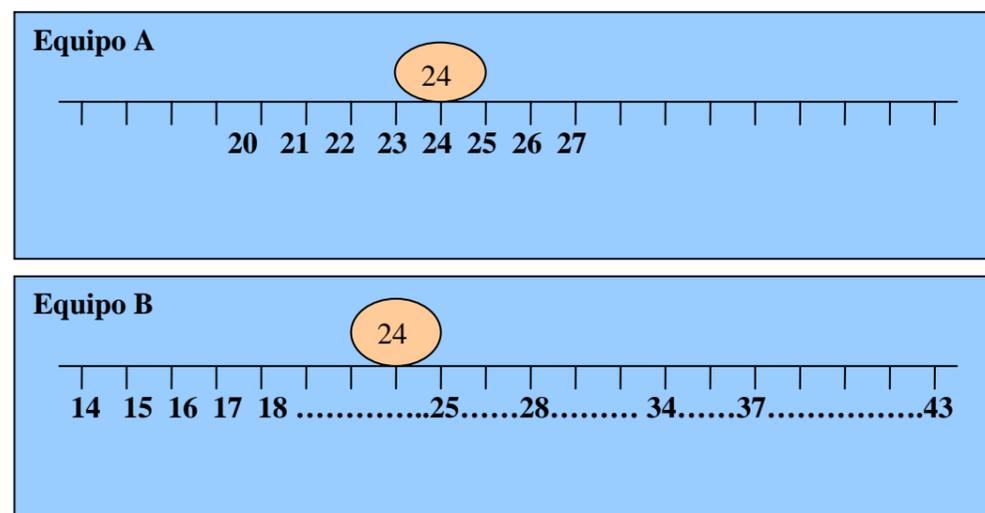
Equipo A: 23 – 27 – 25 – 20 – 26 – 21 – 27 – 23 – 24 – 26 – 22

Equipo B: 34 – 17 – 28 – 18 – 37 – 14 – 15 – 43 – 25 – 16 – 15

La media en ambos casos es 24 años. Sin embargo, si te detienes un momento a observar las edades, te darás cuenta que son equipos con características muy diferentes. En el equipo A todos los jugadores tienen edades próximas a 24 años, en B en cambio las edades fluctúan entre 14 y 43 años. Si nos conformáramos con el valor de la media para informar sobre las edades de los jugadores de cada equipo, la información sería pobre. Necesitaremos algún parámetro estadístico que nos informe cuanto se separan algunos valores de su media. Estos parámetros se llaman **medidas de dispersión**.

En este punto estudiaremos algunos de ellos que nos permite entregar mayor información respecto del problema que se este estudiando.

Observa las siguientes representaciones de las edades



Las edades del equipo A se encuentran más cercanas a la media aritmética, que las edades del equipo B. Es decir, las edades del equipo B se encuentran más **dispersas**.

Las medidas de dispersión nos permitirán realizar un análisis más certero.

4.2.1. Rango o recorrido

Rango (R) es la diferencia entre el mayor y el menor valor que toma el dato que estamos analizando.

En el equipo A el rango es $27 - 20 = 7$ años

En el equipo B el rango es $43 - 14 = 29$ años.

Esto ya nos da más información que conocer solamente la media, sin embargo aunque no es una medida muy significativa, este nos indica cuan dispersos se encuentran los datos entre los valores de los extremos.

$$R = x_{max} - x_{min}$$

4.2.2. Desviación media

Siguiendo con nuestro ejemplo de los equipos de futbol, calcularemos cuanto se desvía de la media cada uno de los valores, que esta dada por $d = x_i - \bar{x}$. Sabemos que la media es 24 años en cada equipo.

Equipo A

x_i	$d = x_i - \bar{x}$	$ x_i - \bar{x} $
20	$20 - 24 = -4$	4
21	$21 - 24 = -3$	3
22	$22 - 24 = -2$	2
23	$23 - 24 = -1$	1
23	$23 - 24 = -1$	1
24	$24 - 24 = 0$	0
25	$25 - 24 = 1$	1
26	$26 - 24 = 2$	2
26	$26 - 24 = 2$	2
27	$27 - 24 = 3$	3
27	$27 - 24 = 3$	3
	$\sum d = 0$	$\sum = 22$

Equipo B

x_i	$d = x_i - \bar{x}$	$ x_i - \bar{x} $
14	$14 - 24 = -10$	10
15	$15 - 24 = -9$	9
15	$15 - 24 = -9$	9
16	$16 - 24 = -8$	8
17	$17 - 24 = -7$	7
18	$18 - 24 = -6$	6
27	$27 - 24 = 3$	3
28	$28 - 24 = 4$	4
34	$34 - 24 = 10$	10
37	$37 - 24 = 13$	13
43	$43 - 24 = 19$	19
	$\sum d = 0$	$\sum = 98$

Observamos que la suma de las desviaciones d es cero, pues las positivas se compensan con las negativas. Son las desviaciones absolutas (valor absoluto de las desviaciones) las que nos interesan.

Llamaremos desviación media a la media (promedio aritmético) de las desviaciones absolutas.

$$D.M. = \frac{\sum |x_i - \bar{x}|}{n}$$

Equipo A

$$D.M. = \frac{22}{11} = 2$$

Equipo B

$$D.M. = \frac{98}{11} = 8,9$$

Para **datos agrupados**, la desviación media es

$$D.M. = \frac{\sum_{i=1}^k |x_i - \bar{x}| n_i}{n}$$

Con esto nos podemos dar cuenta que la dispersión en el equipo B es mucho mayor que en el equipo A.

Ejemplo 16

Los vendedores de una empresa comercial fueron clasificados de acuerdo al volumen de ventas en miles de \$.

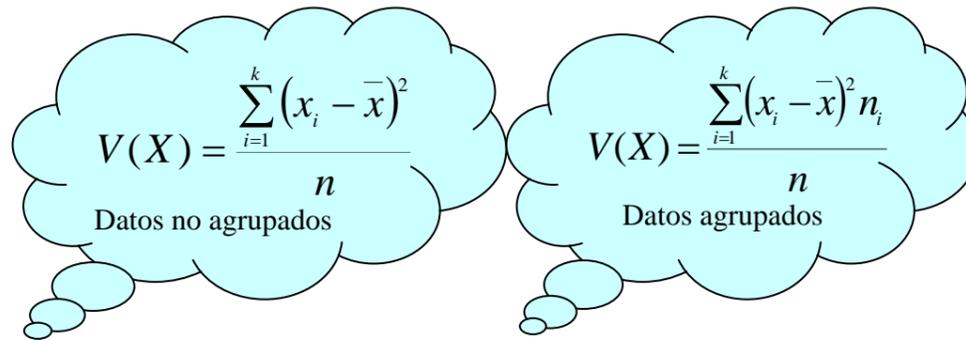
Volumen de ventas	Número de vendedores	x_i	$x_i \cdot n_i$	$ x_i - \bar{x} $	$ x_i - \bar{x} \cdot n_i$
5 – 15	3	10	30	41.2	123.6
15 – 25	24	20	480	31.2	748.8
25 – 45	46	65	2990	13.8	634.8
45 – 75	27	60	1620	8.8	237.6
	100		5120		1744.8

$$\bar{X} = \frac{5120}{100} = 51.2 \qquad D.M. = \frac{1744.8}{100} = 17.448$$

El promedio de las desviaciones absolutas es 17.448

4.2.3. Varianza.

La varianza, denotada por $V(\mathbf{x})$ o $S^2(\mathbf{x})$, es el promedio de los cuadrados, de las desviaciones de los valores de la variable respecto de su media aritmética. Es decir:



Ejemplo 17

Con el objeto de estudiar el retraso de los vuelos de un gran aeropuerto, se toma una muestra cuyos datos, ya agrupados, aparecen en la siguiente tabla:

Minutos de retraso	Número de vuelos
0 – 10	29
10 – 20	23
20 – 30	17
30 – 40	14
40 – 50	11
50 – 60	6

Solución:

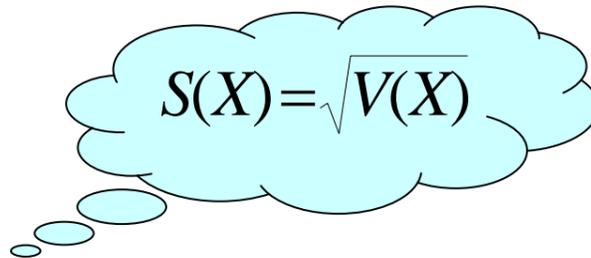
Minutos de retraso	Nº de vuelos (n_i)	x_i	$x_i n_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n_i$
0 – 10	29	5	145	299.29	8679.41
10 – 20	23	15	345	53.29	1225.67
20 – 30	17	25	425	7.29	123.93
30 – 40	14	35	490	161.29	2258.06
40 – 50	11	45	495	515.29	5668.19
50 – 60	6	55	330	1069.29	6415.74
	100		2230		22371

$$\bar{X} = \frac{2230}{100} = 22.3 \rightarrow \text{Minutos de retraso}$$

$$V(X) = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 n_i}{n} = \frac{22371}{100} = 223.71 \rightarrow \text{Minuto de retraso}$$

4.2.4. Desviación estándar o típica

Desviación estándar o típica, expresa el grado de dispersión de los datos con respecto al promedio \bar{x} . Se designa con la letra S y se calculará de la siguiente forma


$$S(X) = \sqrt{V(X)}$$

Entonces la desviación estándar para el ejercicio anterior, es

$$S(X) = \sqrt{223.71} = 14.956$$

Observación:

Mientras mayor sea el valor de la desviación estándar, el grupo de observaciones es más “homogéneo” que si el valor de la desviación estándar fuera más grande. O sea, a menor dispersión mayor homogeneidad y a mayor dispersión, menor homogeneidad.

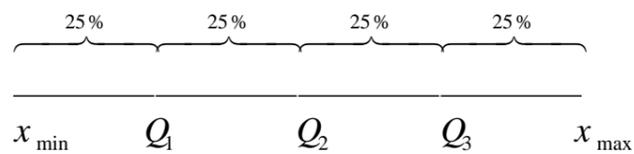
Homogéneo: Dicho de un conjunto: Formado por elementos iguales.

4.3. Medidas de localización: Cuartiles, Percentiles y Deciles.

Como extensión de esta medida de tendencia central, será fácil ampliar el concepto a otras medidas que dividan la masa de información en otras proporciones, y no sólo en mitades.

4.3.1. Cuartiles

Los cuartiles (Q_1, Q_2, Q_3) son 3 valores que dividen a la distribución en cuatro partes de igual número de datos.



Q_1 es el valor de la variable que es superado por no más del 25% de las observaciones y es superado por no más del 75% de los datos restantes.

Q_2 es la mediana

Q_3 es el valor de la variable que supera a no más del 75% de los datos y es superado por no más del 25% restante de las observaciones.

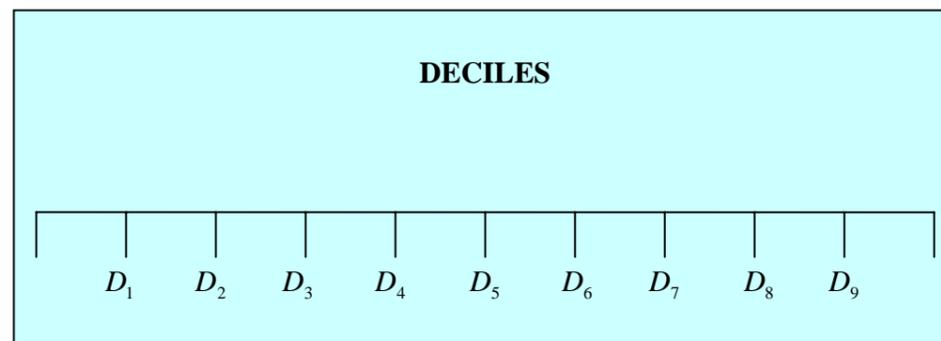
De esta manera los cuartiles, para datos agrupados, se calculan mediante las siguientes fórmulas:

$$Q_1 = X'_{j-1} + \frac{\frac{n}{4} - N_{j-1}}{n_j} \cdot a_j$$

$$Q_3 = X'_{j-1} + \frac{\frac{3n}{4} - N_{j-1}}{n_j} \cdot a_j$$

4.3.2. Deciles

Deciles (D_1, D_2, \dots, D_9) son 9 valores que dividen a la población completa en 10 partes iguales.



Para calcular los deciles, se hace mediante los percentiles, donde el decil 1 equivale al percentil 10, el 2 equivale al percentil 20, y así sucesivamente.

4.3.3. Percentil

Los percentiles dividen a la distribución en unos por cientos. Por lo tanto el primer cuartil es el percentil veinticinco, la mediana es el percentil cincuenta y el tercer cuartil es el percentil setenta y cinco.

P_i es el **i-ésimo percentil** con $i = 1, 2, 3, \dots, 99$ y se define como el valor de la variable que supera a no más de $i\%$ de las observaciones y es superado por no más del $(100 - i)\%$ restante. Se calculan por la fórmula:

$$P_i = x'_{j-1} + \frac{\frac{in}{100} - N_{j-1}}{n_j} \cdot a_j$$

Ejemplo 18

Considerando la tabla de distribución de frecuencias de la inversión de la industria pesquera, se tiene:

INVERSION En millones De dólares	n_i	N_i
0 - 8	8	8
8 - 16	16	24
16 - 24	18	42
24 - 32	12	54
32 - 40	6	60
	60	

Con los datos de la inversión de la industria pesquera, calcule:

- i) P_{20} , Q_3 , P_{80} .
- ii) la mayor inversión del 12% de las empresas que menos invierten.
- iii) la menor inversión del 8% de las empresas que más invierten

Solución.

i) (a) Cálculo de P_{20}

$$\frac{20n}{100} = \frac{20 \cdot 60}{100} = 12 \Rightarrow P_{20} \text{ pertenece al segundo intervalo}$$

$$P_{20} = 8 + \frac{12 - 8}{16} \cdot 8 = 10 \text{ millones de dólares}$$

El 20% de las empresas invierten a lo más 10 millones de dólares

b) Cálculo de Q_3

$$\frac{75n}{100} = \frac{75 \cdot 60}{100} = 45 \Rightarrow Q_3 \text{ pertenece al cuarto intervalo}$$

$$Q_3 = 24 + \frac{45 - 42}{12} \cdot 8 = 26 \text{ millones de dólares}$$

El 75% de las empresas invierten a lo más 26 millones de dólares.

c) Cálculo de P_{80}

$$\frac{80n}{100} = \frac{80 \cdot 60}{100} = 48 \Rightarrow P_{80} \text{ pertenece al cuarto intervalo}$$

$$P_{80} = 24 + \frac{48 - 42}{12} \cdot 8 = 28 \text{ millones de dólares}$$

El 80% de las empresas invierten a lo más 28 millones de dólares.

ii) En este caso se debe calcular el percentil 12

$$\frac{12n}{100} = \frac{12 \cdot 60}{100} = 7,2 \Rightarrow P_{12} \text{ pertenece al primer intervalo}$$

$$P_{12} = 0 + \frac{7,2 - 0}{8} \cdot 8 = 7,2 \text{ millones de dólares}$$

El 12% de las empresas invierte a lo más 7,2 millones de dólares.

iii) Aquí se debe calcular el percentil 92

$$\frac{92n}{100} = \frac{90 \cdot 60}{100} = 55,2 \Rightarrow P_{92} \text{ pertenece al quinto intervalo}$$

$$P_{92} = 32 + \frac{55,2 - 54}{6} \cdot 8 = 33,6 \text{ millones de dólares}$$

El 92% de las empresas invierte a lo menos 33,6 millones de dólares.

4.4. Diagrama de cajas

El diagrama de cajas consiste en un gráfico que muestra simultáneamente diferentes elementos de la distribución de una o más muestras por ejemplo, mediana, rango, cuartiles, deciles, etc.

La principal utilidad de este tipo de gráficos es la comparación de distribuciones de diferentes grupos.

Ejemplo 19

Los siguientes datos corresponden a la masa (en Kg.) de 24 mujeres de 17 años.

Peso	x_i	n_i	N_i
44 – 47	45,5	1	1
47 – 50	48,5	4	5
50 – 53	51,5	5	10
53 – 56	54,5	7	17
56 – 59	57,5	5	22
59 – 62	60,5	2	24
		24	

La Mediana

$$\frac{n}{2} = \frac{24}{2} = 12, \text{ pertenece al cuarto intervalo.}$$

$$Me = 53 + \frac{12 - 10}{7} \cdot 3 = 54$$

$$\text{Rango} = 61 - 44 = 17$$

Q_1

$$\frac{n}{4} = \frac{24}{4} = 6, \text{ corresponde al tercer intervalo}$$

$$Q_1 = 50 + \frac{6-5}{5} \cdot 3 = 50,6 = 51$$

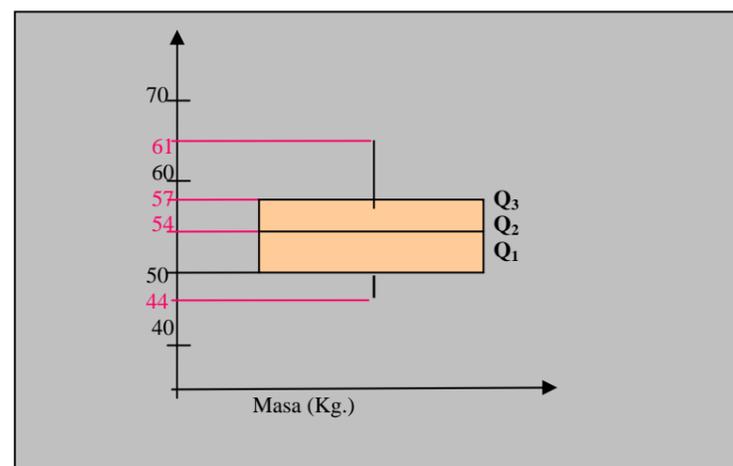
Q_3

$$\frac{3n}{4} = \frac{3 \cdot 24}{4} = 18, \text{ pertenece al quinto intervalo}$$

$$Q_3 = 56 + \frac{18-17}{5} \cdot 3 = 56,6 = 57$$

Diagrama de cajas

Visualizaremos todos los elementos anteriores mediante el siguiente diagrama de caja.



Observa que en el gráfico, los extremos del rectángulo indican los cuartiles Q_1 y Q_3 , mientras que la línea que divide a este horizontalmente indica la mediana.

Ejemplo Excel 8

Las siguientes tablas corresponden a las distribuciones de notas de dos terceros medios en una misma prueba de Matemática. Cada curso está compuesto por treinta alumnos:

Tercer año A

5.3	3.8	4.0	5.6	5.4
4.5	5.3	4.7	5.2	5.1
5.7	6.4	4.9	5.2	7.0
6.5	4.6	4.8	5.2	6.2
6.5	4.9	5.0	6.6	7.0
6.9	4.1	5.2	5.4	5.0

Tercer año B

5.6	5.5	5.5	6.7	5.8
5.1	4.1	4.9	7.0	4.3
7.0	5.1	5.9	4.4	6.5
4.6	5.2	5.2	3.3	7.0
6.8	5.9	6.0	5.4	4.2
5.5	4.6	7.0	4.2	3.7

De acuerdo a la información anterior, ¿podemos saber qué curso tuvo un desempeño más homogéneo en el test de matemática?

Solución: Para contestar la pregunta anterior, podemos hacer uso de la planilla Excel, ingresamos los datos a la planilla, obteniendo como resultado, lo siguiente:

	A	B	C	D	E	F
1	Tercero A					
2						
3	5.3	3.8	4	5.6	5.4	
4	4.5	5.3	4.7	5.2	5.1	
5	5.7	6.4	4.9	5.2	7	
6	6.5	4.6	4.8	5.2	6.2	
7	6.5	4.9	5	6.6	7	
8	6.9	4.1	5.2	5.4	5	
9						
10						

Aquí podemos de inmediato determinar la media aritmética. Para ello, podemos escribir en la celda **A10** el texto "Media Aritmética" y a continuación, situarnos en la celda **C10**. Luego hacemos clic en el botón . Ahí, en **Insertar Función**, escogemos **ESTADÍSTICAS**, y en **Seleccionar una Función**, seleccionamos **PROMEDIO**.

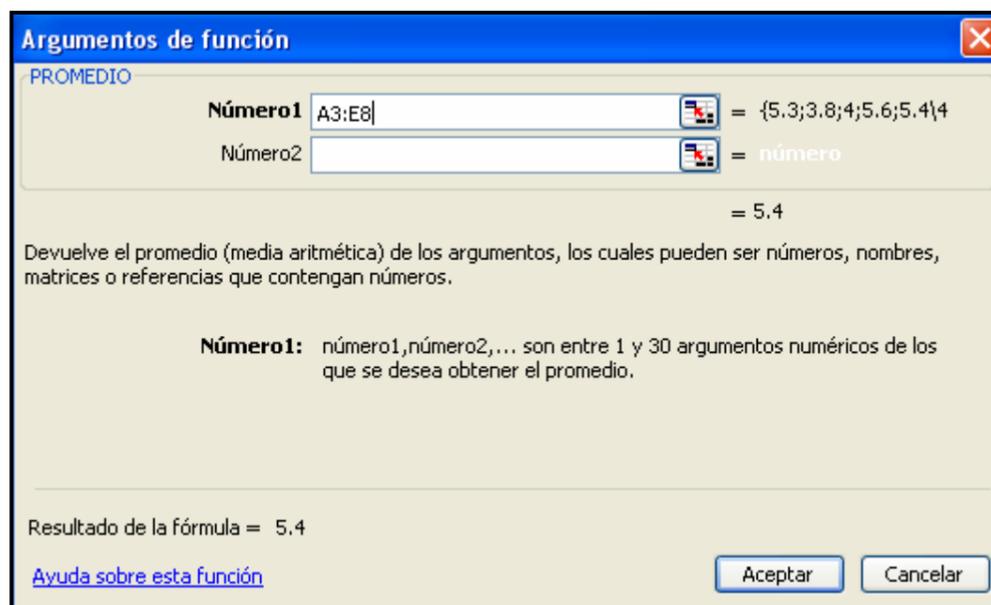
	A	B	C	D	E	F	G	H
1	Tercero A							
2								
3	5.3	3.8	4	5.6	5.4			
4	4.5	5.3	4.7					
5	5.7	6.4	4.9					
6	6.5	4.6	4.8					
7	6.5	4.9	5					
8	6.9	4.1	5.2					
9								
10	Media aritmética		=					
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								

A continuación, hacemos clic en **ACEPTAR**. Después de esto, aparecerá el cuadro de diálogo **Argumentos de función - PROMEDIO**. En **Número1**, seleccionamos las 30 notas del curso en cuestión (En este caso, desde **A3** hasta **E8**). Observar el siguiente cuadro:

Una vez desarrollado lo anterior, hacemos clic en **ACEPTAR**. El resultado se muestra en la siguiente figura:

	A	B	C	D	E
1	Tercero A				
2					
3	5.3	3.8	4	5.6	5.4
4	4.5	5.3	4.7	5.2	5.1
5	5.7	6.4	4.9	5.2	7
6	6.5	4.6	4.8	5.2	6.2
7	6.5	4.9	5	6.6	7
8	6.9	4.1	5.2	5.4	5
9					
10	Media aritmética		5.4		
11					

Si deseamos que todas las notas aparezcan con un decimal (incluso las notas enteras) seleccionamos el menú **Formato – Celdas – Número – Posiciones decimales:1 – Aceptar**. El resultado final será el siguiente:



	A	B	C	D	E
1	Tercero A				
2					
3	5.3	3.8	4.0	5.6	5.4
4	4.5	5.3	4.7	5.2	5.1
5	5.7	6.4	4.9	5.2	7.0
6	6.5	4.6	4.8	5.2	6.2
7	6.5	4.9	5.0	6.6	7.0
8	6.9	4.1	5.2	5.4	5.0
9					
10	Media aritmética		5.4		
11					

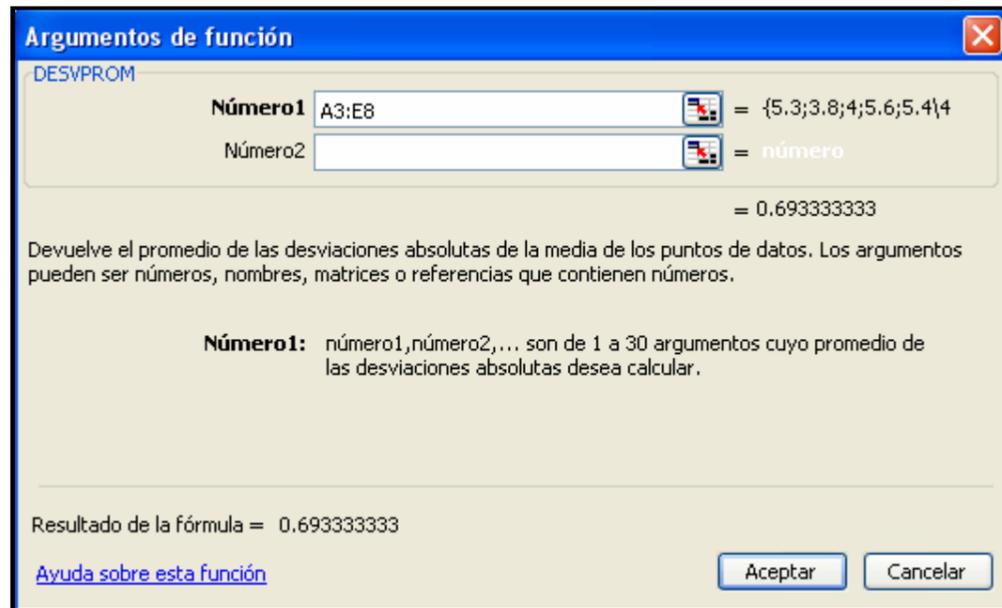
Ahora, en la celda **A11**, escribimos el texto "Desviación Media". Ubicamos el cursor en la celda **C11**, hacemos clic en el menú **Insertar función –Seleccionar una categoría: Estadísticas –Seleccionar una función: DESVPROM – ACEPTAR**.

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E
1	Tercero A				
2					
3	5.3	3.8	4.0	5.6	5.4
4	4.5	5.3	4.7	5.2	5.1
5	5.7	6.4	4.9	5.2	7.0
6	6.5	4.6	4.8	5.2	6.2
7	6.5	4.9	5.0	6.6	7.0
8	6.9	4.1	5.2	5.4	5.0
9					
10	Media aritmética		5.4		
11	Desviación Media	=			

The 'Insertar función' dialog box is open, showing the 'Estadísticas' category selected. The function list includes 'DESV PROM', which is highlighted. The dialog also shows the syntax 'DESV PROM(número1,número2,...)' and a description: 'Devuelve el promedio de las desviaciones absolutas de la media de los puntos de datos. Los argumentos pueden ser números, nombres, matrices o referencias que contienen números.'

Una vez hecho lo anterior, aparecerá el cuadro de diálogo **Argumentos de función -DESV PROM**. En **Número1**, seleccionamos el rango **A3:E8** que contiene a los datos numéricos relativos a las calificaciones del Tercer año A (Ver siguiente cuadro).



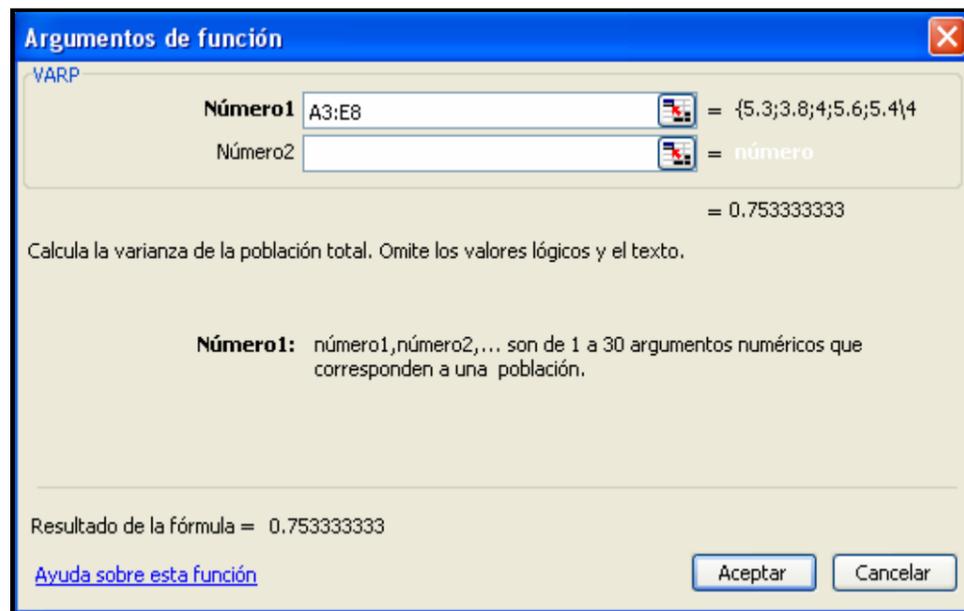
Al hacer clic en **ACEPTAR**, aparece calculado el valor de la desviación estándar en la celda **C11**

	A	B	C	D	E
1	Tercero A				
2					
3	5.3	3.8	4.0	5.6	5.4
4	4.5	5.3	4.7	5.2	5.1
5	5.7	6.4	4.9	5.2	7.0
6	6.5	4.6	4.8	5.2	6.2
7	6.5	4.9	5.0	6.6	7.0
8	6.9	4.1	5.2	5.4	5.0
9					
10	Media aritmética		5.4		
11	Desviación Media		0.693333333		
12					

Ahora, en la celda **A12**, escribimos el texto "Varianza". Ubicamos el cursor en la celda **C12**, hacemos clic en el menú **Insertar función –Seleccionar una categoría: Estadísticas –Seleccionar una función: VARP – ACEPTAR**.



Una vez hecho lo anterior, aparecerá el cuadro de diálogo **Argumentos de función -VARP**. En **Número1**, seleccionamos el rango **A3:E8** que contiene a los datos numéricos relativos a las calificaciones del Tercer año A (Ver siguiente cuadro).



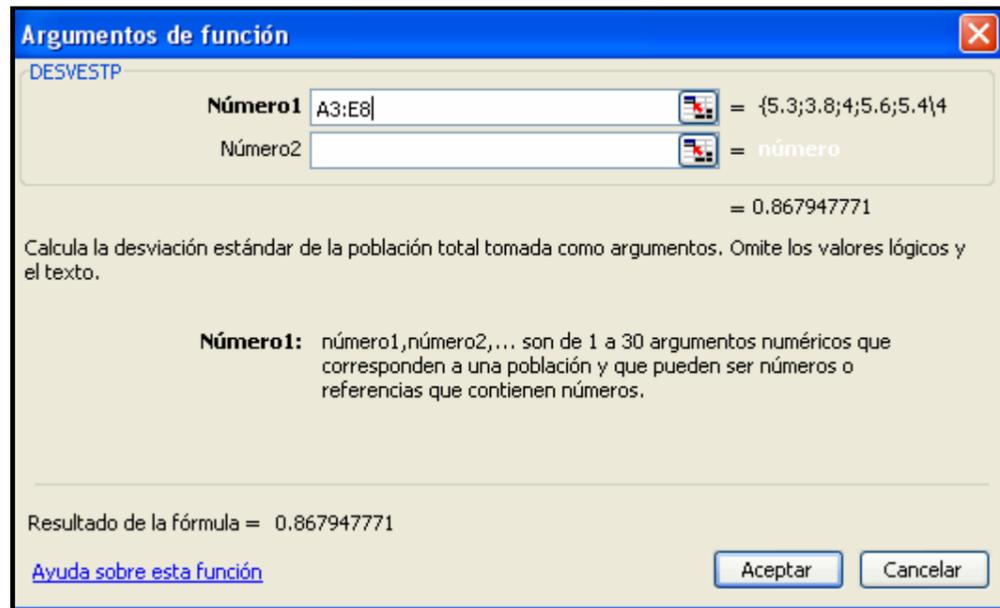
Al hacer clic en **ACEPTAR**, aparece calculado el valor de la varianza en la celda **C12**

	A	B	C	D	E	F
1	Tercero A					
2						
3	5.3	3.8	4.0	5.6	5.4	
4	4.5	5.3	4.7	5.2	5.1	
5	5.7	6.4	4.9	5.2	7.0	
6	6.5	4.6	4.8	5.2	6.2	
7	6.5	4.9	5.0	6.6	7.0	
8	6.9	4.1	5.2	5.4	5.0	
9						
10	Media aritmética		5.4			
11	Desviación Media		0.69333333			
12	Varianza		0.75333333			
13						

Ahora, en la celda **A13**, escribimos el texto “Desviación Estándar”. Ubicamos el cursor en la celda **C13**, hacemos clic en el menú **Insertar función –seleccionar una categoría: Estadísticas –Seleccionar una función: DESVESTP – ACEPTAR**.

	A	B	C	D	E	F	G	H	I	J	
1	Tercero A										
2											
3	5.3	3.8	4.0	5.6	5.4						
4	4.5	5.3	4.7	5.2	5.1						
5	5.7	6.4	4.9	5.2	7.0						
6	6.5	4.6	4.8	5.2	6.2						
7	6.5	4.9	5.0	6.6	7.0						
8	6.9	4.1	5.2	5.4	5.0						
9											
10	Media aritmética		5.4								
11	Desviación Media		0.69333333								
12	Varianza		0.75333333								
13	Desviación estándar										
14											
15											
16											
17											
18											

Una vez hecho lo anterior, aparecerá el cuadro de diálogo **Argumentos de función –DESVESTP**. En **Número1**, seleccionamos el rango **A3:E8** que contiene a los datos numéricos relativos a las calificaciones del Tercer año A (Ver siguiente cuadro).



Al hacer clic en **ACEPTAR**, aparece calculado el valor de la desviación estándar en la celda **C13**.

	A	B	C	D	E
1	Tercero A				
2					
3	5.3	3.8	4.0	5.6	5.4
4	4.5	5.3	4.7	5.2	5.1
5	5.7	6.4	4.9	5.2	7.0
6	6.5	4.6	4.8	5.2	6.2
7	6.5	4.9	5.0	6.6	7.0
8	6.9	4.1	5.2	5.4	5.0
9					
10	Media aritmética		5.4		
11	Desviación Media		0.69333333		
12	Varianza		0.75333333		
13	Desviación estándar		0.86794777		
14					

Si efectuamos los mismos pasos anteriores con las notas del Tercer año B, obtendremos el siguiente resultado:

17	Tercero B				
18					
19	5.6	5.5	5.5	6.7	5.8
20	5.1	4.1	4.9	7	4.3
21	7	5.1	5.9	4.4	6.5
22	4.6	5.2	5.2	3.3	7
23	6.8	5.9	6	5.4	4.2
24	5.5	4.6	7	4.2	3.7
25					
26	Media aritmética		5.4		
27	Desviación Media		0.84666667		
28	Varianza		1.06866667		
29	Desviación estándar		1.03376335		
30					

¿Qué conclusión es posible formular en base a los cálculos anteriores? Notemos que ambos cursos presentan la misma media aritmética, lo que nos llevaría a concluir que ambos cursos tuvieron igual desempeño en el test. Sin embargo, debemos tener en cuenta no sólo la tendencia central de un conjunto de observaciones: también debemos considerar la dispersión de dichas distribuciones. El **Tercer año A** posee una desviación estándar igual a 0,867947771, mientras que el **Tercer año B** presenta una desviación estándar igual a 1,033763351. Como vemos, el Tercer año B presenta una mayor dispersión promedio que el Tercer año A, lo que nos llevaría a concluir que en general el Tercer año A tuvo un desempeño más homogéneo en el test que el Tercer año B.

Para tener una mayor certeza de nuestra conclusión anterior, podemos determinar otro tipo de medidas, que llevan el nombre de medidas de localización. Veamos como determinar el cuartil 1, cuartil 2, y cuartil 3 para el tercer año A, recordemos que el cuartil 2 corresponde a la mediana, ubicamos el curso en la celda **A14** y escribimos "Cuartil 1". Luego, situamos el cursor en la celda **C14**, y hacemos clic en **Insertar Función –Seleccionar una categoría: Estadísticas –Seleccionar una Función: Cuartil** (Ver figura).

	A	B	C	D	E
1	Tercero A				
2					
3	5.3	3.8	4.0	5.6	5.4
4	4.5	5.3	4.7	5.2	5.1
5	5.7	6.4	4.9	5.2	7.0
6	6.5	4.6	4.8	5.2	6.2
7	6.5	4.9	5.0	6.6	7.0
8	6.9	4.1	5.2	5.4	5.0
9					
10	Media aritmética		5.4		
11	Desviación Media		0.69333333		
12	Varianza		0.75333333		
13	Desviación estándar		0.86794777		
14	Cuartil 1		=		
15					
16					
17					
18					

Insertar función

Buscar una función:

Escriba una breve descripción de lo que desea hacer y, a continuación, haga clic en Ir

O seleccionar una categoría: Estadísticas

Seleccionar una función:

CONTAR.SI
CONTARA
COVAR
CRECIMIENTO
CUARTIL
CURTOSIS
DESVEST

CUARTIL(matriz, cuartil)
Devuelve el cuartil de un conjunto de datos.

[Ayuda sobre esta función](#) Aceptar Cancelar

Al hacer clic en **ACEPTAR**, aparecerá el cuadro de diálogo **Argumentos de función- Cuartil**. En **MATRIZ**, ingresamos el rango que contiene a las calificaciones del Tercer año A, y en el valor **cuartil**, se escribe el valor 1 (que corresponde al primer cuartil).

Argumentos de función

CUARTIL

Matriz A3:E8 = {5.3;3.8;4;5.6;5.4\4

Cuartil 1 = 1

= 4.9

Devuelve el cuartil de un conjunto de datos.

Cuartil es un número: valor mínimo = 0; primer cuartil = 1; valor de la mediana = 2; tercer cuartil = 3; valor máximo = 4.

Resultado de la fórmula = 4.9

[Ayuda sobre esta función](#) Aceptar Cancelar

Una vez hecho esto, hacer clic en **ACEPTAR**. Aparecerá en la celda **C14** el primer cuartil, igual a 4.9.

	A	B	C	D	E
1	Tercero A				
2					
3	5.3	3.8	4.0	5.6	5.4
4	4.5	5.3	4.7	5.2	5.1
5	5.7	6.4	4.9	5.2	7.0
6	6.5	4.6	4.8	5.2	6.2
7	6.5	4.9	5.0	6.6	7.0
8	6.9	4.1	5.2	5.4	5.0
9					
10	Media aritmética		5.4		
11	Desviación Media		0.69333333		
12	Varianza		0.75333333		
13	Desviación estándar		0.86794777		
14	Cuartil 1		4.9		
15					

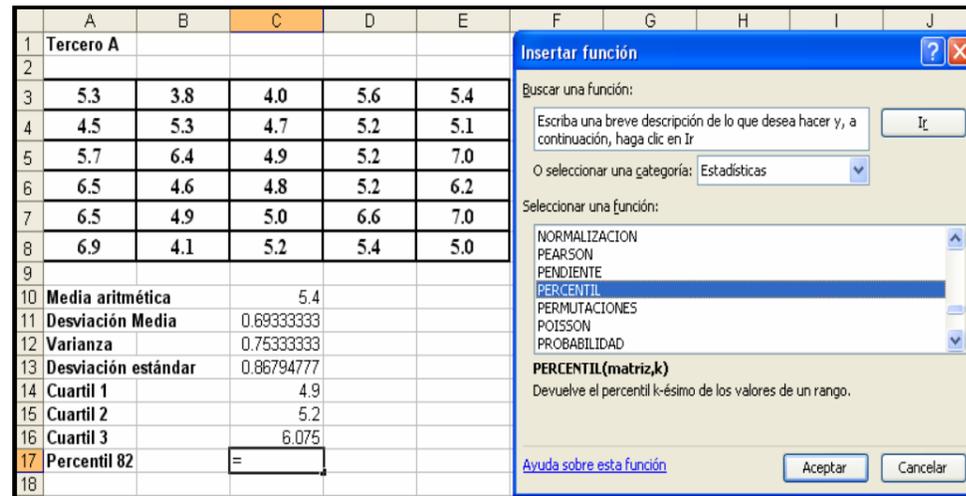
¿Qué significa que el valor **Q1** sea igual a 4.9? Significa que aproximadamente el 25% de los estudiantes tuvieron notas inferiores a 4.9.

Para calcular el cuartil 2 y 3 se realizan los mismos pasos anteriores, la única diferencia es al ingresar el valor del **cuartil**, en el cuartil 2 será **2** y en el caso del cuartil 3 será **3**

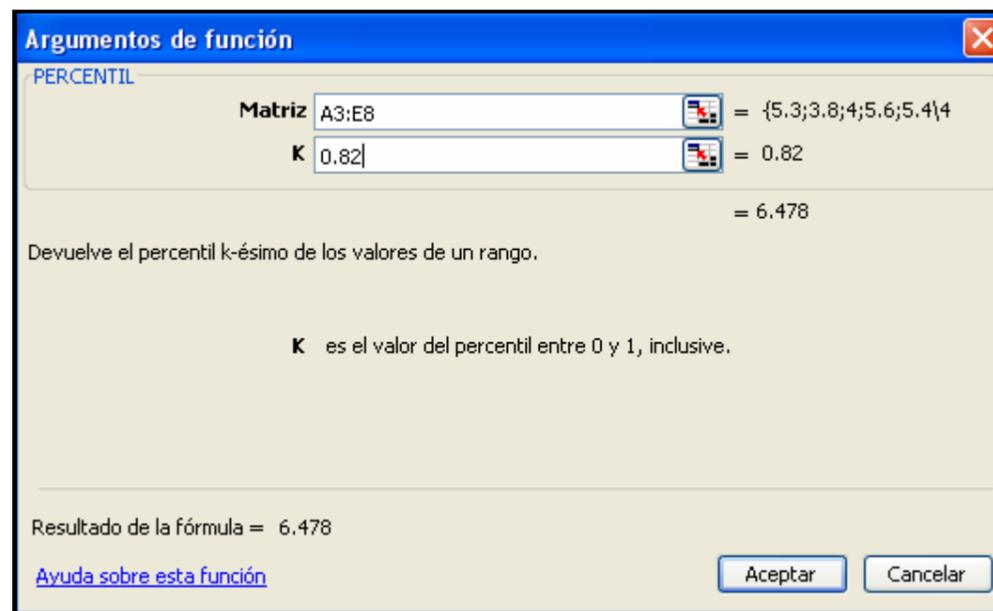
Los resultados se muestran en la siguiente figura:

	A	B	C	D	E
1	Tercero A				
2					
3	5.3	3.8	4.0	5.6	5.4
4	4.5	5.3	4.7	5.2	5.1
5	5.7	6.4	4.9	5.2	7.0
6	6.5	4.6	4.8	5.2	6.2
7	6.5	4.9	5.0	6.6	7.0
8	6.9	4.1	5.2	5.4	5.0
9					
10	Media aritmética		5.4		
11	Desviación Media		0.69333333		
12	Varianza		0.75333333		
13	Desviación estándar		0.86794777		
14	Cuartil 1		4.9		
15	Cuartil 2		5.2		
16	Cuartil 3		6.075		
17					

Ahora calcularemos el percentil 82, en la celda **A17** escribimos "Percentil 82". Luego, situamos el cursor en la celda **C15**, y hacemos clic en **Insertar Función – seleccionar una categoría: Estadísticas – Seleccionar una Función: Percentil** (Ver figura).



Al hacer clic en **ACEPTAR**, aparecerá el cuadro de diálogo **Argumentos de función-Percentil**. En **MATRIZ**, ingresamos el rango que contiene a las calificaciones del Tercer año A, y en el valor **k**, se escribe el valor 0.82 (Se ingresan valores entre 0 y 1).



Una vez hecho esto, hacer clic en **ACEPTAR**. Aparecerá en la celda **C17** el percentil 82, igual a 6.478.

	A	B	C	D	E
1	Tercero A				
2					
3	5.3	3.8	4.0	5.6	5.4
4	4.5	5.3	4.7	5.2	5.1
5	5.7	6.4	4.9	5.2	7.0
6	6.5	4.6	4.8	5.2	6.2
7	6.5	4.9	5.0	6.6	7.0
8	6.9	4.1	5.2	5.4	5.0
9					
10	Media aritmética		5.4		
11	Desviación Media		0.69333333		
12	Varianza		0.75333333		
13	Desviación estándar		0.86794777		
14	Cuartil 1		4.9		
15	Cuartil 2		5.2		
16	Cuartil 3		6.075		
17	Percentil 82		6.478		

¿Qué significa que el valor Percentil 82 sea igual a 6.478? Significa que aproximadamente el 82% de los estudiantes tuvieron notas inferiores a 6.478.

Al seguir los mismos pasos anteriores, determinamos el cuartil 1, 2 y 3 y el percentil 82, para el Tercer año B. Se obtiene lo siguiente:

24	Tercero B				
25					
26	5.6	5.5	5.5	6.7	5.8
27	5.1	4.1	4.9	7	4.3
28	7	5.1	5.9	4.4	6.5
29	4.6	5.2	5.2	3.3	7
30	6.8	5.9	6	5.4	4.2
31	5.5	4.6	7	4.2	3.7
32					
33	Media aritmética		5.4		
34	Desviación Media		0.84666667		
35	Varianza		1.06866667		
36	Desviación estándar		1.03376335		
37	Cuartil 1		4.6		
38	Cuartil 2		5.45		
39	Cuartil 3		5.975		
40	Percentil 82		6.656		

Ejemplo Excel 9

La siguiente tabla corresponde a la distribución de las ventas diarias (en miles de pesos) de un grupo de empleados de una tienda:

Ventas	Frecuencia
200 – 300	6
300 – 400	15
400 – 500	13
500 – 600	9

Una vez cargado el programa Microsoft Excel, se procede a ingresar los datos a la planilla. En una hoja, ingresamos los valores correspondientes a los límites inferior y superior de los intervalos (idealmente, en columnas distintas. En este ejemplo, en las columnas **A** y **B**), y sus correspondientes frecuencias absolutas, en otra columna (en este caso columna **D**):

	A	B	C	D	E	F	G	H
1								
2	Limite	Limite	Marca de	Frecuencia	$x_i \cdot n_i$	$(x_i - m.a.)^2 \cdot n_i$	$[x_i - m.a] \cdot n_i$	Frec. Abs.
3	Inferior	Superior	clase (x_i)	absoluta (n_i)				Acumulada
4	200	300		6				
5	300	400		15				
6	400	500		13				
7	500	600		9				
8				43				
9								

Ahora usamos la función **PROMEDIO** para determinar la marca de clase, para cada intervalo.

Para calcular los valores de la columna $x_i \cdot n_i$, introducimos en **E4** la fórmula $=C4 \cdot D4$, y arrastramos el Mouse desde **E4** hasta **E7**, luego para mayor comodidad, en la celda **E8** podemos calcular la suma de $x_i \cdot n_i$, con la opción **AUTOSUMA**. Dicha suma aparece en negrita. Con esto podemos calcular la media aritmética, en la celda **A10** escribimos el texto "media aritmética", luego situamos el cursor en la celda **C10** y escribimos la formula $=E8/D8$.

Como lo vemos en la figura.

	A	B	C	D	E	F	G	H
1								
2	Limite	Limite	Marca de	Frecuencia	$x_i \cdot n_i$	$(x_i - m.a.)^2 \cdot n_i$	$(x_i - m.a) \cdot n_i$	Frec. Abs.
3	Inferior	Superior	clase (x_i)	absoluta (n_i)				Acumulada
4	200	300	250	6	1500			
5	300	400	350	15	5250			
6	400	500	450	13	5850			
7	500	600	550	9	4950			
8				43	17550			
9								
10	Media aritmética		408.14					
11								

De manera similar, para completar la columna $(x_i - m.a.)^2 \cdot n_i$ (para calcular la varianza y luego la desviación estándar), escribimos en la celda **F4** la fórmula = **((C4- \$C\$10)^2)·D4**, y presionamos **ENTER**. Al arrastrar el Mouse desde **F4** hasta **F7**, queda completa la columna, Luego calculamos la suma con la opción **AUTOSUMA**. (El signo peso es para mantener fija la celda, en este caso corresponde a la celda de la media aritmética).

	A	B	C	D	E	F	G	H
1								
2	Limite	Limite	Marca de	Frecuencia	$x_i \cdot n_i$	$(x_i - m.a.)^2 \cdot n_i$	$(x_i - m.a) \cdot n_i$	Frec. Abs.
3	Inferior	Superior	clase (x_i)	absoluta (n_i)				Acumulada
4	200	300	250	6	1500	150049		
5	300	400	350	15	5250	50703		
6	400	500	450	13	5850	22780		
7	500	600	550	9	4950	181120		
8				43	17550	404651		
9								
10	Media aritmética		408.14					
11								

Ahora en la celda **A11** escribimos "varianza" y en **C11** escribimos la formula: **=F8/D8**, como se ve en la siguiente figura

	A	B	C	D	E	F	G	H
1								
2	Limite	Limite	Marca de	Frecuencia	$x_i \cdot n_i$	$(x_i - m.a.)^2 \cdot n_i$	$(x_i - m.a) \cdot n_i$	Frec. Abs.
3	Inferior	Superior	clase (x_i)	absoluta (n_i)				Acumulada
4	200	300	250	6	1500	150049		
5	300	400	350	15	5250	50703		
6	400	500	450	13	5850	22780		
7	500	600	550	9	4950	181120		
8				43	17550	404651		
9								
10	Media aritmética		408.14					
11	Varianza		9410				$\sum (x_i - \bar{x})^2 \cdot n_i$	
12								

Ahora en la celda A12 escribimos “desviación estándar” y en la celda C12, ingresamos la siguiente formula = RAIZ (C11), dando como resultado

C12		=RAIZ(C11)						
	A	B	C	D	E	F	G	H
1								
2	Limite	Limite	Marca de	Frecuencia	$x_i \cdot n_i$	$(x_i - m.a.)^2 \cdot n_i$	$ x_i - m.a \cdot n_i$	Frec. Abs.
3	Inferior	Superior	clase (xi)	absoluta (ni)				Acumulada
4	200	300	250	6	1500	150049		
5	300	400	350	15	5250	50703		
6	400	500	450	13	5850	22780		
7	500	600	550	9	4950	181120		
8				43	17550	404651		
9								
10	Media aritmética		408.14					
11	Varianza		9410					
12	Desviación estándar		97					
13								

Para calcular la desviación media, debemos completar la columna $[x_i - m.a] \cdot n_i$, entonces en G4 ingresamos la siguiente formula = ABS (C4-\$C\$10)*D4, y presionamos ENTER. Al arrastrar el Mouse desde G4 hasta G7, queda completa la columna. (La función ABS es para obtener el valor absoluto). Luego calculamos la suma con la opción AUTOSUMA.

Luego en A13 escribimos “desviación media” y en C13 ingresamos la formula =G8/D8, dando como resultado lo siguiente

	A	B	C	D	E	F	G	H
1								
2	Limite	Limite	Marca de	Frecuencia	$x_i \cdot n_i$	$(x_i - m.a.)^2 \cdot n_i$	$ x_i - m.a \cdot n_i$	Frec. Abs.
3	Inferior	Superior	clase (xi)	absoluta (ni)				Acumulada
4	200	300	250	6	1500	150049	949	
5	300	400	350	15	5250	50703	872	
6	400	500	450	13	5850	22780	544	
7	500	600	550	9	4950	181120	1277	
8				43	17550	404651	3642	
9								
10	Media aritmética		408.14					
11	Varianza		9410					
12	Desviación estándar		97					
13	Desviación Media		85					
14								
15								

$$\sum |x_i - \bar{x}| \cdot n_i$$

Para calcular los cuartiles y los Percentiles debemos completar la columna de las frecuencias absolutas acumuladas, y en la celda A14 escribimos “Cuartil 1”, luego vemos que la frecuencia acumulada que supera a $43/4 = 10.75$ es 21, entonces el cuartil 1 está en el intervalo (300,400].

Para calcularla, usamos la expresión

$$Q_1 = X'_{j-1} + \frac{\frac{n}{4} - N_{j-1}}{n_j} \cdot a_j$$

En este ejemplo particular, en la celda C14 vemos que la fórmula en Excel será

`=A5+(((D8/4)-H4)/D5)*(B5-A5)`, cuyo valor es

	A	B	C	D	E	F	G	H
1								
2	Limite	Limite	Marca de	Frecuencia	xi*ni	(xi - m.a.) ² *ni	[xi - m.a]*ni	Frec. Abs.
3	Inferior	Superior	clase (xi)	absoluta (ni)				Acumulada
4	200	300	250	6	1500	150049	949	6
5	300	400	350	15	5250	50703	872	21
6	400	500	450	13	5850	22780	544	34
7	500	600	550	9	4950	181120	1277	43
8				43	17550	404651	3642	
9								
10	Media aritmética		408.14					
11	Varianza		9410					
12	Desviación estándar		97					
13	Desviación Media		85					
14	Cuartil 1		332					
15								

Ahora calcularemos el cuartil 3 de la misma forma, para este usamos la expresión

$$Q_3 = X'_{j-1} + \frac{\frac{3n}{4} - N_{j-1}}{n_j} \cdot a_j$$

Escribimos en la celda **A15** "Cuartil 3", luego la frecuencia absoluta acumulada que supera a $3 \cdot 43 / 4 = 32.25$ es 34, entonces el cuartil 3 está en el intervalo (400,500].

Para calcularla, en la celda C15 escribimos la siguiente formula:

`=A6+(((3*D8/4)-H5)/D6)*(500-400)`, cuyo valor es:

	A	B	C	D	E	F	G	H
1								
2	Limite	Limite	Marca de	Frecuencia	$x_i \cdot n_i$	$(x_i - m.a.)^2 \cdot n_i$	$(x_i - m.a) \cdot n_i$	Frec. Abs.
3	Inferior	Superior	clase (x_i)	absoluta (n_i)				Acumulada
4	200	300	250	6	1500	150049	949	6
5	300	400	350	15	5250	50703	872	21
6	400	500	450	13	5850	22780	544	34
7	500	600	550	9	4950	181120	1277	43
8				43	17550	404651	3642	
9								
10	Media aritmética		408.14					
11	Varianza		9410					
12	Desviación estándar		97					
13	Desviación Media		85					
14	Cuartil 1		332					
15	Cuartil 3		487					
16								

Por último calcularemos el Percentil 70, debemos ver que la frecuencia absoluta acumulada que supera a $70 \cdot 43 / 100 = 30.1$ es 34, entonces el percentil 70 está en el intervalo (400,500]. Para calcularlo, usamos la expresión

$$P_i = x'_{j-1} + \frac{\frac{i \cdot n}{100} - N_{j-1}}{n_j} \cdot a_j$$

Entonces en la celda **A16**, escribimos Percentil 70, y en la celda **C16**, escribimos la siguiente formula: `=400+(((70*D8/100)-H5)/D6)*(B6-A6)`

Dando como resultado:

	A	B	C	D	E	F	G	H
1								
2	Limite	Limite	Marca de	Frecuencia	$x_i \cdot n_i$	$(x_i - m.a.)^2 \cdot n_i$	$(x_i - m.a) \cdot n_i$	Frec. Abs.
3	Inferior	Superior	clase (x_i)	absoluta (n_i)				Acumulada
4	200	300	250	6	1500	150049	949	6
5	300	400	350	15	5250	50703	872	21
6	400	500	450	13	5850	22780	544	34
7	500	600	550	9	4950	181120	1277	43
8				43	17550	404651	3642	
9								
10	Media aritmética		408.14					
11	Varianza		9410					
12	Desviación estándar		97					
13	Desviación Media		85					
14	Cuartil 1		332					
15	Cuartil 3		487					
16	Percentil 70		470					
17								

4.5. Propuesta de trabajo Excel

La tabla de frecuencia representa los sueldos de los empleados de la empresa ENCAL LTDA. medidos en miles de pesos:

Intervalos de sueldo	Frecuencia
400 - 420	10
421 - 441	15
442 - 462	20
463 - 483	30
484 - 504	5
505 - 525	10

- Complete la tabla.
- Calcula Media Aritmética.
- Desviación Media.
- Varianza.
- Desviación Estándar.
- Cuartil 1 y 3.
- Percentil 79.

Respuesta Esperada

1								
2	Limite	Limite	Marca de	Frecuencia				Frec. Abs.
3	Inferior	Superior	Clase (xi)	Absoluta (ni)	$xi \cdot ni$	$(xi - m.a)^2 \cdot ni$	$(xi - m.a) \cdot ni$	Acumulada
4	400	420	410	10	4100	25167	501,67	10
5	421	441	431	15	6465	12760	437,50	25
6	442	462	452	20	9040	1334	163,33	45
7	463	483	473	30	14190	4941	385,00	75
8	484	504	494	5	2470	5723	169,17	80
9	505	525	515	10	5150	30067	548,33	90
10				90	41415	79993	2205	
11								
12								
13	Media Aritmetica		460,17					
14	Varianza		888,81					
15	Desviacion Estandar		29,81					
16	Desviacion Media		24,5					
17	Cuartil 1		437,67					
18	Cuartil 3		478					
19	Percentil 79		480,4					

4.6. PONGAMOS A PRUEBA LO APRENDIDO

Actividad 1

En una empresa los salarios han tenido un promedio semanal de \$5.000, con una desviación estándar de 50. ¿Qué sucede con la varianza de los salarios si?:

- a) Se aumentan en \$450
- b) Se aumentan en 10%
- c) ¿Efectúan ambos aumentos simultáneamente?

Actividad 2

En una empresa los sueldos tienen un promedio de \$100.000 con una desviación estándar de \$180.000. ¿Qué sucede con la media y la varianza si?:

- a) Se aumentan todos los sueldos en \$8.500
- b) Se aumentan todos los sueldos en un 15%
- c) ¿Se aumenta todos los sueldos en un 8% más \$4.000?

Actividad 3

Un estudiante obtuvo 84 puntos en el examen final de matemáticas. El curso obtuvo una puntuación media de 76 y una desviación típica de 10. En el examen final de estadística con una media general de 82 puntos y una desviación estándar de 16 para el curso, él obtuvo una puntuación de 90.

¿Qué asignatura presenta mayor desviación estándar?

Actividad 4

La siguiente tabla representa el sueldo de un grupo de personas en U.F.

SUELDOS	n_i
0 - 20	30
20 - 40	20
40 - 60	15
60 - 80	13
80 - 100	12
100 - 120	10

- i) Calcule el mayor sueldo del 12,5% de las personas que menos ganan.
- ii) Calcule el menor sueldo del 18% de las personas que más ganan.
- iii) Calcule la desviación media, la varianza y la desviación estándar.

Actividad 5

La demanda diaria en unidades de un producto durante 30 días de trabajo es:

25 - 28 - 32 - 33 - 35 - 36 - 38 - 42 - 44 - 44 - 47 - 48 - 48
- 49 - 51 - 52 - 53 - 56 - 57 - 58 - 58 - 59 - 61 - 63 - 66 -
67 - 69 - 72 - 76 - 78

- i) Complete la tabla sabiendo que hay siete intervalos de amplitud constante.
- ii) Calcule : P_{35} , Q_1 , Q_3 , Mediana, Desviación Media y Desviación estándar
- iii) Construye un diagrama de caja.

4.7. SOLUCIONARIO

Actividad 1:

Sea X_A : "Salario semanal, en pesos, de un trabajador de una empresa".

$$\bar{X} = 5000 \quad \sigma_X = 50$$

Según cada caso, el nuevo salario semanal del trabajador X_p estará dado por:

a) Para el caso de la varianza se tiene que

$$X_p = X_A + 450 \quad \Rightarrow \quad \sigma_p^2 = \sigma_A^2$$

Con lo que:

$$\sigma_p^2 = \sigma_A^2 = (50)^2 = 2500 \quad \therefore \quad \sigma_p^2 = 2500 \quad [\$^2]$$

b) Para el caso de la varianza se tiene que

$$X_p = (1,1) \cdot X_A \quad \Rightarrow \quad \sigma_p^2 = (1,1)^2 \cdot \sigma_A^2$$

Con lo que:

$$\sigma_p^2 = (1,1)^2 \cdot \sigma_A^2 = (1,1)^2 \cdot (50)^2 = 3025 \quad \therefore \quad \sigma_p^2 = 3025 \quad [\$^2]$$

c) Para el caso de la varianza se tiene que

$$X_p = (1,1) \cdot X_A + 4000 \quad \Rightarrow \quad \sigma_p^2 = (1,1)^2 \cdot \sigma_A^2$$

Con lo que:

$$\sigma_p^2 = (1,1)^2 \cdot \sigma_A^2 = (1,1)^2 \cdot (50)^2 = 3025 \quad \therefore \quad \sigma_p^2 = 3025 \quad [\$^2]$$

Actividad 2:

Sea X_A : "Salario, en pesos, de un trabajador de una empresa".

$$\bar{X} = 100000 \quad \sigma_X = 180000$$

Según cada caso, el nuevo salario del trabajador X_p estará dado por:

a) Para el caso de la media se tiene que

$$X_p = X_A + 8500 \Rightarrow \bar{X}_p = \bar{X}_A + 8500$$

Con lo que:

$$\bar{X}_p = 100000 + 8500 = 108500 \quad \therefore \quad \bar{X}_p = 108500 \quad [\$]$$

Para el caso de la varianza se tiene que

$$X_p = X_A + 8500 \Rightarrow \sigma_p^2 = \sigma_A^2$$

Con lo que:

$$\sigma_p^2 = \sigma_A^2 = (180000)^2 = 32400000000 \quad \therefore \quad \sigma_p^2 = 32400000000 \quad [\$^2]$$

b) Para el caso de la media se tiene que

$$X_p = (1,15) \cdot X_A \Rightarrow \bar{X}_p = (1,15) \cdot \bar{X}_A$$

Con lo que:

$$\bar{X}_p = (1,15) \cdot 100000 = 115000 \quad \therefore \quad \bar{X}_p = 115000 \quad [\$]$$

Para el caso de la varianza se tiene que

$$X_p = (1,15) \cdot X_A \Rightarrow \sigma_p^2 = (1,15)^2 \cdot \sigma_A^2$$

Con lo que:

$$\sigma_p^2 = (1,15)^2 \cdot \sigma_A^2 = (1,15)^2 \cdot (180000)^2 = 42849000000 \\ \therefore \quad \sigma_p^2 = 42849000000 \quad [\$^2]$$

c) Para el caso de la media se tiene que

$$X_p = (1,08) \cdot X_A + 4000 \Rightarrow \bar{X}_p = (1,08) \cdot \bar{X}_A + 4000 \quad \text{con lo que:}$$

$$\bar{X}_p = (1,08) \cdot 100000 + 4000 = 112000 \quad \therefore \quad \bar{X}_p = 112000 \quad [\$]$$

Para el caso de la varianza se tiene que

$$X_p = (1,08) \cdot X_A + 4000 \Rightarrow \sigma_p^2 = (1,08)^2 \cdot \sigma_A^2 \quad \text{con lo que:}$$

$$\sigma_p^2 = (1,08)^2 \cdot \sigma_A^2 = (1,08)^2 \cdot (180000)^2 = 37791360000$$

Actividad 3:

Sea X: "Nota de un alumno en el ramo de matemática".

Y: "Nota de un alumno en el ramo de estadística".

$$\bar{X} = 76 \quad \sigma_x = 10$$

$$\bar{Y} = 82 \quad \sigma_y = 16$$

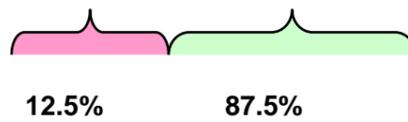
a) $\sigma_x < \sigma_y$ Por lo tanto, el ramo de matemática presenta menor desviación típica que el ramo de estadística.

Actividad 4:

SUELDO S	n_i	x_i	Ni	fi	$x_i \cdot n_i$ i	$ x_i - \bar{x} $	$ x_i - \bar{x} \cdot n_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot n_i$
0 - 20	30	10	30	0.3	30	27.5	825	756.25	22687.5
20 - 40	20	30	50	0.2	20	7.5	150	52.25	1045
40 - 60	15	50	65	0.15	15	12.5	187.5	156.25	2343.75
60 - 80	13	70	78	0.13	13	32.5	422.5	1056.2	13731.2
80 - 100	12	90	90	0.12	12	52.5	630	2756.2	33075
100 - 120	10	110	100	0.1	10	72.5	725	5256.2	52562.5
	100			1	375		2940	10033.	125445

a)

- Percentil 12.5



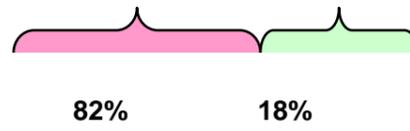
$$\frac{i \cdot n}{100} = \frac{12.5 \cdot 100}{100} = 12.5 \text{ Por lo tanto corresponde al primer intervalo, Luego}$$

$$P_{12.5} = 0 + \frac{12.5 - 0}{30} \cdot 20 = 8.33$$

El mayor sueldo del 12.5 % de las personas que menos ganan es 8.33 UF.

b)

- **Percentil 82**



$$\frac{i \cdot n}{100} = \frac{82 \cdot 100}{100} = 82 \text{ Por lo tanto corresponde al quinto intervalo, Luego}$$

$$P_{82} = 80 + \frac{82 - 78}{12} \cdot 20 = 86.67$$

El menor sueldo del 18 % de las personas que mas ganan es 86.67 UF

c)

$$\bar{X} = \frac{3750}{100} = 37.5$$

- **Desviación Media**

$$D.M = \frac{2940}{100} = 29.4$$

- **Varianza**

$$V(X) = \frac{125445}{100} = 1254.45$$

- **Desviación Estándar**

$$S(X) = \sqrt{1254.45} = 35.42$$

Actividad 5:

a)

Demand	n_i	x_i	$x_i \cdot n_i$	N_i	$ x_i - \bar{x} $	$ x_i - \bar{x} \cdot n_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot n_i$
25 - 33	3	29	87	3	23.2	69.6	538.24	1614.72
33 - 41	4	37	148	7	15.2	60.8	231.04	924.16
41 - 49	6	45	270	13	7.2	43.2	51.84	311.04
49 - 57	5	53	265	18	0.8	4	0.64	3.2
57 - 65	6	61	366	24	8.8	52.8	77.44	464.64
65 - 73	4	69	276	28	16.8	67.2	282.24	1128.96
73 - 81	2	77	154	30	24.8	49.6	615.04	1230.08
	30		1566			347.2		5676.8

b)

• **Percentil 35**

$$\frac{i \cdot n}{100} = \frac{35 \cdot 30}{100} = 10.5 \text{ Por lo tanto corresponde al tercer intervalo, Luego}$$

$$P_{35} = 41 + \frac{10.5 - 7}{6} \cdot 8 = 45.67$$

El 35% de las demandas diarias en unidades de un producto durante 30 días de trabajo son a lo más 45,67 unidades.

• **Cuartil 1**

$$\frac{n}{4} = \frac{30}{4} = 7.5 \text{ Por lo tanto corresponde al tercer intervalo, Luego}$$

$$Q_1 = 41 + \frac{7.5 - 7}{6} \cdot 8 = 41.67$$

El 25% de las demandas diarias en unidades de un producto durante 30 días de trabajo son a lo más 41.67 unidades.

• **Cuartil 3**

$$\frac{3n}{4} = \frac{3 \cdot 30}{4} = 22.5 \text{ Por lo tanto corresponde al quinto intervalo, Luego}$$

$$Q_3 = 57 + \frac{22.5 - 18}{6} \cdot 8 = 63$$

El 75% de las demandas diarias en unidades de un producto durante 30 días de trabajo son a lo más 63 unidades.

- **Mediana**

$$\frac{n}{2} = \frac{30}{2} = 15, \text{ Por lo tanto corresponde al cuarto intervalo.}$$

$$Me = 49 + \frac{15 - 13}{5} \cdot 8 = 52.2$$

- **Desviación Media**

$$\bar{X} = \frac{1566}{30} = 52.2$$

$$DM = \frac{347.2}{30} = 11.57$$

- **Desviación Estándar**

$$V(X) = \frac{5676.8}{30} = 189.23$$

$$S(X) = \sqrt{189.23} = 13.76$$

- **Diagrama de caja**

Mediana	Cuartiles	Valor mínimo	Valor máximo	Rango
93.8	$Q_1 = 41.67$ $Q_2 = 52.2$ $Q_3 = 63$	25	81	56

Diagrama de caja

