ESTADÍSTICA BÁSICA
CON APLICACIONES EN MS EXCEL

Introducción a la estadística, distribuciones de frecuencias, gráficos estadísticos, medidas de tendencia central, dispersión, posición y forma, con ejemplos resueltos en Microsoft Excel®

JUAN CARLOS VERGARA SCHMALBACH
VÍCTOR MANUEL QUESADA IBARGUEN

GMCG Grupo Métodos Cuantitativos de Gestión

Programa de Administración Industrial
Universidad de Cartagena

Universidad de Cartagena

ESTADÍSTICA BÁSICA CON APLICACIONES EN MS EXCEL®

JUAN CARLOS VERGARA SCHMALBACH

VÍCTOR MANUEL QUESADA IBARGÜEN

CONTENIDO

PRÓLOGO 9
LOS AUTORES 10
LIBRO: MÉTODOS CUANTITATIVOS CON WINQSB 11
INTRODUCCIÓN 12
1. INTRODUCCIÓN A LA ESTADÍSTICA 14
1.1 UN POCO DE HISTORIA 14
1.2 DEFINICIÓN Y CLASIFICACIÓN 15
1.2.1 Clasificación de la estadística 16
1.3 LA INVESTIGACIÓN ESTADÍSTICA 17
2. LAS TABLAS DE FRECUENCIA ..... **¡Error! Marcador no definido.**
2.1 CONSTRUCCIÓN Y CARACTERÍSTICAS DE LAS TABLAS TIPO A. **¡Error! Marcador no definido.**
2.1.1 Ejemplo tabla de frecuencia tipo A ..... **¡Error! Marcador no definido.**
2.1.2 Características de las tablas Tipo A ..... **¡Error! Marcador no definido.**
2.1.3 Construcción de las tablas Tipo A en Excel ..... **¡Error! Marcador no definido.**
2.2 CONSTRUCCIÓN Y CARACTERÍSTICAS DE LAS TABLAS TIPO B. **¡Error! Marcador no definido.**
2.2.1 Ejemplo 1: tablas de frecuencia tipo B .... **¡Error! Marcador no definido.**
2.2.2 El Problema de la Ambigüedad en las Tablas de Frecuencia Tipo B ..... **¡Error! Marcador no definido.**
2.2.3 Ejemplo 2: tablas de frecuencia tipo B .... **¡Error! Marcador no definido.**
2.2.4 Características de las tablas tipo B ..... **¡Error! Marcador no definido.**
2.2.4 Construcción de las tablas tipo B en Excel ..... **¡Error! Marcador no definido.**
2.3 EJERCICIOS PROPUESTOS ..... **¡Error! Marcador no definido.**
2.4 CASO: LA GESTIÓN DEL GOBIERNO ..... **¡Error! Marcador no definido.**
2.5 CUESTIONARIO DE REPASO ..... **¡Error! Marcador no definido.**
3. GRÁFICOS ESTADÍSTICOS ..... **¡Error! Marcador no definido.**
3.1 GRÁFICOS DE SECTORES ..... **¡Error! Marcador no definido.**

3.1.1 Ejemplo de gráficos de sectores ..... **¡Error! Marcador no definido.**
3.1.2 Características de los gráficos de sectores..... **¡Error! Marcador no definido.**
3.1.3 Construcción de los gráficos de sectores en Excel . **¡Error! Marcador no definido.**
3.2 GRÁFICOS DE COLUMNAS..... **¡Error! Marcador no definido.**
3.2.1 Ejemplo de gráficos de columnas..... **¡Error! Marcador no definido.**
3.2.2 Características de los gráficos de columnas ..... **¡Error! Marcador no definido.**
3.2.3 Construcción de los gráficos de columnas en Excel **¡Error! Marcador no definido.**
3.3 HISTOGRAMA..... **¡Error! Marcador no definido.**
3.3.1 Ejemplo de histogramas..... **¡Error! Marcador no definido.**
3.3.2 Características de los histogramas ..... **¡Error! Marcador no definido.**
3.3.3 Construcción de histogramas en Excel ... **¡Error! Marcador no definido.**
3.4 POLÍGONOS DE FRECUENCIAS..... **¡Error! Marcador no definido.**
3.4.1 Ejemplo de polígonos de frecuencias..... **¡Error! Marcador no definido.**
3.4.2 Características de los polígonos de frecuencias ..... **¡Error! Marcador no definido.**
3.4.3 Construcción de los polígonos de frecuencias en Excel ..... **¡Error! Marcador no definido.**
3.5 CURVAS SUAVIZADAS O CURVAS DE FRECUENCIAS . **¡Error! Marcador no definido.**
3.5.1 Construcción de las curvas suavizadas en Excel.... **¡Error! Marcador no definido.**
3.5.2 Características de las curvas suavizadas **¡Error! Marcador no definido.**
3.6 OJIVAS..... **¡Error! Marcador no definido.**
3.6.1 Ejemplo de ojivas ..... **¡Error! Marcador no definido.**
3.6.2 Características de las ojivas..... **¡Error! Marcador no definido.**
3.6.3 Interpretando la información en las ojivas **¡Error! Marcador no definido.**
3.6.4 Construcción de ojivas en Excel ..... **¡Error! Marcador no definido.**
3.7 PICTOGRAMAS ..... **¡Error! Marcador no definido.**

3.7.1 Ejemplo de pictogramas ..... **¡Error! Marcador no definido.**
3.7.2 Características de los pictogramas ..... **¡Error! Marcador no definido.**
3.8 EJERCICIOS PROPUESTOS ..... **¡Error! Marcador no definido.**
3.9 CASO: EL PROVEEDOR DE TUBOS DE ACERO ..... **¡Error! Marcador no definido.**
3.10 CUESTIONARIO DE REPASO..... **¡Error! Marcador no definido.**
CAPITULO 4: MEDIDAS DE TENDENCIA CENTRAL **¡Error! Marcador no definido.**
4.1 LA MEDIA ARITMÉTICA ..... **¡Error! Marcador no definido.**
4.1.1 Media aritmética para datos no agrupados ..... **¡Error! Marcador no definido.**
4.1.2 Ejemplo: la media aritmética para datos no agrupados **¡Error! Marcador no definido.**
4.1.3 Media aritmética para datos agrupados .. **¡Error! Marcador no definido.**
4.1.4 Ejemplo: media aritmética para datos agrupados en tablas tipo A **¡Error! Marcador no definido.**
4.1.5 Ejemplo: media aritmética para datos agrupados en tablas tipo B **¡Error! Marcador no definido.**
4.1.6 Ejemplo: comparativa entre el cálculo de la media aritmética para datos no agrupados y datos agrupados en tablas tipo B ..... **¡Error! Marcador no definido.**
4.1.7 Cálculo de la media aritmética en Excel.. **¡Error! Marcador no definido.**
4.1.8 Ventajas ..... **¡Error! Marcador no definido.**
4.1.9 Desventajas ..... **¡Error! Marcador no definido.**
4.2 LA MEDIANA ..... **¡Error! Marcador no definido.**
4.2.1 Ejemplo: mediana para datos no agrupados (cantidad de datos impar) ..... **¡Error! Marcador no definido.**
4.2.2 Ejemplo: mediana para datos no agrupados (cantidad de datos par) ..... **¡Error! Marcador no definido.**
4.2.3 Ejemplo: mediana para datos agrupados en tablas tipo A ..... **¡Error! Marcador no definido.**

4.2.4 Ejemplo: mediana para datos agrupados en tablas tipo B ..... **¡Error! Marcador no definido.**
4.2.5 La fórmula para calcular la mediana ..... **¡Error! Marcador no definido.**
4.2.6 Ubicando la mediana en el gráfico de ojiva..... **¡Error! Marcador no definido.**
4.2.7 Calculo de la mediana en Excel ..... **¡Error! Marcador no definido.**
4.2.8 Ventajas ..... **¡Error! Marcador no definido.**
4.2.9 Desventajas ..... **¡Error! Marcador no definido.**
4.3 LA MODA..... **¡Error! Marcador no definido.**
4.3.1 Ejemplo: moda para datos no agrupados **¡Error! Marcador no definido.**
4.3.2 Ejemplo: moda para datos agrupados..... **¡Error! Marcador no definido.**
4.3.3 Calculo de la moda mediante fórmula ..... **¡Error! Marcador no definido.**
4.3.4 Calculo de la mediana en Excel ..... **¡Error! Marcador no definido.**
4.3.5 Ventajas ..... **¡Error! Marcador no definido.**
4.3.6 Desventajas ..... **¡Error! Marcador no definido.**
4.4 EJERCICIOS PROPUESTOS ..... **¡Error! Marcador no definido.**
4.5 CASO: POBLACIÓN Y MUESTRA..... **¡Error! Marcador no definido.**
4.6 CUESTIONARIO DE REPASO..... **¡Error! Marcador no definido.**
CAPITULO 5: MEDIDAS DE DISPERSIÓN ..... **¡Error! Marcador no definido.**
5.1 DESVIACIÓN MEDIA ..... **¡Error! Marcador no definido.**
5.1.1 Ejemplo: Desviación media para datos no agrupados . **¡Error! Marcador no definido.**
5.1.2 Ejemplo: Desviación media para datos agrupados . **¡Error! Marcador no definido.**
5.1.3 Cálculos de la desviación media en Excel **¡Error! Marcador no definido.**
5.2 LA VARIANZA ..... **¡Error! Marcador no definido.**
5.2.1 Ejemplo: Varianza para datos no agrupados ..... **¡Error! Marcador no definido.**
5.2.2 Ejemplo: Varianza para datos agrupados **¡Error! Marcador no definido.**
5.2.3 Cálculo de la varianza en Excel ..... **¡Error! Marcador no definido.**
5.3 DESVIACIÓN ESTÁNDAR ..... **¡Error! Marcador no definido.**

5.3.1 Ejemplo: Desviación estándar para datos no agrupados ..... **¡Error! Marcador no definido.**
5.3.2 Ejemplo: Desviación estándar para datos agrupados .. **¡Error! Marcador no definido.**
5.3.3 Cálculo de la Desviación estándar en Excel ..... **¡Error! Marcador no definido.**
5.4 COEFICIENTE DE VARIACIÓN ..... **¡Error! Marcador no definido.**
5.4.1 Ejemplo: Desviación estándar para datos no agrupados ..... **¡Error! Marcador no definido.**
5.4.2 Calculo del coeficiente de variación en Excel ..... **¡Error! Marcador no definido.**
5.5 EJERCICIOS PROPUESTOS ..... **¡Error! Marcador no definido.**
5.6 CASO: EL RÍO MAGDALENA ..... **¡Error! Marcador no definido.**
CAPITULO 6: MEDIDAS DE POSICIÓN..... **¡Error! Marcador no definido.**
6.1 PERCENTILES..... **¡Error! Marcador no definido.**
6.1.1 Ejemplo: Calculo de percentiles ..... **¡Error! Marcador no definido.**
6.1.2 La fórmula para calcular percentiles..... **¡Error! Marcador no definido.**
6.2 DECILES ..... **¡Error! Marcador no definido.**
6.2.1 Ejemplo: Calculo de deciles ..... **¡Error! Marcador no definido.**
6.2.2 La fórmula para calcular deciles..... **¡Error! Marcador no definido.**
6.3 CUARTILES..... **¡Error! Marcador no definido.**
6.3.1 Ejemplo: Calculo de cuartiles ..... **¡Error! Marcador no definido.**
6.3.2 La fórmula para calcular cuartiles..... **¡Error! Marcador no definido.**
6.4 APLICACIÓN DE PERCENTILES Y CUARTILES EN EXCEL ..... **¡Error! Marcador no definido.**
6.5 LAS MEDIDAS DE POSICIÓN Y EL GRÁFICO DE OJIVA **¡Error! Marcador no definido.**
6.6 EJERCICIOS PROPUESTOS ..... **¡Error! Marcador no definido.**
6.7 CUESTIONARIO DE REPASO..... **¡Error! Marcador no definido.**
CAPITULO 7: MEDIDAS DE FORMA ..... **¡Error! Marcador no definido.**
7.1 TIPOS DE DISTRIBUCIONES DE FRECUENCIA MÁS COMUNES .. **¡Error! Marcador no definido.**

7.1.1 Distribución simétrica ..... **¡Error! Marcador no definido.**
7.1.2 Distribución asimétrica ..... **¡Error! Marcador no definido.**
7.2 RELACIÓN ENTRE LA MEDIA, MEDIANA Y MODA .... **¡Error! Marcador no definido.**
7.2.1 Ejemplo: Relación entre la media, mediana y moda **¡Error! Marcador no definido.**
7.3 COEFICIENTE DE ASIMETRÍA ..... **¡Error! Marcador no definido.**
7.3.1 Ejemplo: Cálculo del coeficiente de asimetría..... **¡Error! Marcador no definido.**
7.4 CURTOSIS ..... **¡Error! Marcador no definido.**
7.4.1 Ejemplo: Cálculo de la Curtosis ..... **¡Error! Marcador no definido.**
7.5 EJERCICIOS PROPUESTOS ..... **¡Error! Marcador no definido.**
ANEXO A. RESUMEN DE SIMBOLOGÍA Y FÓRMULAS ESTADÍSTICAS UNIDAD II ..... **¡Error! Marcador no definido.**
ANEXO B. RESUMEN DE SIMBOLOGÍA Y FÓRMULAS ESTADÍSTICAS UNIDAD IV ..... **¡Error! Marcador no definido.**
ANEXO C. RESUMEN DE SIMBOLOGÍA Y FÓRMULAS ESTADÍSTICAS UNIDAD V ..... **¡Error! Marcador no definido.**
ANEXO D. RESUMEN DE SIMBOLOGÍA Y FÓRMULAS ESTADÍSTICAS UNIDAD VI ..... **¡Error! Marcador no definido.**
ANEXO E. RESUMEN DE SIMBOLOGÍA Y FÓRMULAS ESTADÍSTICAS UNIDAD VII ..... **¡Error! Marcador no definido.**

PRÓLOGO

El lector seguramente se preguntará el porqué y para qué de un nuevo texto de estadística básica, con la abundancia de manuales de estadística como existen hoy día, de manera que se hace necesario, desde estas primeras páginas ponerle en conocimiento las diferencias de este nuevo texto con cualquier otro que haya tenido a mano.

En efecto, en cuanto a contenido, es posible que usted, señor estudiante, docente, empresario o, en general, usuario de esta obra, sólo encuentre temas comunes a cualquier libro de estadística; no obstante encontrará un aspecto diferenciador que constituye un verdadero valor añadido que le permitirá agilizar el estudio de las técnicas estadísticas ahorrándole tiempo apreciable en el manejo de datos y por ende dejándole un mayor espacio disponible para el análisis de la información requerida para la toma de decisiones.

Es común en la enseñanza de la estadística en las escuelas de ingeniería, administración y demás disciplinas que la utilizan, que los docentes reclamen la adquisición de software especializado para el manejo de su asignatura, a lo que no siempre las instituciones responden con la debida diligencia. Pero aún cuando haya respuesta positiva en este sentido, que las instituciones se preocupen por mantenerse actualizadas en materia de software licenciado, en ocasiones éstos revisten tal complejidad en su manejo que tanto estudiantes como profesores desisten de su uso, permaneciendo la enseñanza de esta materia en una situación de “manualidad” que obliga a que la mayor parte del tiempo presupuestado para su desarrollo se invierta en la llamada “carpintería” y muy poco en el análisis, cual es la finalidad última si se desea hacer uso óptimo de la información disponible.

La obra que hoy ofrecemos a la comunidad académica y empresarial cuenta con la ventaja de estar basada en Excel®, un software al alcance de cualquier institución o persona y de fácil manejo por parte de cualquier usuario, de tal forma que al tiempo que se imparte la asignatura se logra tanto el dominio de la estadística como del Excel®, una valiosa herramienta para la gestión de procesos administrativos a cualquier nivel.

Los estudiantes de la estadística descriptiva encontrarán en este texto los conceptos básicos y la metodología para la manipulación de datos para producir la información relevante para el uso requerido.

Esperamos que el libro tenga la acogida que se merece pues, siendo un tanto inmodestos, pretendemos que constituya un aporte significativo a la mejora de los métodos de enseñanza de esta importante asignatura,

Los autores

LOS AUTORES

VÍCTOR MANUEL QUESADA IBARGUEN: Docente Titular de la Universidad de Cartagena en el área de los métodos cuantitativos del programa de Administración Industrial. Ingeniero Industrial de la Universidad INCCA, Especialista en Finanzas U. del Valle, Especialista en Investigación U. de Cartagena, Magíster en Economía de la Universidad Nacional de Colombia, Ph.D. Ingeniería de Organización, Universidad de Sevilla España. Perteneciente al Grupo de Investigación de Métodos Cuantitativos de Gestión (***GMCG***). Libros publicados: Programación Lineal (S/ISBN), Programación Lineal y Entera. ISBN 958 – 33 – 0588 – X (1997), Productividad y Eficiencia en la Empresa: Un Enfoque Práctico ISBN-958-9230-19-9 (2003), Métodos Cuantitativos con WINQSB ISBN-978-84-690-3681-5 (2007).

Correo electrónico: quesastoque@une.com.co

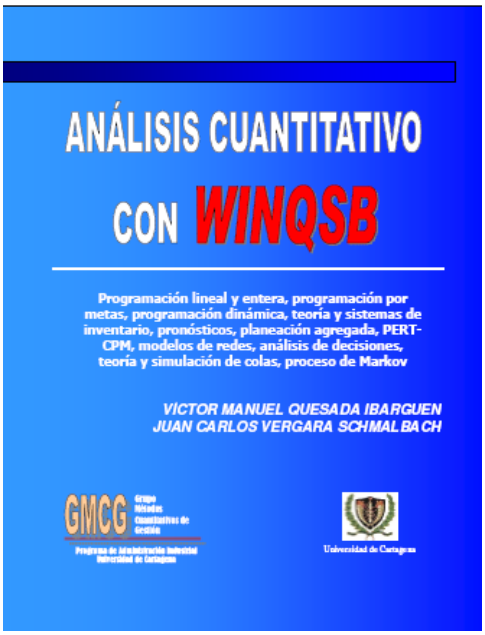
JUAN CARLOS VERGARA SCHMALBACH: Ingeniero Industrial de la Universidad Tecnológica de Bolívar, especialista en finanzas de la Universidad de Cartagena y magister en Administración de Empresas de la Universidad Nacional. Docente de tiempo completo de la Universidad de Cartagena en el área de los métodos cuantitativos del programa de Administración Industrial. Perteneciente al Grupo de Investigación de Métodos Cuantitativos de Gestión (***GMCG***). Libros publicados: Métodos Cuantitativos con WINQSB ISBN- 978-84-690-3681-5 (2006).

Correo Electrónico: juancarlosvergaras@yahoo.com.mx. Página WEB: http://juancarlosvergara.50webs.org.

Página WEB grupo métodos cuantitativos de gestión:

http://metodoscuantitativos.50webs.org

LIBRO: MÉTODOS CUANTITATIVOS CON WINQSB



El libro **MÉTODOS CUANTITATIVOS CON WINQSB** lo introducirá en el apasionante mundo de la solución de problemas complejos mediante el uso de software para computadoras.

WINQSB es una aplicación versátil que permite la solución de una gran cantidad de problemas desde niveles administrativos, producción, recurso humano hasta la dirección de proyectos.

Debido a su facilidad y potencia de manejo, este libro se convierte en una herramienta indispensable para el estudiante de pregrado o postgrado que participa en materias relacionadas como la investigación de operaciones, los métodos de trabajo, planeación de la producción, evaluación de proyectos, control de calidad, simulación, estadística, entre otras.

Los módulos tratados en este libro son:

- Programación Lineal y Entera
- Programación por Metas
- PERT – CPM
- Planeación Agregada
- Pronósticos
- Teoría y Sistemas de Inventario
- Análisis de Decisiones
- Planeación de Requerimiento de Materiales (MRP)
- Programación Dinámica
- Modelos de Redes
- Teoría y simulación de colas
- Cadenas de Markov

Puede conseguir la versión electrónica en la página web http://www.eumed.net o en http://metodoscuantitativos.50webs.org

INTRODUCCIÓN

El libro **ESTADÍSTICA BÁSICA CON APLICACIONES EN MS EXCEL®** esta diseñado como un texto introductorio hacia la Estadística Descriptiva, a partir de ejercicios resueltos paso a paso, utilizando como complemento Microsoft Excel®.

Al terminar el curso, el estudiante comprenderá el origen de la Estadística, los métodos cuantitativos básicos para el tratamiento de datos y un manejo en las funciones estadísticas ofrecidas por Microsoft Excel®.

El libro cuenta con siete capítulos donde se presenta una introducción teórica, ejercicios resueltos paso a paso, ejercicios propuestos, un cuestionario y un resumen de fórmulas utilizadas por capitulo. La temática tratada se resume en:

Capitulo 1 - Introducción a la estadística: Incluye una breve historia del origen y desarrollo de la estadística. Al igual que los conceptos básicos necesarios para iniciar el curso.

Capitulo 2 – Tablas de frecuencia: Tabulación de datos en tablas simples (llamadas tipo A) y con intervalos de clases (tipo B).

Capitulo 3 – Gráficos estadísticos: Gráficos construidos a partir de las tablas de frecuencias.

Capitulo 4 – Medidas de tendencia central: Calculo de la media, mediana y moda.

Capitulo 5 – Medidas de dispersión: Cálculo de la desviación media, varianza y desviación estándar.

Capitulo 6 – Medidas de posición: Cálculo de percentiles, Deciles y cuartiles.

Capitulo 7 – Medidas de forma: Cálculo de indicadores que identifican la forma en que se distribuyen los datos.

El libro cuenta con pequeños iconos que identifican secciones especiales:

Identificador de definiciones de términos estadísticos

Ejercicio resuelto en Microsoft Excel®

Formato de la función estadística empleada en Microsoft Excel®

CAPITULO 1

INTRODUCCIÓN A LA ESTADÍSTICA

En este capitulo se hará un pequeño recorrido por la historia de la estadística, mencionando algunos de los personajes que la impulsaron, recalcando sus progresos y aportes a través del tiempo.

Analizaremos los conceptos que adoptan los distintos autores sobre la definición de estadísticas y su clasificación.

Por último, entenderemos la estadística como una herramienta de apoyo a la investigación de tipo cuantitativa, la cual se hace participe desde la recolección de datos hasta el análisis de los mismos.

1. INTRODUCCIÓN A LA ESTADÍSTICA

1.1 UN POCO DE HISTORIA

El uso de herramientas cuantitativas para el tratamiento de datos, tiene origen en épocas remotas. Se tiene información de hace más 3000 años antes de Cristo, donde las antiguas civilizaciones, como la Egipcia, aplicaron continuamente censos que ayudaban a la organización del estado y la construcción de las pirámides.

El antiguo testamento nos sugiere que Moisés ordenó un "*Censo*" a la población Israelita para identificar los miembros de las familias. En la antigua Grecia y el Imperio Romano, era común la aplicación de censos para la planificación de impuestos y la prestación del servicio militar.

La palabra estadística deriva del latín moderno *statisticum collegium* ("consejo de estado"), del latín antiguo *status* ("posición", "forma de gobierno"), de la palabra italiana moderna *statista* ("estadista", "político") y del italiano antiguo *stato* ("estado"). En 1749, el alemán, Gottfried Achenwall (1719-1792) usa el término *Statistik* en su libro titulado "*Staatswissenschaft der vornehmen Europäischen Reiche und Republiken*", quien originalmente designó la palabra estadística para el análisis de los datos de un gobierno, definiéndola como la "Ciencia del Estado". A Gottfried Achenwall se le conoce como el "Padre de la Estadística".

La primera persona que introdujo el término estadística en Inglaterra fue Sir John Sinclair (1754-1835) con su trabajo "*Statistical Account of Scotland*" (1791-,1799) trabajo compilado en 21 volumenes. El autor explica en su libro, que la palabra estadística la adoptó gracias al estudio de investigaciones realizadas en Alemania, como una palabra novedosa que llamaría la atención de los ingleses; a diferencia, de que en Alemania la estadística se usa como instrumento para medir la fortaleza de un estado, mientras que Sinclair, la emplearía como generadora de información interna para encontrar falencias y proponer mejoras en el país. A este trabajo le siguieron dos publicaciones: la segunda edición elaborada entre 1834 y 1845; la tercera edición comienza después de la segunda guerra mundial comprendiendo los periodos entre 1951 y 1992.[1]

A comienzos del siglo XIX, la palabra estadística adopta un significado más generalizado hacia la recolección y clasificación de cualquier tipo de datos cuantitativos.

William Playfair (1759-1823) expone su idea de que los gráficos permiten una comunicación más eficiente que las tablas de frecuencia. Es considerado como el inventor de los gráficos lineales, de barras y de sectores. Playfair publicó el libro titulado "*The Commercial and Political Atlas*" (1786) el cual contiene 43 gráficos de

[1] Para obtener el texto completo de las ediciones puede dirigirse al siguiente link http://edina.ac.uk/stat-acc-scot

series de tiempo y por primera vez, es usado un gráfico de barras. En 1801 utiliza el primer gráfico de sectores en su obra "*Playfair's Statistical Breviary*".

Sir Francis Galton (1822-1911) creó el concepto estadístico de regresión y correlación, y fue el primero en aplicar métodos estadísticos para estudiar las diferencias humanas basado en el uso de cuestionarios y entrevistas para recolectar los datos.

Herman Hollerith (1860-1929) fue un estadístico estadounidense quien desarrollo la primera máquina tabuladora basada en tarjetas perforadas y mecanismos eléctrico-mecánicos para el tratamiento rápido de millones de datos. Su máquina fue usada en el censo de 1890 en estados unidos que redujo la tabulación de los datos de 7 años (censo de 1880) a 2.5 años. Creó la firma "*Computing Tabulating Recording Corporation (CTR)*", que bajo la presidencia de Thomas J. Watson fue renombrada a "*International Business Machines (IBM)*" en 1924.

Major Greenwwod (1880-1949) investiga los problemas de salud asociados al trabajo en fábricas. Desarrolló la Epidemiología y en 1919 creó el Ministerio de la Salud en Inglaterra, responsable de datos estadísticos médicos.

1.2 DEFINICIÓN Y CLASIFICACIÓN

Como vimos en el apartado anterior, la estadística a variado su significado a través del tiempo, pasando de ser una herramienta usada solo para la administración de los gobiernos, a una ciencia con un sin fin de aplicaciones en diferentes disciplinas.

Estadística: La enciclopedia Británica define la estadística como la ciencia encargada de recolectar, analizar, presentar e interpretar datos.

La estadística pasa a ser una ciencia básica cuyo objetivo principal es el procesamiento y análisis de grandes volúmenes de datos, resumiéndolos en tablas, gráficos e indicadores (estadísticos), que permiten la fácil compresión de las características concernientes al fenómeno estudiado.

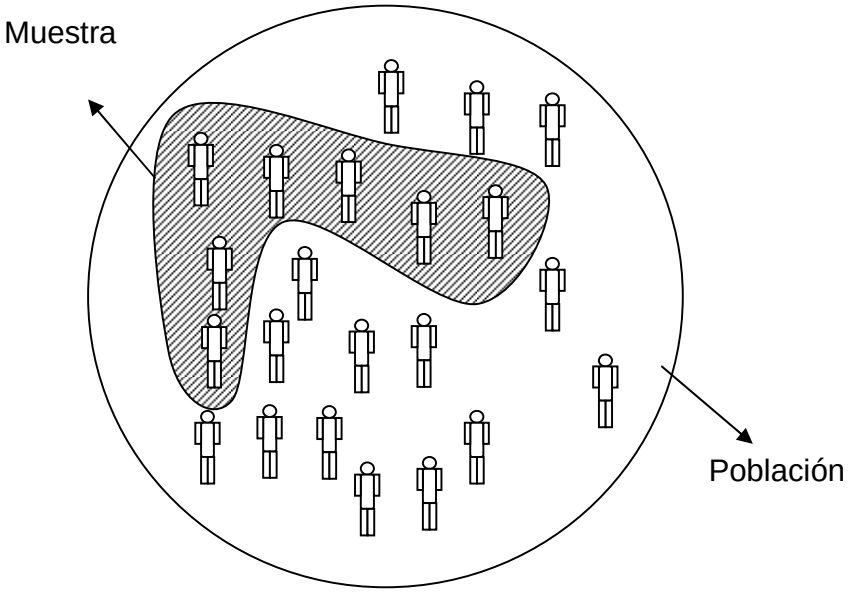
Estadística: El famoso diccionario Ingles Word Reference define la estadística como un área de la matemática aplicada orientada a la recolección e interpretación de datos cuantitativos y al uso de la teoría de la probabilidad para calcular los parámetros de una población.

Estadístico: Cualquier característica medible calculada sobre una muestra o población.

Los datos pueden provenir de una población o muestra. Esto datos deben ser cuantitativos, para así poder aplicar sobre ellos, operaciones aritméticas.

Muestra: Es un subconjunto de una población. Una muestra es representativa cuando los elementos son seleccionados de tal forma que pongan de manifiesto las características de una población. Su característica más importante es la representatividad.

La selección de los elementos que conforman una muestra pueden ser realizados de forma probabilística o aleatoria (al azar), o no probabilística.



1.2.1 Clasificación de la estadística

La estadística se puede clasificar en dos grandes ramas:

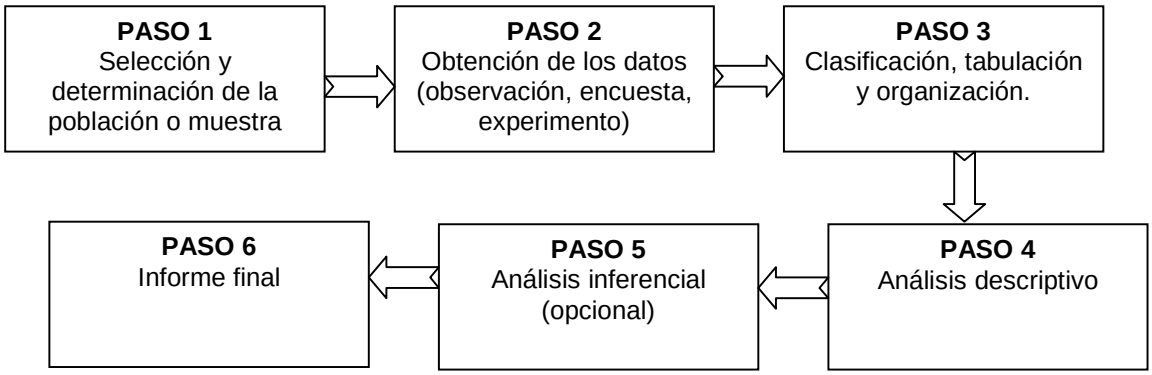
- Estadística descriptiva o deductiva.
- Estadística inferencial o inductiva.

La primera se emplea simplemente para resumir de forma numérica o gráfica un conjunto de datos. Se restringe a describir los datos que se analizan. Si aplicamos las herramientas ofrecidas por la estadística descriptiva a una muestra, solo nos limitaremos a describir los datos encontrados en dicha muestra, no se podrá generalizar la información hacia la población. La estadística inferencial permite realizar conclusiones o inferencias, basándose en los datos simplificados y analizados de una muestra hacia la población o universo. Por ejemplo, a partir de una muestra representativa tomada a los habitantes de una ciudad, se podrá inferir la votación de todos los ciudadanos que cumplan los requisitos con un error de aproximación.

1.3 LA INVESTIGACIÓN ESTADÍSTICA

El proceso de aplicación de la estadística implica una serie de pasos:

1. Selección y determinación de la población o muestra y las características contenidas que se desean estudiar. En el caso de que se desee tomar una muestra, es necesario determinar el tamaño de la misma y el tipo de muestreo a realizar (probabilístico o no probabilístico).
2. Obtención de los datos. Esta puede ser realizada mediante la observación directa de los elementos, la aplicación de encuestas y entrevistas, y la realización de experimentos.
3. Clasificación, tabulación y organización de los datos. La clasificación incluye el tratamiento de los datos considerados anómalos que pueden en un momento dado, falsear un análisis de los indicadores estadísticos. La tabulación implica el resumen de los datos en tablas y gráficos estadísticos.
4. Análisis descriptivo de los datos. El análisis se complementa con la obtención de indicadores estadísticos como las medidas: de tendencia central, dispersión, posición y forma.
5. Análisis inferencial de los datos. Se aplican técnicas de tratamiento de datos que involucran elementos probabilísticos que permiten inferir conclusiones de una muestra hacia la población (opcional).
6. Elaboración de conclusiones. Se construye el informe final.



CAPITULO 2

TABLAS DE FRECUENCIA

El principal objetivo de la estadística descriptiva es sintetizar conjuntos de datos mediante tablas o gráficos resumen, con el fin de poder identificar el comportamiento característico de un fenómeno y facilitar su análisis exhaustivo.

Cualquier investigación que se emprenda puede conducir a la acumulación de valores cuantitativos y cuasi-cualitativos correspondientes a las diversas medidas efectuadas. Esta posibilidad, convierte a la estadística en una herramienta vital para el tratamiento de volúmenes de datos mediante tablas resúmenes conocidas como **"*Tablas de Frecuencia*"**. Cuando los datos son agrupados, la interpretación resulta ser más sencilla.

2. LAS TABLAS DE FRECUENCIA

Tablas de Frecuencia: Tablas estadísticas que agrupan diversos valores de una variable, simplificando los datos.

Para entender como funcionan las tablas de frecuencia, analicemos el siguiente ejemplo:

Una persona lanza una moneda 10 veces, y registra si el lado superior cae en cara (C) o sello (S). Los resultados del experimento se muestran a continuación:

C, S, S, C, C, S, S, C, S, C

La forma de simplificar los datos anteriores equivale a contar cuantas veces se repite cada lado de la moneda. A esta operación la conoceremos como **"*frecuencia Absoluta*"**.

Frecuencia Absoluta (f): Numero de veces que se repite un valor dentro de un conjunto de datos.

El lado cara se repitió 5 veces y el lado sello 5 veces. Nótese que la suma de las frecuencias equivale al total de lanzamientos.

Lado	Frecuencia (f)
Cara (C)	5
Sello (S)	5

Podemos identificar dos tipos de tablas de frecuencias, las cuales denotaremos como tablas tipo A y B.

2.1 CONSTRUCCIÓN Y CARACTERÍSTICAS DE LAS TABLAS TIPO A

Las Tablas Tipo A se caracterizan por manejar un conjunto pequeño de posibles resultados de una variable dentro de la muestra o población. Por lo general, su uso tiende al manejo de datos cualitativos o variables cuantitativas discretas.

2.1.1 Ejemplo tabla de frecuencia tipo A

Una empresa decide medir el grado de aceptación de 10 clientes sobre un nuevo producto que hace poco salió al mercado. Para tal fin, se les pide que valoren, empleando una escala del 1 al 5, su opinión frente al producto. (**1 = Muy Malo, 2 = Malo, 3 = Regular, 4 = Bueno y 5 = Excelente**). Las respuestas tabuladas de los 10 clientes son:

Cliente	Respuesta
1	2
2	5
3	4
4	5
5	4
6	3
7	4
8	5
9	3
10	5

En presencia de estos puntajes, la persona encargada del proyecto, pide que se simplifiquen y luego se interpreten los datos.

SOLUCIÓN

Como podemos observar, el numero de resultados que puede alcanzar la variable grado de aceptación son relativamente pocos (solo cinco posibilidades), por lo cual identificaremos la tabla de frecuencia resultante como una Tabla Tipo A.

Otra forma de catalogar los datos es conociendo la distancia o variación que hay entre el valor menor (Xmin) y el valor mayor (Xmax), diferencia que de ahora en adelante la conoceremos como “Rango”.

Rango (R): Diferencia existente entre el valor Máximo (X_{max}) y el valor Mínimo (X_{min}) de un conjunto de datos. La fórmula empleada es:

$$R = X_{\max} - X_{\min}$$

En nuestro ejemplo R seria igual a 4.

$$R = 5 - 1$$
$$R = 4$$

Si el rango manejado es pequeño, bastara representar los datos con una tabla Tipo A. Para crear esta tabla deberemos seguir los siguientes pasos:

PASO 1: Contar las veces que se repite cada valor dentro de la muestra.

PASO 2: Ubicar estas frecuencias en una tabla ordenada.

Grado de Aceptación	Frecuencia (f)
1	0
2	1
3	2
4	3
5	4
TOTAL	**10**

Ninguno de los clientes valoró al producto como muy malo (grado de aceptación igual a 1), mientras que la mayoría de las respuestas se centraron en valorar al producto como Excelente y Bueno (grado de aceptación iguales a 5 y 4 respectivamente).

Observando los datos resumidos, podemos concluir que la mayoría de las personas encuestadas tienen una visión favorable del nuevo producto. Queda claro, como la tabla de frecuencia agiliza el análisis de los datos.

Nótese que la sumatoria de las frecuencias es igual al número de personas encuestadas (10), por lo cual podemos llegar a la siguiente conclusión:

$$\sum_{i=1}^{Nc} f_i = n$$

Donde Nc representa el número de posibles resultados tabulados en la tabla (que de ahora en adelante se le conocerán como “clases”). En el caso de que se entreviste a toda la población, la fórmula se adaptaría así:

$$\sum_{i=1}^{Nc} f_i = N$$

La estadística considera otros tipos de frecuencias auxiliares que complementan el análisis de las tablas de frecuencia.

Frecuencia Absoluta Acumulada (F): Presenta un saldo acumulado de las frecuencias de los intervalos. Esta frecuencia se calcula sumando el acumulado de las frecuencias de los intervalos anteriores más la frecuencia absoluta del intervalo actual.

$$F_i = F_{i-1} + f_i$$

La Frecuencia Absoluta Acumulada del último intervalo es igual al tamaño de la muestra (o población). Siguiendo con el ejemplo, tenemos:

Grado de aceptación 1: $F_1 = 0$

Grado de aceptación 2: $F_2 = 0 + 1 = 1$

Grado de aceptación 3: $F_3 = 1 + 2 = 3$

Grado de aceptación 4: $F_4 = 3 + 3 = 6$

Grado de aceptación 5: $F_5 = 6 + 4 = 10$

Esta frecuencia no proporciona de inmediato el número de casos que queda por debajo de cada clase. La F_4, por ejemplo, nos dice que seis personas opinaron que el producto se encontraba entre muy malo y bueno.

Frecuencia Relativa (h): Equivale a la razón de las frecuencias de cada intervalo sobre la totalidad de los datos (n o N, dependiendo del caso). Matemáticamente se expresa:

$$h_i = \frac{f_i}{n}$$

Para el ejemplo, las frecuencias relativas son:

Grado de aceptación 1: $h_1 = \frac{0}{10} = 0$

Grado de aceptación 2: $h_2 = \frac{1}{10} = 0{,}1$ ó 10%

Grado de aceptación 3: $h_3 = \frac{2}{10} = 0{,}2$ ó 20%

Grado de aceptación 4: $h_4 = \frac{3}{10} = 0{,}3$ ó 30%

Grado de aceptación 5: $h_5 = \frac{4}{10} = 0{,}4$ ó 40%

La sumatoria de las frecuencias relativas debe ser igual a 1 (si se trabaja estos valores como porcentaje, equivaldría al 100% de los datos).

$$\sum_{i=1}^{Nc} h_i = 1$$

El 40% de las personas encuestadas (h_5), opinaron que el producto es excelente.

Frecuencia Relativa Acumulada (H): Presenta un saldo acumulado de las frecuencias relativas de cada intervalo de clase. Su cálculo resulta de la suma del acumulado de las frecuencias relativas de los intervalos

anteriores más la frecuencia relativa del intervalo actual.

$$H_i = H_{i-1} + h_i$$

La última de las Frecuencias Relativas Acumuladas equivale a 1. Las tablas de frecuencias suelen mostrar tanto las frecuencias absolutas, como relativas.

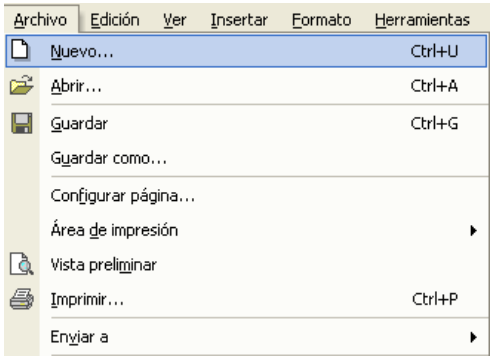
Grado de Aceptación (Clase)	F	F	h	H
1	0	0	0,0	0,0
2	1	1	0,1	0,1
3	2	3	0,2	0,3
4	3	6	0,3	0,6
5	4	10	0,4	1,0
TOTAL	**10**		**1,0**	

2.1.2 Características de las tablas Tipo A

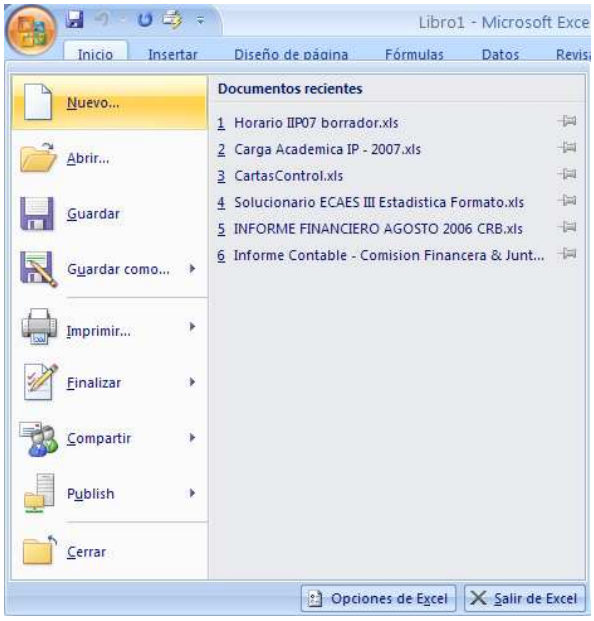
- El número de posibles valores que toma la variable debe ser reducido. (Rango pequeño).
- Suele ser utilizada en la cuantificación de las variables cuasi-cualitativas.
- Maneja variables cuantitativas cuyos valores son preferiblemente discretos.
- Su construcción es sencilla.
- La interpretación equivale a especificar la frecuencia de cada resultado.

2.1.3 Construcción de las tablas Tipo A en Excel

Creemos un archivo nuevo en Microsoft ® Excel pulsando sobre la opción Nuevo en el menú Archivo o pulsando sobre el botón que se encuentra en la barra de herramientas.



Aquellos que posean la versión 2007, deberán oprimir el botón **ARCHIVO** que se encuentra en la esquina superior izquierda de la ventana.



Ubiquémonos a partir de la celda **B2** y escribamos los siguientes datos (una palabra en cada celda):

Azul	Verde	Rojo	Blanco	Rojo	Verde	Rojo
Rojo	Azul	Azul	Blanco	Verde	Azul	Rojo
Verde	Azul	Azul	Rojo	Verde	Azul	Rojo
Verde	Verde	Azul	Rojo	Negro	Negro	Negro
Azul	Verde	Verde	Verde	Blanco	Azul	Rojo
Negro	Negro	Azul	Negro	Blanco	Rojo	Azul
Blanco	Azul	Negro	Azul	Azul	Blanco	Azul
Rojo	Blanco	Rojo	Blanco	Rojo	Blanco	Verde
Blanco	Azul	Verde	Azul	Rojo	Verde	Negro
Blanco	Rojo	Azul	Azul	Rojo	Azul	Blanco

En Excel debería verse así:

G16

	A	B	C	D	E	F	G	H	I
1									
2		Azul	Verde	Rojo	Blanco	Rojo	Verde	Rojo	
3		Rojo	Azul	Azul	Blanco	Verde	Azul	Rojo	
4		Verde	Azul	Azul	Rojo	Verde	Azul	Rojo	
5		Verde	Verde	Azul	Rojo	Negro	Negro	Negro	
6		Azul	Verde	Verde	Verde	Blanco	Azul	Rojo	
7		Negro	Negro	Azul	Negro	Blanco	Rojo	Azul	
8		Blanco	Azul	Negro	Azul	Azul	Blanco	Azul	
9		Rojo	Blanco	Rojo	Blanco	Rojo	Blanco	Verde	
10		Blanco	Azul	Verde	Azul	Rojo	Verde	Negro	
11		Blanco	Rojo	Azul	Azul	Rojo	Azul	Blanco	
12									
13									

Construyamos la estructura de nuestra tabla de frecuencia dejando espacio para la frecuencia absoluta, absoluta acumulada, relativa y relativa acumulada a partir de la celda **B13**:

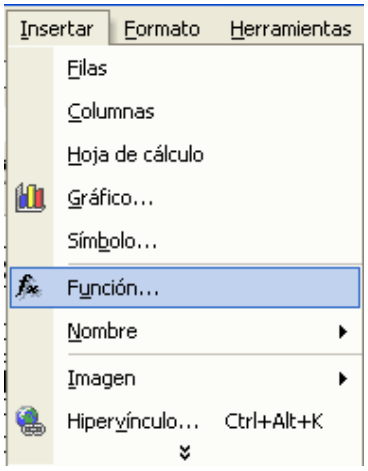
	A	B	C	D	E	F	G
12							
13		**Color**	**f**	**F**	**h**	**H**	
14		Azul					
15		Rojo					
16		Verde					
17		Negro					
18		Blanco					
19							
20							

Para determinar la cantidad de colores azul que existen en los datos iniciales emplearemos la siguiente fórmula:

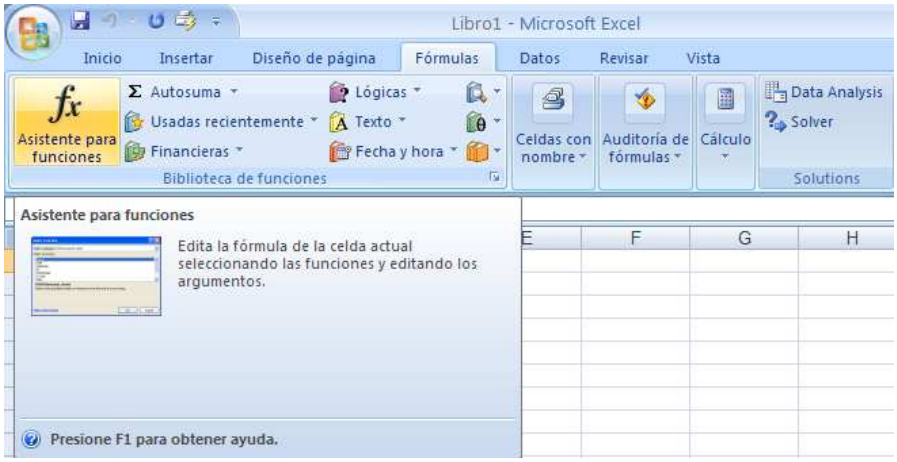
CONTAR.SI: Permite contar una palabra, número o carácter especial dentro de un conjunto de datos.

Formato: CONTAR.SI(rango;criterio)
Categoría: Estadística

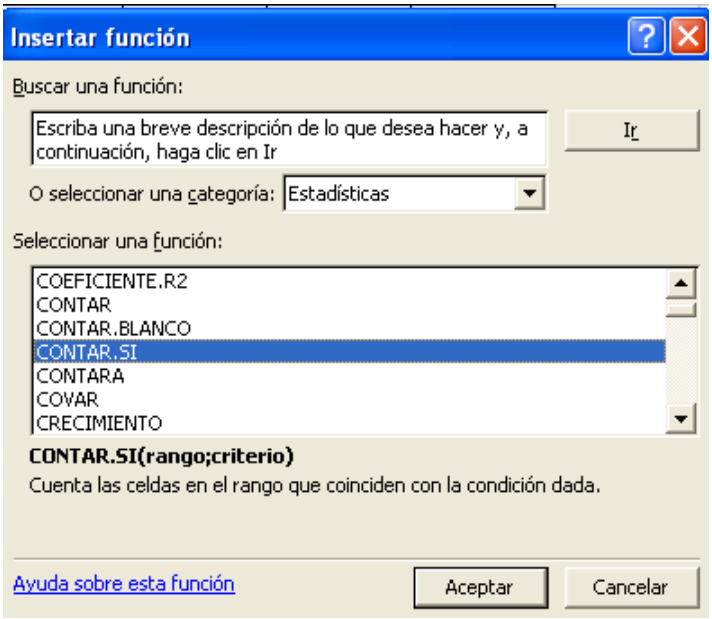
Ubicándonos en la celda **C14** activaremos la fórmula **CONTAR.SI** pulsando sobre el botón o en la opción Función en el menú Insertar.



En Microsoft Excel 2007 deberá ingresar en la ficha **FÓRMULAS** y luego pulsar sobre el botón **ASISTENTE PARA FUNCIONES**.

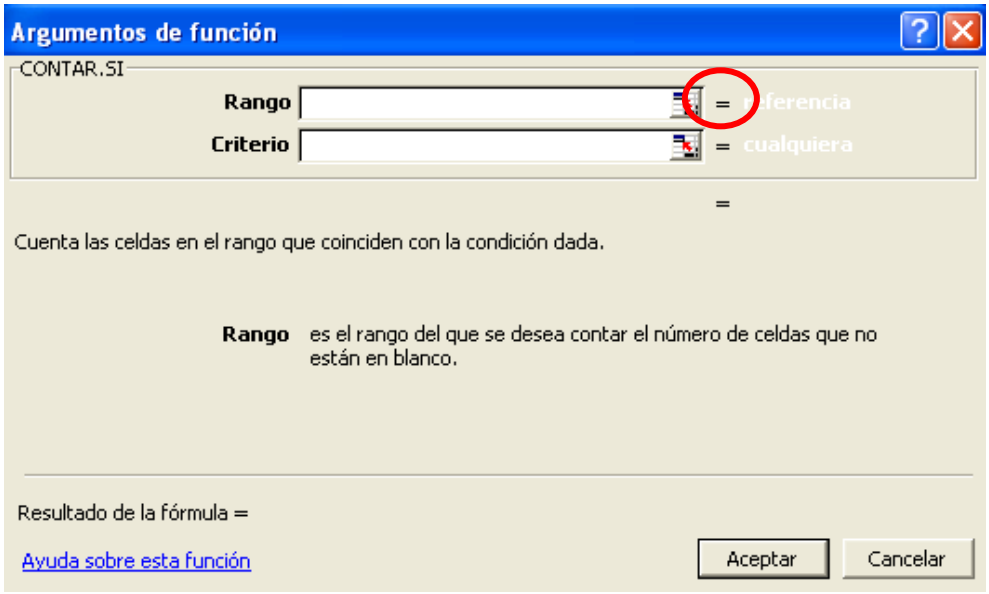


Esto generará una ventana donde se encuentran todas las funciones disponibles en Excel. Localicemos nuestra función y pulsemos en el botón Aceptar.



La función **CONTAR.SI** requiere de dos parámetros:

- El rango, equivalente a los datos iniciales.
- El criterio, haciendo alusión al valor que se desea contar

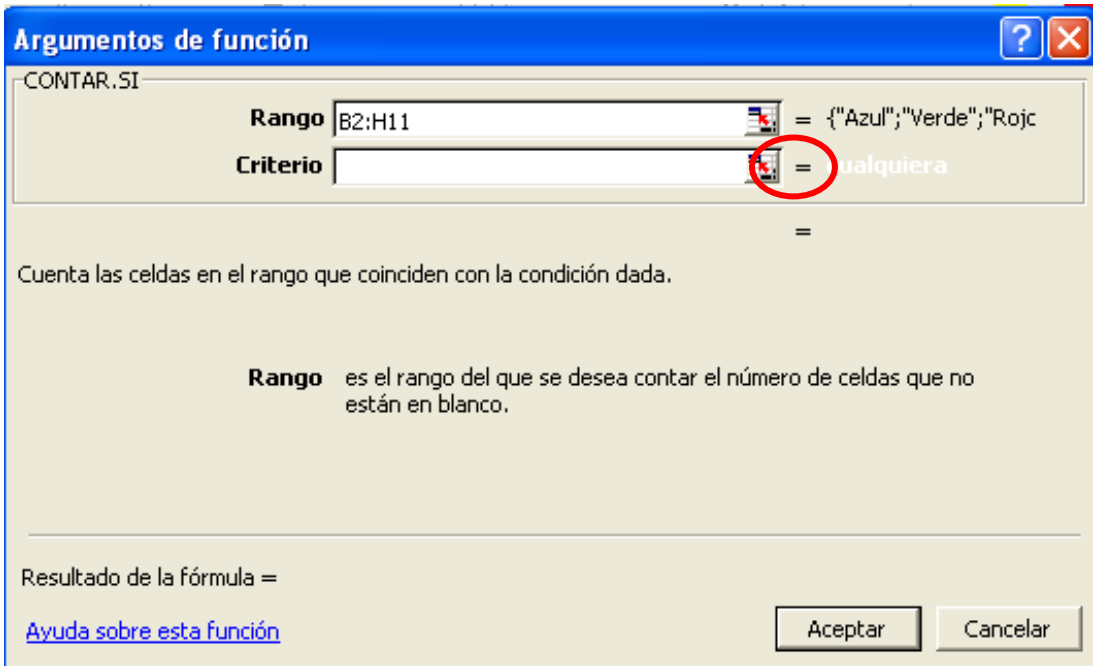


Pulsemos en el botón designado para el Rango, el cual reduce la ventana de la función. Luego con ayuda del Mouse, seleccionamos los datos iniciales del ejemplo (B2:H11):

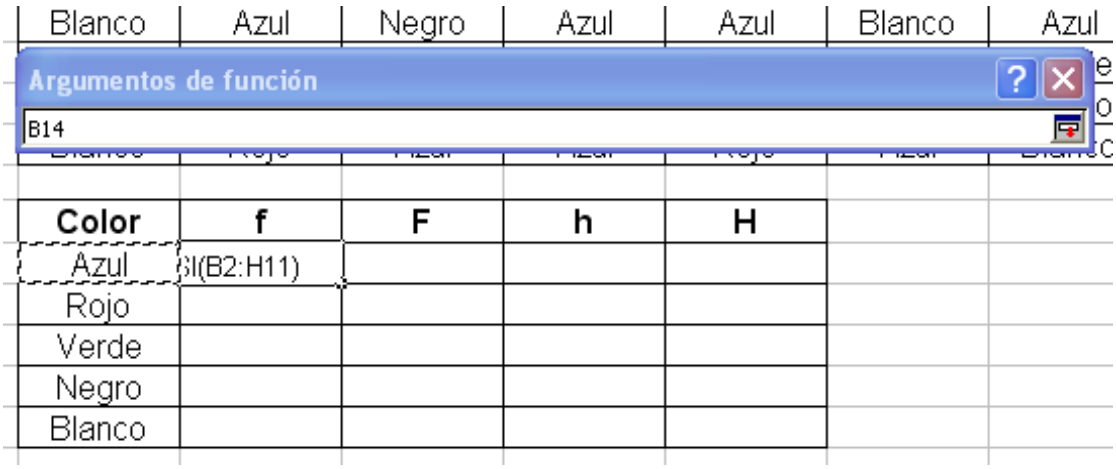
Azul	Verde	Rojo	Blanco	Rojo	Verde	Rojo
Rojo	Azul	Azul	Blanco	Verde	Azul	Rojo
Verde	Azul	Azul	Rojo	Verde	Azul	Rojo
Verde	Verde	Azul	Rojo	Negro	Negro	Negro
Azul	Verde	Verde	Verde	Blanco	Azul	Rojo
Negro	Negro	Azul	Negro	Blanco	Rojo	Azul
Blanco	Azul	Negro	Azul	Azul	Blanco	Azul
Rojo	Blanco	Rojo	Blanco	Rojo	Blanco	Verde
Blanco	Azul	Verde	Azul	Rojo	Verde	Negro
Blanco	Rojo	Azul	Azul	Rojo	Azul	Blanco

Pulsemos la tecla Enter o en el botón para regresar a la ventana de la función.

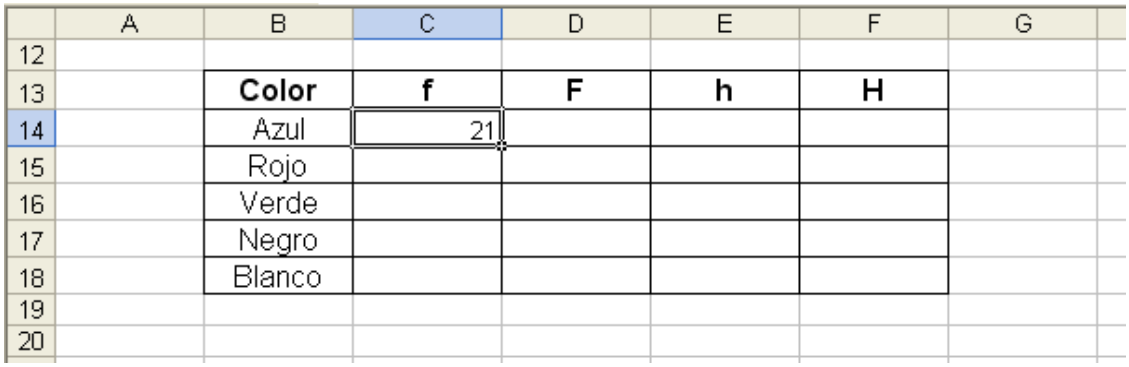
Luego seguimos los mismos pasos para el parámetro Criterio, teniendo en cuenta que solo seleccionaremos el color que deseamos contar dentro de los datos iniciales (el Azul):



Nos valdremos de la celda B13 en nuestra tabla de frecuencia donde se encuentran el listado de los colores:



Regresamos a la ventana de la función y pulsamos en el botón Aceptar:



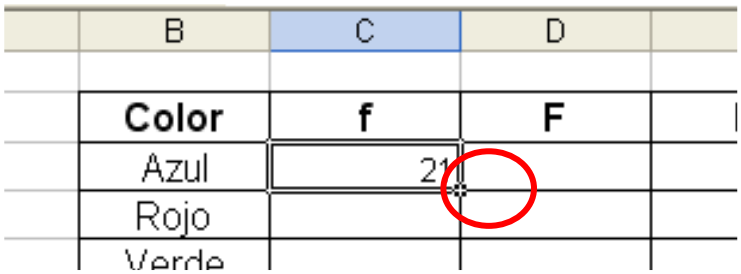
Debe aparecer el valor 21, indicando que existen 21 aciertos en los datos iniciales. Para calcular la frecuencia del color Rojo debemos seguir los mismos pasos que para el color Azul.

Para ahorrarnos este trabajo, Excel tiene la opción de arrastrar fórmulas a otras celdas. En nuestro caso arrastraríamos la fórmula a las celdas C15, C16, C17 y C18. Para que esto tenga efecto debemos ajustar la fórmula inicial:

El rango B2:H11 marca los datos iniciales, mientras que la celda B14 indica el color que se va a contar. La celda B14 debe variar para poder asignar los otros colores, mientras que el rango de datos iniciales se debe mantener fijo (los datos iniciales serán siempre los mismos). Ubiquemos el cursor en la Barra de Fórmulas al comienzo de la celda B2 y pulsemos la tecla F4; de igual horma hacemos esto para la celda H11. El resultado debería verse como sigue:

Lo que hicimos se conoce como ***Referencia Fija de Celda*** y permite que a la hora de arrastrar la fórmula, las celdas fijadas no varíen.

Para arrastrar la celda debemos ubicar el Mouse en la esquina inferior derecha de la celda C14 hasta que el puntero cambie a una cruz negra delgada. Justo en ese momento, pulsaremos el botón izquierdo del Mouse y arrastramos hacia la celda C18:



El resultado final debería verse como sigue:

	A	B	C	D	E	F	G
12							
13		**Color**	**f**	**F**	**h**	**H**	
14		Azul	21				
15		Rojo	16				
16		Verde	13				
17		Negro	8				
18		Blanco	12				
19							
20							

Recordemos que la sumatoria de las Frecuencias Absolutas equivale al tamaño del conjunto de los datos iniciales. Para verificar esto emplearemos una nueva función llamada **SUMA**.

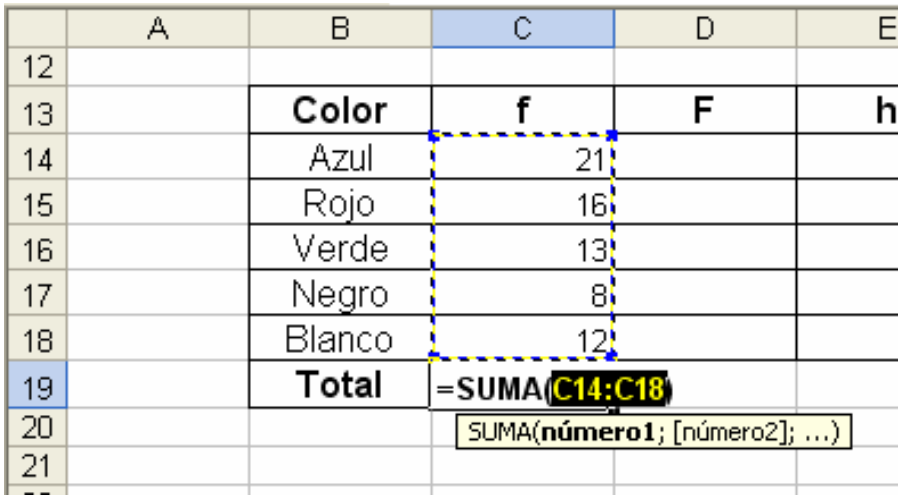
SUMA: Suma todos los números en un rango de celdas

Formato: SUMA(número1;número2;…)
Categoría: Matemáticas y Trigonométricas

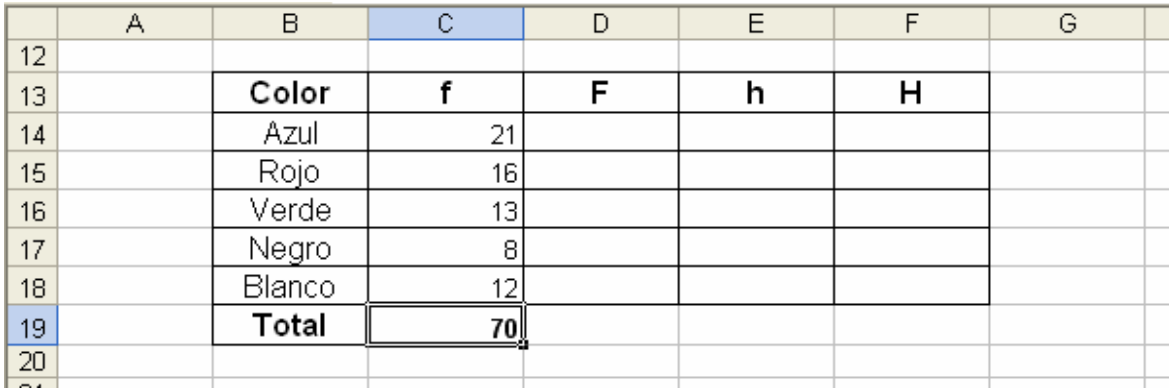
Ampliemos la tabla de frecuencia para incluir una fila al final cuyo encabezado sea Total:

	A	B	C	D	E	F	G
12							
13		**Color**	**f**	**F**	**h**	**H**	
14		Azul	21				
15		Rojo	16				
16		Verde	13				
17		Negro	8				
18		Blanco	12				
19		**Total**					
20							
21							

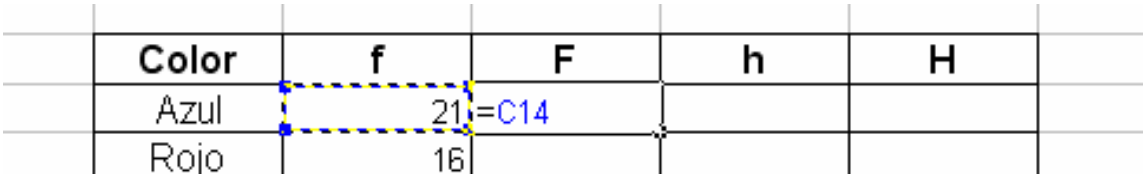
En la celda C19, ubiquemos la función SUMA de la forma explicada en la función anterior o pulsando sobre el botón Σ. De forma automática se seleccionará el rango de celdas C14:C18.



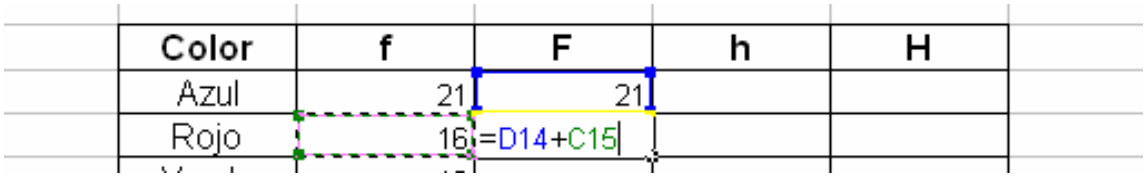
Pulsemos la tecla ENTER para ver el resultado de la sumatoria:



La primera Frecuencia Absoluta Acumulada será igual a la primera Frecuencia Absoluta (21). En la celda D14 colocaremos el signo igual y luego pulsaremos sobre la celda C14 para trasladar su valor (recuerde pulsar la tecla Enter):



La Frecuencia Absoluta Acumulada para el color Rojo (D15) equivale a la Frecuencia Absoluta Acumulada del color Azul (D14) más la Frecuencia Absoluta del color Rojo (C15). En Excel se vería como sigue:



Para calcular el resto de Frecuencias, arrastraremos la fórmula que esta en D15 hasta la celda D18. El resultado final se muestra a continuación:

	A	B	C	D	E	F	G
12							
13		**Color**	**f**	**F**	**h**	**H**	
14		Azul	21	21			
15		Rojo	16	37			
16		Verde	13	50			
17		Negro	8	58			
18		Blanco	12	70			
19		**Total**	**70**				
20							

Para calcular las Frecuencias Relativas (h) tomaremos cada Frecuencia Absoluta y la dividiremos sobre el total de datos (C19).

	A	B	C	D	E	F	G
12							
13		**Color**	**f**	**F**	**h**	**H**	
14		Azul	21	21	=C14/C19		
15		Rojo	16	37			

Nótese que para poder arrastrar la fórmula debemos fijar primero la celda C19 (el total no varía).

fx =C14/C19

El calculo de la Frecuencia Relativa Acumulada (H) lo haremos de forma similar que el calculo de la Frecuencia Absoluta Acumulada (F). La tabla final de frecuencia es (se cambiaron los formatos):

	A	B	C	D	E	F	G
12							
13		**Color**	**f**	**F**	**h**	**H**	
14		Azul	21	21	30,0%	30,0%	
15		Rojo	16	37	22,9%	52,9%	
16		Verde	13	50	18,6%	71,4%	
17		Negro	8	58	11,4%	82,9%	
18		Blanco	12	70	17,1%	100,0%	
19		**Total**	**70**		**100,0%**		
20							

2.2 CONSTRUCCIÓN Y CARACTERÍSTICAS DE LAS TABLAS TIPO B

Este tipo de tablas suelen ser utilizadas cuando el número de resultados posibles que puede obtener una variable son tan amplios, que una Tabla Tipo A haría muy poco en resumirlos (estos datos representan un rango muy amplio).

Debido a esta cantidad de valores, será necesario agruparlos mediante intervalos (la estadística los llama "Intervalos de Clases").

Por ejemplo, en el caso de contar con una valoración del 1 al 100 (un rango equivalente a 99), una tabla de frecuencia Tipo A se encargaría de buscar cuantas veces se repite cada uno de los 99 posibles resultados en un conjunto de datos, teniendo una función contraria a la de resumir los datos.

Agrupar los valores de la variable en intervalos podría simplificar estas fuentes de datos. Por ejemplo, podríamos hablar de las frecuencias para los valores comprendidos entre 0-20, 20-40, 40-60, 60-80 y 80-100.

En el intervalo 0-20 (que de ahora en adelante le llamaremos intervalo de clase), se sumaran las frecuencias de los datos cuyos resultados estén entre 0 y 20.

Intervalo de clase: Intervalos empleados en las Tablas de Frecuencias Estadísticas, capaz de contener diversas medidas de una variable. Consta de un límite inferior (Lm) y un límite superior (Ls).

Otro punto importante que el estadista debe definir, es la cantidad de intervalos de clase que empleará en la tabla de frecuencia. Esta cantidad de intervalos no deberían ser muchos, debido a que no se cumpliría el objetivo de resumir la información, y no tan pocos intervalos, ya que se perdería mucha información.

No existe una formula, ni unos principios únicos para establecer el numero de intervalos. Para nuestro libro, optaremos por manejar un número de intervalos convenientes entre 5 y 15.

Algunos autores han propuestos formulas que permiten ayudar en la tarea de conseguir el numero ideal de intervalos.

Numero de intervalos (Nc): Cantidad de intervalos con los cuales se compone una tabla de frecuencia.

La primera, la más conocida, establece el número de intervalos al obtener la raíz cuadrada del total de elementos considerados en el estudio.

$$Nc = \sqrt{n}$$ [1]

Cuando se trabajan con muestras mayores a 225, la formula obtiene un Nc superior a 15, por tanto, recomendaremos para estos casos la siguiente formula:

$$1 + 3{,}22 \log n$$

Si en ambas formulas obtenemos un Nc mayor a 15, simplemente tomaremos 15 intervalos. El estadista podrá omitir los resultados de las formulas y conseguirá seleccionar el numero de intervalos que crea son los mas adecuados, de acuerdo al objeto del estudio o las características que desea mostrar de la variable.

Cada intervalo posee un número máximo de resultados que puede agrupar. A este valor lo conoceremos como el "Ancho del Intervalo de Clase (A)".

Ancho del intervalo de Clase (A): Equivale a la diferencia entre el Limite superior (Ls) y el Limite inferior (Lm) de cada intervalo. Matemáticamente se expresa:

$$A = L_s - L_m$$

Su cálculo resulta de la división del Rango (R) entre el Número de Intervalos (Nc)

$$A = R / Nc$$

Hay que aclarar, que el ancho puede variar entre los intervalos, pero por razones estéticas, comprensión y para facilitar el análisis, se recomienda manejar un ancho común.

A continuación expondremos un ejemplo completo de tablas tipo B.

2.2.1 Ejemplo 1: tablas de frecuencia tipo B

Un sondeo realizado en la Universidad de Cartagena sobre 30 alumnos del sexto semestre de Administración Industrial, pretende mostrar que edad es la más representativa.

[1] En el caso de que hablemos de la población, reemplazaremos n por N.

Las edades de los alumnos fueron:

17	17	19	19	31
21	18	27	21	22
24	19	25	24	24
23	20	29	21	19
21	22	21	20	20
19	19	23	20	21

Elabore una tabla de frecuencia que resuma los resultados.

SOLUCIÓN

Antes de elaborar la tabla de frecuencia, debemos definir cual de los dos tipos propuestos es el que mejor se adapta (Tipo A y Tipo B).

Si resumimos los datos en una tabla tipo A, tendríamos una tabla muy extensa, en la cuales algunas frecuencias de las edades serian 0. Esto se debe a que el rango manejado es muy amplio (R = 31 - 17 = 14).

Edad	**f**
17	2
18	1
19	6
20	4
21	6
22	2
,,,	…
31	1
Total	**30**

En el caso de que queramos agrupar aun más estos datos, trabajaríamos con el concepto de intervalos de clase (Tabla Tipo B).

PASO 1: Determinar el numero de intervalos (Nc).

Optaremos por utilizar la primera formula expuesta:

$$Nc = \sqrt{n}$$

$$Nc = \sqrt{30} = 5{,}477 \cong 6 \text{ Intervalos}$$

Se debe siempre aproximar el número de intervalos al entero más próximo, recordando que este valor no será menor a 5, ni un valor mayor a 15. Nuestra tabla estará constituida por seis intervalos.

Paso 2: Determinar el ancho de cada intervalo.

Antes de hallar el ancho de los intervalos de clase, debemos calcular el rango (R) como primera medida.

$$R = 14$$

Con el Rango y el número de intervalos, podremos hallar el ancho:

$$A = \frac{R}{Nc} = \frac{14}{6}$$
$$A = 2{,}333$$

El ancho se debe ajustar para trabajar con el mismo número de decimales que en el conjunto de datos tratados. Como los datos son valores enteros (variable discreta), aproximamos al entero superior.

$$A \cong 3$$

El ajuste del Ancho no podrá ser menor al valor obtenido inicialmente.

Paso 3: Determinar el nuevo Rango (R').

En el momento de realizar el ajuste del ancho del intervalo, el rango se incrementa automáticamente. Este "Nuevo Rango" lo denotaremos como R':

$$R' = A \; x \; Nc$$
$$R' = 3 \; x \; 6 = 18$$

Nuevo Rango (R'): Rango que es convenido por el Ancho de los intervalos a los decimales que son manejados en los datos objeto del estudio. Su calculo se realiza multiplicando el Ancho ajustado por el Número de Intervalos:

$$R' = A \; x \; Nc$$

El rango se incremento en cuatro años. El incremento se le sumará al valor Máximo (Xmax') o restará al valor Mínimo (Xmin'). En este caso optaremos por aumentar el valor Máximo y reducir el valor Mínimo en dos.

$$\text{Incremento} = R' - R = 18 - 14 = 4$$
$$X_{\max}' = 31 + 2 = 33$$
$$X_{\min}' = 17 - 2 = 15$$

El alumno podrá repartir el incremento de la forma que crea más conveniente.

Este procedimiento permite encontrar los valores máximos y mínimos cuya resta sea igual al nuevo Rango (R')

$$R' = X_{\max}' - X_{\min}' = 18$$

Paso 4: Determinar los intervalos de clases iniciales.

Con los valores máximos y mínimos, y el ancho, podremos armar cada intervalo de clase. El primer intervalo parte del valor mínimo, al cual le agregamos el ancho.

Ni	Lm	Ls
1	15	18

El segundo intervalo parte del límite superior del intervalo anterior.

Ni	Lm	Ls
1	15	18
2	18	21

Seguimos realizando este proceso hasta alcanzar el valor máximo:

Ni	Lm	Ls
1	15	18
2	18	21
3	21	24
4	24	27
5	27	30
6	30	33

Esta primera distribución presenta algunos inconvenientes al momento de repartir las frecuencias a cada intervalo de clase, por ejemplo, existen 6 personas del total de encuestados que tienen una edad de 21 años, los cuales podrían ser clasificados en el intervalo dos o en el tres.

Ni	Lm	Ls
2	18	21
3	21	24

Este caso se le conoce como el "Problema de la Ambigüedad", y el cual debe ser solucionado antes de terminar la tabla de frecuencia tipo B.

2.2.2 El Problema de la Ambigüedad en las Tablas de Frecuencia Tipo B

Propondremos dos soluciones diferentes para resolver el problema de la ambigüedad.

2.2.2.1 Primera Solución

Se trabajan con intervalos cuyos límites Superiores e Inferiores tendrán un decimal adicional sobre el número de decimales manejados en los datos.

Por ejemplo, si el Limite Superior del primer intervalo es 21 y los datos trabajados son valores enteros, el nuevo límite superior será 21,1. Si los datos trabajan con un decimal, el nuevo Limite Superior sería 21,01.

El primer límite Inferior (Valor Mínimo) y el último límite Superior (Valor Máximo) se mantendrán sin modificación.

El problema quedaría solucionado de la siguiente manera:

Ni	Lm	Ls
2	18.1	21.1
3	21.1	24.1

Las seis personas que tienen 21 años quedarían registradas en el intervalo número 2.

2.2.2.2 Segunda Solución

Se convierten los Limites Superior e Inferior en Límites Abiertos y Cerrados. Se considera como Límite Abierto aquel que admite un número superior, más no igual, al valor indicado. El Límite Cerrado puede admitirse así mismo.

Los límites que son abiertos se identifican con el Paréntesis y los Límites Cerrados con el Corchete.

La solución a nuestro problema quedaría:

Ni	Lm	Ls
2	(18	21]
3	(21	24]

El valor 21 se ubica en el intervalo dos. Otra forma de colocar los intervalos es:

Ni	Lm	Ls
2	[18	21)
3	[21	24)

El valor 21 se ubica ahora en el intervalo número tres. Continuando con el ejemplo anterior:

Paso 5: Determinar los intervalos de clases reales.

Ni	Lm	Ls
1	15,0	18,1
2	18,1	21,1
3	21,1	24,1
4	24,1	27,1
5	27,1	30,1
6	30,1	33,0

Paso 6: Determinar las frecuencias absolutas, frecuencias relativas y marcas de clases.

Un valor representativo de los intervalos en las tablas de frecuencia son las "Marcas de Clase".

Marcas de Clase (Mc): Se define como el punto medio de un intervalo de clase.

$$Mc = \frac{L_s + L_m}{2}$$

Las marcas de clase son muy utilizadas en algunas gráficas estadísticas y en cálculos que serán vistos posteriormente.

Ni	Lm	Ls	F	F	H	H	MC
1	15,0	18,1	3	3	0,10	0,10	16,6
2	18,1	21,1	16	19	0,53	0,63	19,6
3	21,1	24,1	7	26	0,23	0,87	22,6
4	24,1	27,1	2	28	0,07	0,93	25,6
5	27,1	30,1	1	29	0,03	0,97	28,6
6	30,1	33,0	1	30	0,03	1,00	31,6
		Total	**30**		**1,00**		

2.2.3 Ejemplo 2: tablas de frecuencia tipo B

Crear una tabla tipo B que resuma los siguientes datos:

96,65	118,94	353,18	831,52	170,72	136,76
546,56	949,14	717,34	189,10	226,96	888,39
376,43	97,94	72,06	897,99	510,13	774,02
358,48	835,14	146,19	992,42	722,36	56,06
718,43	869,57	251,83	473,74	253,90	852,44
859,76	950,77	742,90	243,41	558,50	965,75
705,55	461,15	167,49	174,51	919,39	784,01
73,16	673,45	137,28	490,94	87,95	763,32
731,09	235,69	927,49	43,07	224,61	829,01

Paso 1: Determinar el número de intervalos (Nc).

Aplicamos la primera fórmula para determinar el número de intervalos de clase.

$$Nc = \sqrt{n}$$
$$Nc = \sqrt{54} = 7{,}348 \cong 8 \text{ Intervalos}$$

Paso 2: Determinar el ancho de cada intervalo.

Se determina el rango como primera medida.

$$X_{mzx} = 992{,}42$$
$$X_{\min} = 72{,}06$$
$$R = 992{,}42 - 72{,}06 = 920{,}36$$

Con el Rango y el número de intervalos, podremos hallar el ancho:

$$A = \frac{R}{Nc} = \frac{920{,}36}{8}$$

$$A = 115{,}045$$

El ancho se debe ajustar para trabajar con el mismo número de decimales que en el conjunto de datos tratados.

$$A \cong 115{,}05$$

Paso 3: Determinar el nuevo Rango (R').

Como el ancho fue ajustado, se procede a hallar el nuevo rango (R').

$$R' = A \; x \; Nc$$
$$R' = 115{,}05 \; x \; 8 = 920{,}40$$

El incremento entre el nuevo rango (R') y el rango inicial (R), se reparte entre el valor mínimo y el valor máximo

$$\text{Incremento} = R' - R = 920{,}40 - 920{,}36 = 0{,}04$$
$$X_{max}' = 992{,}42 + 0{,}02 = 992{,}44$$
$$X_{min}' = 72{,}06 - 0{,}02 = 72{,}04$$

Paso 4: Determinar los intervalos de clases iniciales.

Ni	Lm	Ls
1	72,04	187,09
2	187,09	302,14
3	302,14	417,19
4	417,19	532,24
5	532,24	647,29
6	647,29	762,34
7	762,34	877,39
8	877,39	992,44

Paso 5: Determinar los intervalos de clases reales.

Ni	Lm	Ls
1	72,040	187,091
2	187,091	302,141
3	302,141	417,191
4	417,191	532,241
5	532,241	647,291
6	647,291	762,341
7	762,341	877,391
8	877,391	992,440

Paso 6: Determinar las frecuencias absolutas, frecuencias relativas y marcas de clases.

Ni	Lm	Ls	f	F	h	H	MC
1	72,040	187,091	14	14	0,26	0,26	129,57
2	187,091	302,141	7	21	0,13	0,39	244,62
3	302,141	417,191	3	24	0,06	0,44	359,67
4	417,191	532,241	4	28	0,07	0,52	474,72
5	532,241	647,291	2	30	0,04	0,56	589,77
6	647,291	762,341	7	37	0,13	0,69	704,82
7	762,341	877,391	9	46	0,17	0,85	819,87
8	877,391	992,440	8	54	0,15	1,00	934,92
		Total	54		1,00		

2.2.4 Características de las tablas tipo B

- El número de posibles valores que toma la variable es elevado. (Rango grande).
- Se utiliza para el tratamiento de variables cuantitativas (discretas y continuas).
- Su construcción es más compleja que en las tablas tipo A.
- La interpretación equivale a especificar la frecuencia de cada intervalo de clase.
- Presenta un componente adicional: las marcas de clase.

2.2.4 Construcción de las tablas tipo B en Excel

Desarrollemos los seis pasos para la construcción de tablas tipo B en Excel a partir del siguiente conjunto de datos (digítelos a partir de la celda B2):

30	32	17	28	79	20	5
8	69	1	72	24	76	29
47	11	22	67	60	4	39
71	19	27	64	28	56	41
70	4	44	24	62	65	30
76	3	28	78	6	78	79
1	13	29	64	16	37	3

Paso 1: Determinar el número de intervalos (Nc).

El número de intervalos depende del tamaño de la muestra o población de datos. Para obtener esta información utilizaremos una nueva función llamada CONTAR:

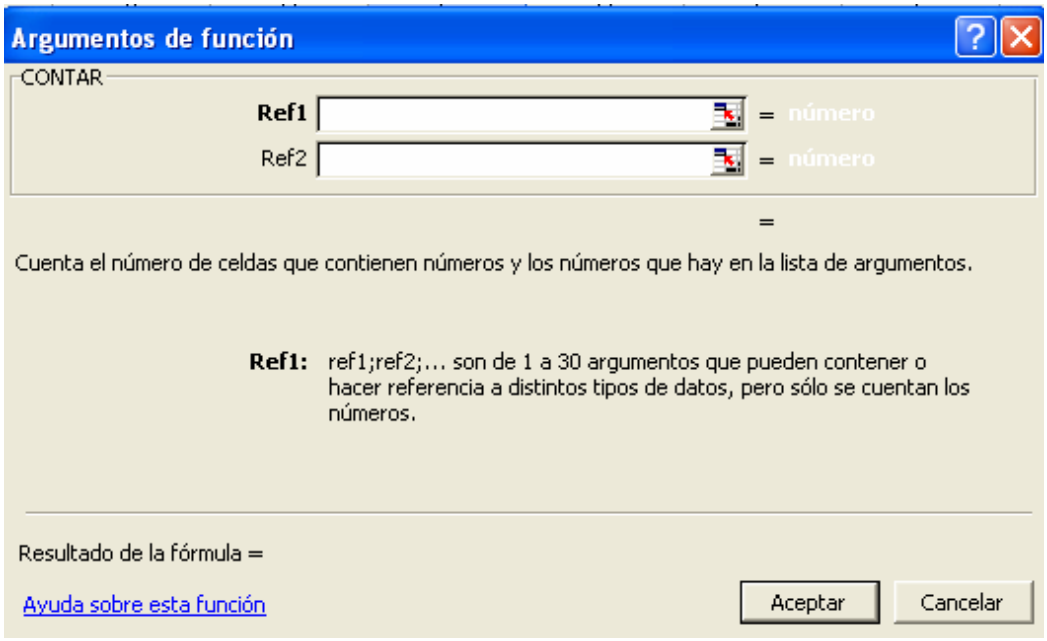
CONTAR: Cuenta un conjunto de celdas que posean números en su contenido

Formato: CONTAR(ref1;ref2;…)
Categoría: Estadísticas

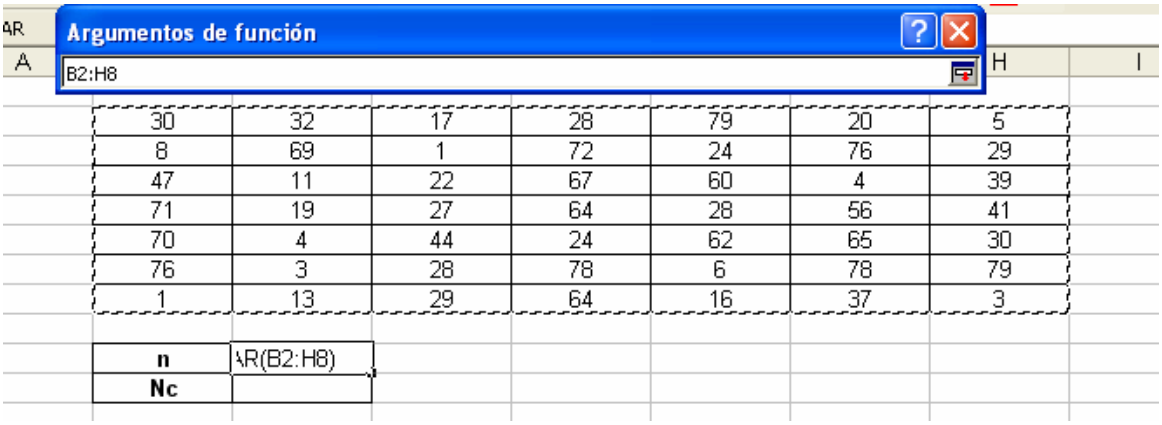
Para aplicar la fórmula, primero creemos una tabla resumen a partir de la celda B10, que empiece por la cantidad de datos (n) y el número de intervalos

	A	B	C	D	E	F	G	H	I
1									
2		30	32	17	28	79	20	5	
3		8	69	1	72	24	76	29	
4		47	11	22	67	60	4	39	
5		71	19	27	64	28	56	41	
6		70	4	44	24	62	65	30	
7		76	3	28	78	6	78	79	
8		1	13	29	64	16	37	3	
9									
10		**n**							
11		**Nc**							
12									

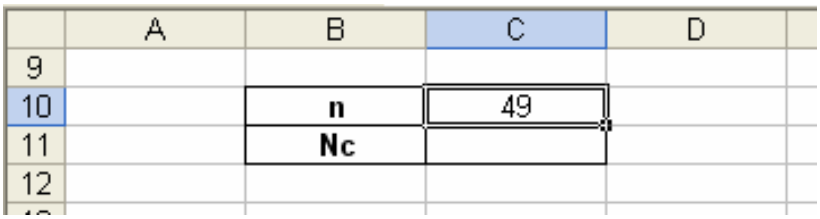
En la celda C10, ingresamos la función CONTAR:



En Ref1, seleccionaremos el rango de celdas equivalentes a la totalidad de los datos:



Al pulsar Enter y luego el botón Aceptar, tendremos como resultado el conteo de las celdas que tienen números (49 datos en total).



Aplicando la fórmula:

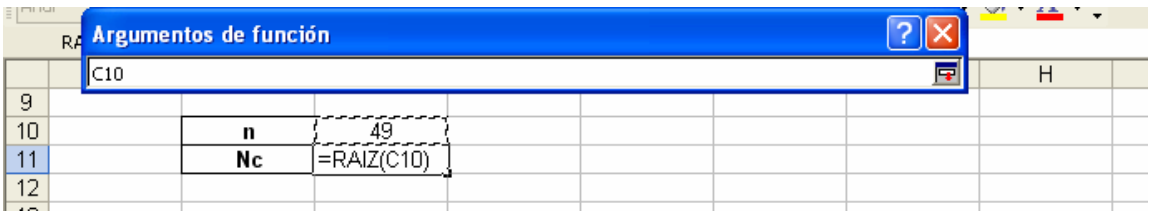
$$Nc = \sqrt{n}$$

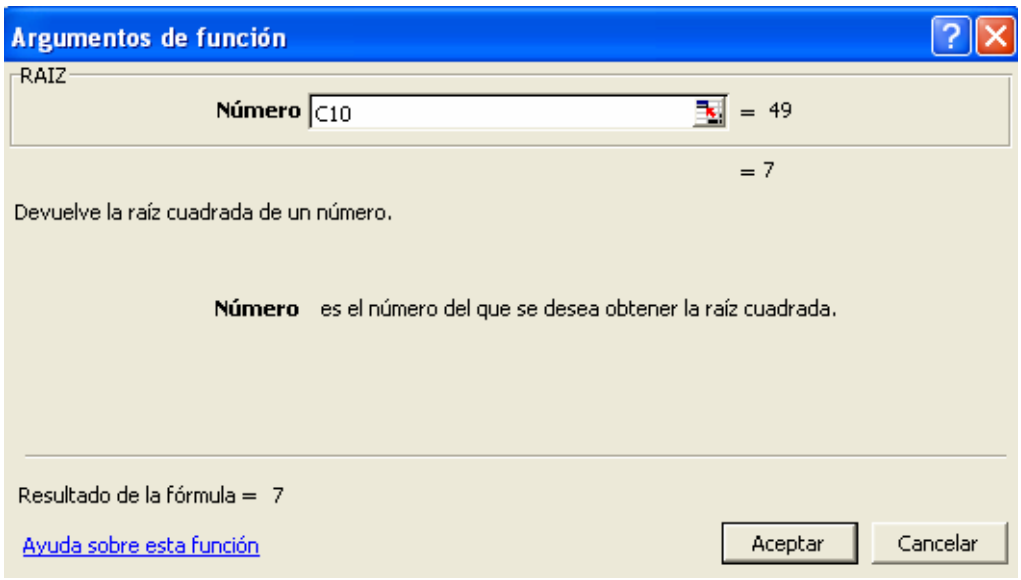
Obtendremos un número de intervalos. La raíz cuadrada se consigue con la fórmula RAIZ:

RAIZ: Calcula la raíz cuadrada de un número.

Formato: RAIZ(número)
Categoría: Matemáticas y Trigonométricas

Ubiquémonos en la celda C11 y activemos esta función. El parámetro número corresponde a la celda C10, cuyo valor es 49:





Al pulsar en Aceptar tendremos como resultado el valor 7, indicando que nuestra tabla tendrá 7 intervalos de clase.

Paso 2: Determinar el ancho de cada intervalo.

Aumentemos nuestra tabla resumen con cuatro nuevas filas: valor mínimo (Xmin), valor máximo (Xmax), rango (R) y ancho del intervalo de clase (A):

Para determinar el valor mínimo y máximo utilizaremos las dos siguientes fórmulas:

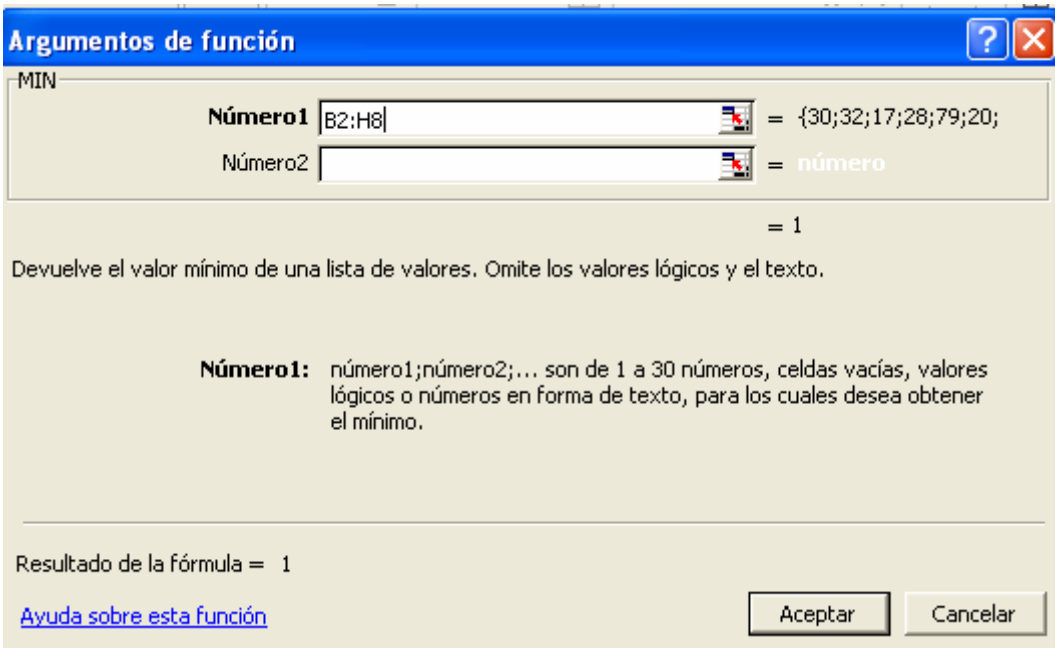
MIN: Localiza y muestra el valor mínimo de un conjunto de números.

Formato: MIN(número1;número2;…)
Categoría: Estadísticas

MAX: Localiza y muestra el valor máximo de un conjunto de números.

Formato: MAX(número1;número2;…)
Categoría: Estadísticas

En la casilla número1 de la función **MIN**, ubicaremos el rango correspondiente a los datos del ejercicio. Procedemos a hacer lo mismos con la función **MAX**.



La tabla resumen debe quedar:

	A	B	C	D
9				
10		**n**	49	
11		**Nc**	7	
12		**Xmin**	1	
13		**Xmax**	79	
14		**R**		
15		**A**		
16				

El rango se calcula con una simple fórmula dada por la resta de C13 y C12. El ancho equivale a la división del rango (C14) y el número de intervalos (C11).

	A	B	C	D	
9					
10		**n**	49		
11		**Nc**	7		
12		**Xmin**	1		
13		**Xmax**	79		
14		**R**	78	←	**=C13-C12**
15		**A**	11,1428571	←	**=C14/C11**
16					

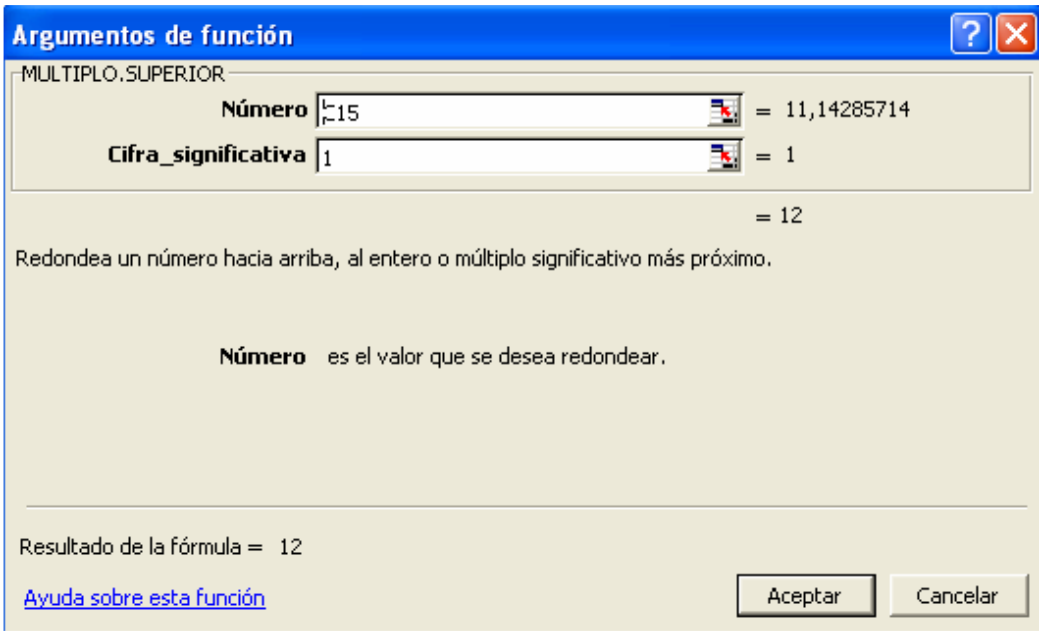
Paso 3: Determinar el nuevo Rango (R').

Ajustemos primero el ancho del intervalo a 12 (para trabajar con valores enteros) con la siguiente función:

MULTIPLO.SUPERIOR: Redondea un número hacia arriba.

Formato: MULTIPLO.SUPERIOR(número;cifra_significativa)
Categoría: Matemáticas y Trigonométricas

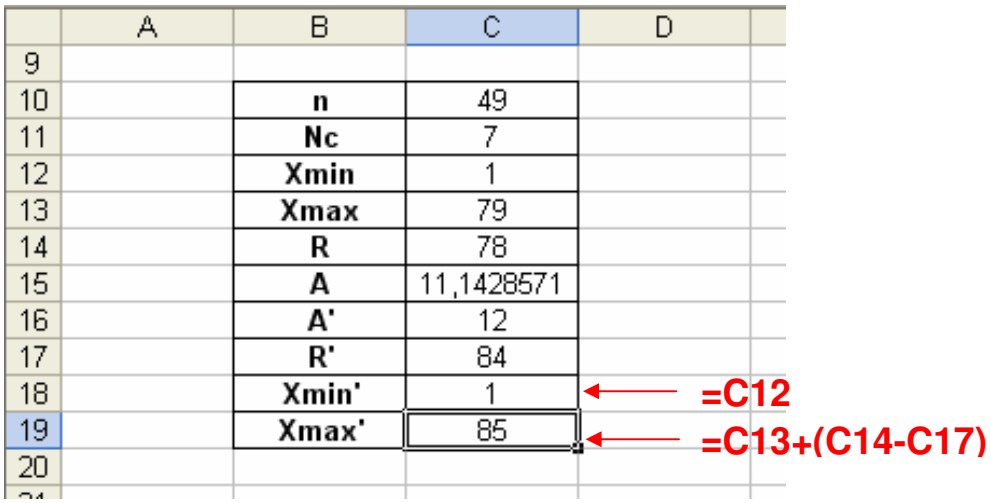
Con una nueva fila encabezada por A' (indica el ancho ajustado), activaremos esta función. En el parámetro número seleccionaremos la celda donde se encuentra el ancho sin ajustar (C15), y en cifra_significativa (equivale al múltiplo al que se desea redondear), el valor de 1.



El nuevo rango resulta de la multiplicación entre la celda C16 y C11:

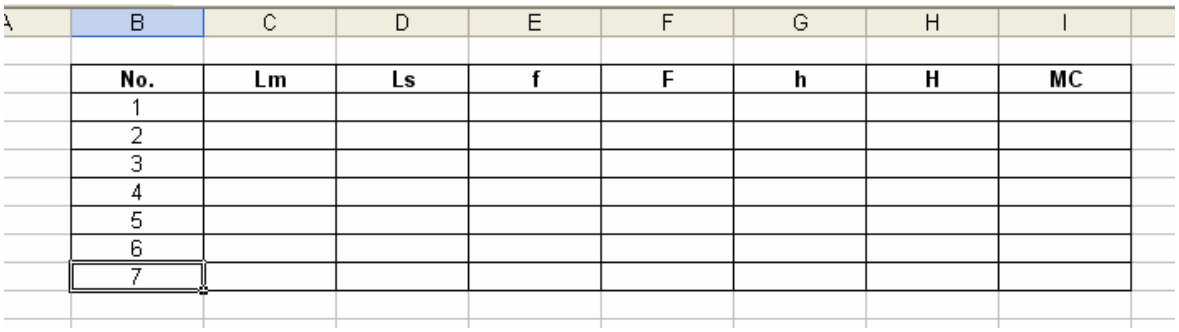
	A	B	C	D
9				
10		**n**	49	
11		**Nc**	7	
12		**Xmin**	1	
13		**Xmax**	79	
14		**R**	78	
15		**A**	11,1428571	
16		**A'**	12	
17		**R'**	84	
18				

Sumaremos la diferencia entre R' y R al valor máximo (para no afectar el valor mínimo):

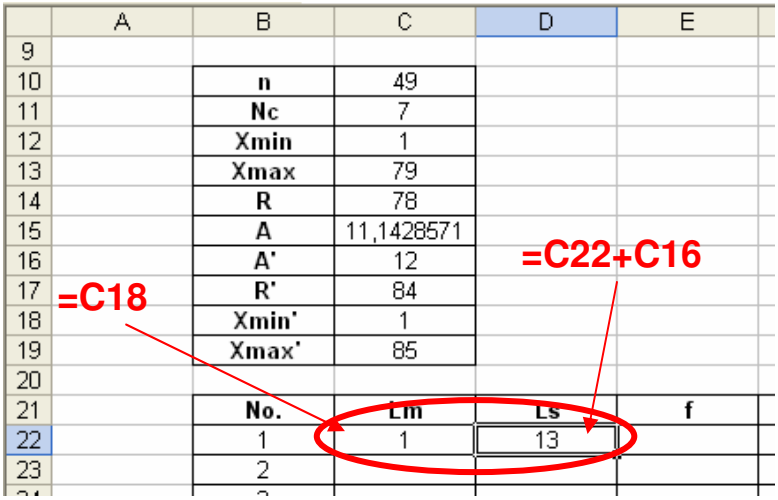


Paso 4 y paso 5: Determinar los intervalos de clases iniciales y reales.

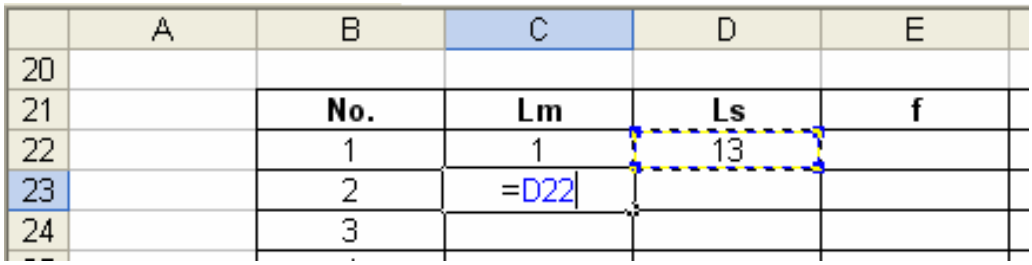
Construyamos la tabla de frecuencia para 7 intervalos de clase:



El límite inferior para el primer intervalo de clase es 1 (Xmin'), siendo su límite superior 13 (Xmin' más el ancho de clase).

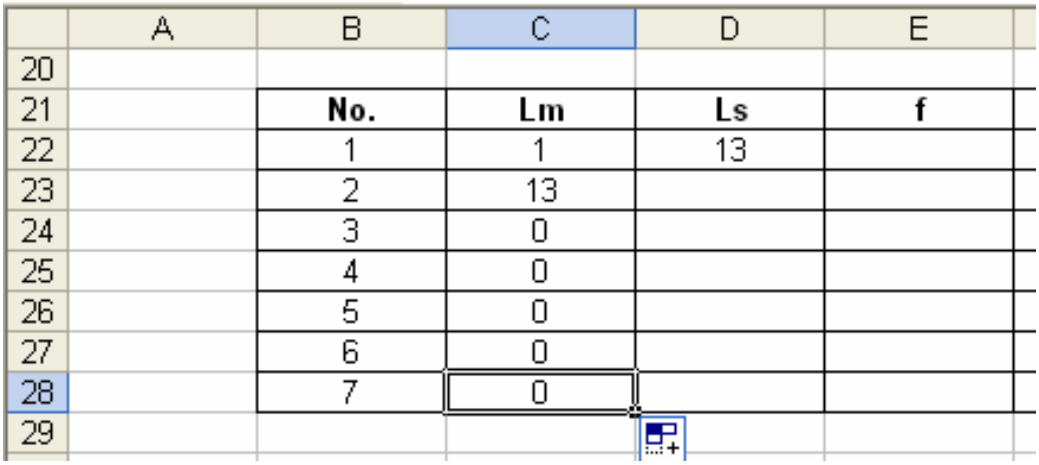


El límite inferior de las siguientes clases es igual al límite superior de su clase anterior:



	A	B	C	D	E
20					
21		**No.**	**Lm**	**Ls**	**f**
22		1	1	13	
23		2	=D22		
24		3			

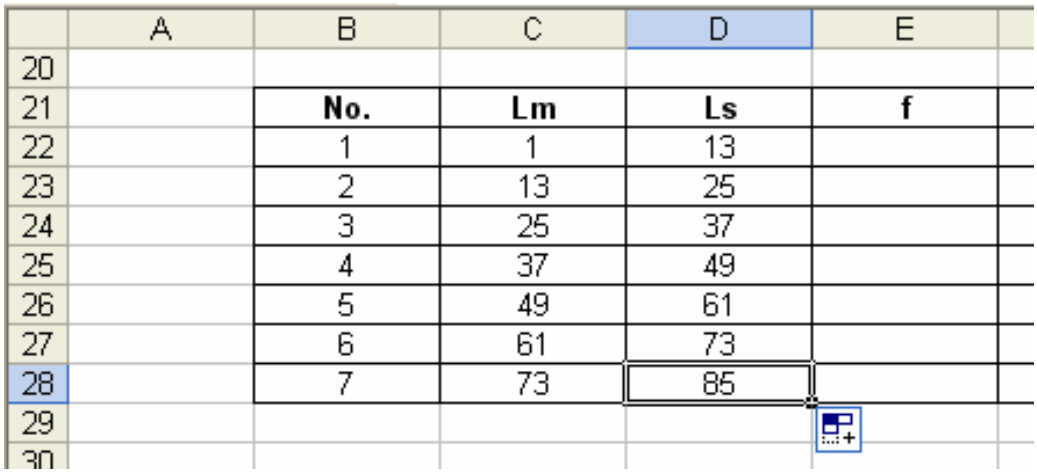
Podremos arrastrar esta fórmula hasta el último intervalo de clase (C28):



	A	B	C	D	E
20					
21		**No.**	**Lm**	**Ls**	**f**
22		1	1	13	
23		2	13		
24		3	0		
25		4	0		
26		5	0		
27		6	0		
28		7	0		
29					

Para calcular los límites superior bastará con arrastrar la fórmula que esta en la celda D22, fijando de antemano, la celda C16 que hace referencia al tamaño de la clase:

El resultado de los intervalos iniciales es el siguiente:



	A	B	C	D	E
20					
21		**No.**	**Lm**	**Ls**	**f**
22		1	1	13	
23		2	13	25	
24		3	25	37	
25		4	37	49	
26		5	49	61	
27		6	61	73	
28		7	73	85	
29					
30					

El problema de la ambigüedad se corregirá agregando al primer límite superior un valor de 0,1 y restando al último este mismo valor:

	A	B	C	D	E
20					
21		**No.**	**Lm**	**Ls**	**f**
22		1	1,0	13,1	
23		2	13,1	25,1	
24		3	25,1	37,1	
25		4	37,1	49,1	
26		5	49,1	61,1	
27		6	61,1	73,1	
28		7	73,1	85,0	
29					

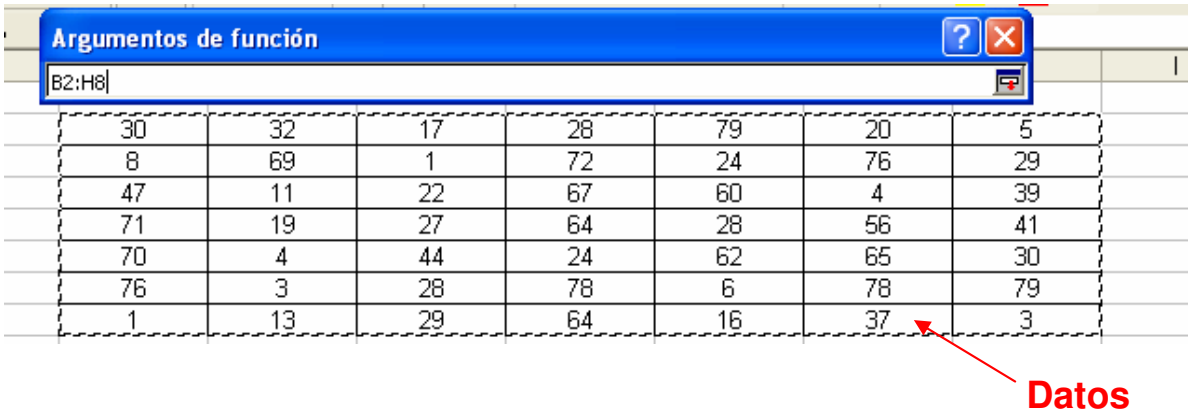
Paso 6: Determinar las frecuencias absolutas, frecuencias relativas y marcas de clases.

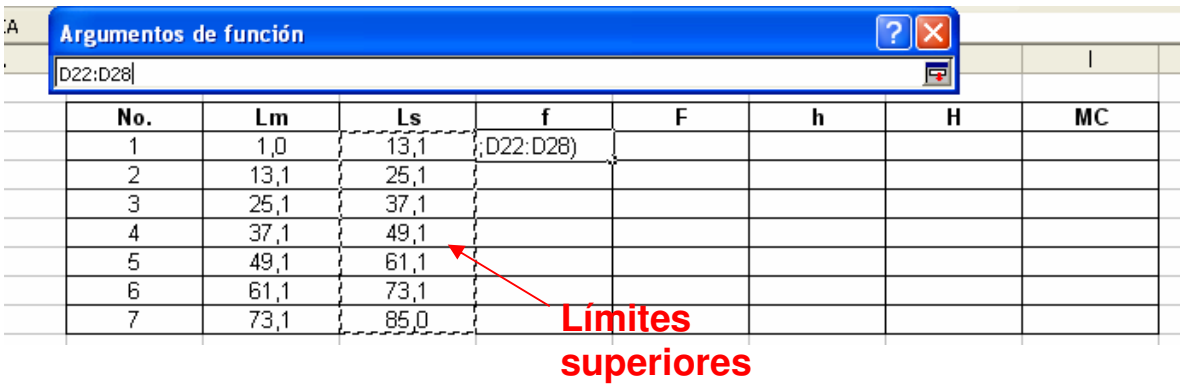
Comencemos con la frecuencia absoluta (f). Para trabajar con intervalos emplearemos una nueva función:

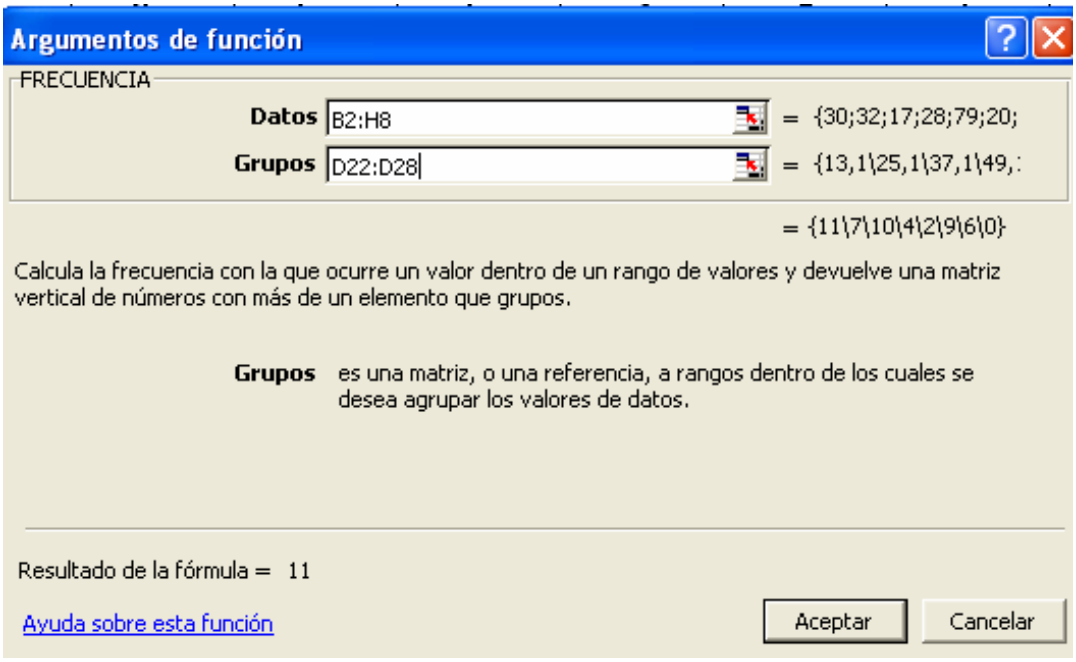
FRECUENCIA: Muestra el número de veces que se repite un número .dentro de un rango de celdas.

Formato: FRECUENCIA(datos;grupos)
Categoría: Estadísticas

A partir de la celda E22, activamos la función FRECUENCIA mostrando una ventana que pide dos requisitos: datos, que equivale al rango de celda donde está los datos iniciales y grupos, correspondiente en nuestro caso, a los límites superiores de la tabla de frecuencia.







Al pulsar en Aceptar, Excel mostrará la frecuencia para el primer intervalo de clase:

	A	B	C	D	E
20					
21		**No.**	**Lm**	**Ls**	**f**
22		1	1,0	13,1	11
23		2	13,1	25,1	
24		3	25,1	37,1	
25		4	37,1	49,1	
26		5	49,1	61,1	
27		6	61,1	73,1	
28		7	73,1	85,0	
29					

Para el cálculo de las frecuencias restantes deberemos seguir los siguientes pasos (dado que es una fórmula matricial):

- Ubicados desde la celda E22, seleccione las celdas consecutivas hasta C28.
- Pulse la tecla F2.
- Luego pulse de forma conjunta las teclas Control + Mayúsculas + Enter.

El resultado final es:

	A	B	C	D	E	F
20						
21		**No.**	**Lm**	**Ls**	**f**	**F**
22		1	1,0	13,1	11	
23		2	13,1	25,1	7	
24		3	25,1	37,1	10	
25		4	37,1	49,1	4	
26		5	49,1	61,1	2	
27		6	61,1	73,1	9	
28		7	73,1	85,0	6	
29						

Determine el resto de las frecuencias empleando los mismos pasos vistos para las tablas tipo A. El cálculo de las marcas de clase se hace tomando la suma de los dos límites dividido entre dos. La primera marca de clase es el resultado de:

La tabla definitiva (con algunos cambios en el formato) es:

	A	B	C	D	E	F	G	H	I
20									
21		**No.**	**Lm**	**Ls**	**f**	**F**	**h**	**H**	**MC**
22		1	1,0	13,1	11	11	22,4%	22,4%	7,05
23		2	13,1	25,1	7	18	14,3%	36,7%	19,1
24		3	25,1	37,1	10	28	20,4%	57,1%	31,1
25		4	37,1	49,1	4	32	8,2%	65,3%	43,1
26		5	49,1	61,1	2	34	4,1%	69,4%	55,1
27		6	61,1	73,1	9	43	18,4%	87,8%	67,1
28		7	73,1	85,0	6	49	12,2%	100,0%	79,05
29		**TOTAL**			**49**		**100,0%**		
30									

2.3 EJERCICIOS PROPUESTOS

2.3.1 Realice una tabla de frecuencia que resuma los siguientes datos:

1, 6, 8, 4, 5, 3, 4, 1, 1, 5, 3, 8, 7, 4, 6, 2, 8, 9, 3, 4, 10, 2.

a. ¿Cuál es el dato que mas se repite?
b. ¿Cuál es el dato que menos se repite?
c. ¿Cuál es el Rango?
d. ¿Qué tipo de tabla sería la más conveniente para agrupar estos datos?

2.3.2 Crear una tabla de frecuencia que permita agrupar los siguientes datos cualitativos

Rojo	Verde	Azul	Verde
Negro	Amarillo	Azul	Rojo
Rojo	Verde	Negro	Azul
Blanco	Negro	Verde	Rojo
Negro	Rojo	Blanco	Azul
Rojo	Verde	Verde	Negro

2.3.3 Un grupo de personas valora la gestión del departamento de servicio al cliente de un supermercado catalogándolo como: Excelente (E), Bueno (B), Regular (R) o Malo (M). Los resultados obtenidos son:

E	B	B	R	E
M	B	E	B	R
R	R	M	B	B
E	M	E	R	R
B	B	E	R	R
B	B	E	R	M
E	E	B	E	B
B	R	M	R	E

a) Elabore una tabla de frecuencia que permita resumir los datos
b) ¿Que porcentaje de personas valoró la Gestión del Departamento como Buena?
c) ¿Cuantas personas valoraron la gestión como Excelente y Buena?
d) Interprete f3, F3, h3 y H3.

2.3.4 Agrupe los siguientes datos en una Tabla de Frecuencia

23	21	43	41	19
29	17	33	35	30
25	11	28	40	22
45	43	23	29	32
9	47	47	31	12

2.3.5 Agrupe los siguientes datos en una tabla de frecuencia

11,3	14,2	21	20,5	29,9
31,2	33,7	22,5	27,6	20,3
29,4	31,4	21	12,1	30
29,9	15,6	32,2	43	17,7
27,6	22,5	41,1	19,1	13,6
47,3	11	15,6	33,3	15,4
38,1	35,3	39,8	30	15

2.3.6 Los siguientes datos representan el diámetro interno en cm. de 30 tubos para acueducto tomados como muestra dentro de un programa de calidad estatal.

14,1	14,2	13,9	14,7	12,9
15	14,1	14,5	14,9	13,6
14,5	14	15,1	14,7	13,8
14,2	14,2	14,7	13,9	13
14,6	14,1	14	14,8	14,7
15,2	13,5	14,2	14,8	14,5

a) Elabore una tabla de frecuencia que agrupe los datos. Justifique la elección del Tipo de Tabla usada.
b) Interprete F2, F4 - F2, F4 - f3, f3 + f2.
c) Interprete h1 + h2, H3 - H2, H3 - h2

2.3.7 Elabore una tabla de frecuencia que agrupe los siguientes datos.

200,23	145,81
178,15	133,9
149,11	211,64
176,59	124,45
194,58	144,32
157,21	174,38
121,04	193,2
139,45	201,55
174,73	147,83
230,99	212,71

2.3.8 A continuación se muestran los ingresos registrados en 50 familias seleccionadas al azar de estrato 3 en una importante ciudad:

$ 582.400,00	$ 841.210,00	$ 458.882,00	$ 535.211,00
$ 283.427,00	$ 433.792,00	$ 413.914,00	$ 485.925,00
$ 463.710,00	$ 848.607,00	$ 417.028,00	$ 550.409,00
$ 291.932,00	$ 538.597,00	$ 438.579,00	$ 223.878,00
$ 280.678,00	$ 947.218,00	$ 240.334,00	$ 391.814,00
$ 622.441,00	$ 781.633,00	$ 503.314,00	$ 700.010,00
$ 762.212,00	$ 681.517,00	$ 593.656,00	$ 221.135,00
$ 570.688,00	$ 382.734,00	$ 431.972,00	$ 718.487,00
$ 688.648,00	$ 775.123,00	$ 592.240,00	$ 317.555,00
$ 341.204,00	$ 1.147.607,00	$ 303.165,00	$ 716.003,00
$ 696.637,00	$ 926.773,00	$ 380.497,00	$ 647.222,00
$ 315.031,00	$ 584.599,00	$ 635.302,00	$ 345.931,00
$ 753.701,00	$ 500.558,00	$ 617.137,00	$ 285.715,00
$ 262.217,00	$ 1.115.432,00	$ 551.668,00	$ 698.338,00
$ 398.000,00	$ 655.230,00	$ 821.100,00	$ 293.312,00
$ 346.031,00	$ 415.560,00	$ 558.260,00	$ 306.300,00
$ 394.019,00	$ 347.485,00	$ 341.103,00	$ 400.496,00
$ 469.799,00	$ 773.411,00	$ 357.441,00	$ 192.019,00
$ 191.021,00	$ 810.474,00	$ 504.725,00	$ 489.025,00
$ 382.482,00	$ 353.289,00	$ 376.616,00	$ 480.314,00
$ 385.656,00	$ 705.757,00	$ 978.298,00	$ 272.328,00
$ 452.318,00	$ 736.707,00	$ 576.128,00	$ 310.504,00
$ 384.916,00	$ 584.196,00	$ 465.540,00	$ 828.898,00
$ 690.247,00	$ 920.824,00	$ 310.504,00	$ 480.314,00
$ 750.317,00	$ 1.115.851,00	$ 828.898,00	$ 272.328,00

Construya una tabla de frecuencia que agrupe los datos en 10 intervalos de clase.

2.3.9 A partir de la siguiente tabla de frecuencias con datos parciales:

Nc	Lm	Ls	f	F	h	H
1	(10	14]		10		
2	(14	18]		15		
3	(18	22]		31		
4	(22	26]		42		
5	(26	30]		55		
		TOTAL				

a) Calcule las frecuencias faltantes (f, h y H)
b) Modifique la tabla de frecuencia para que incluya los siguientes datos: 14, 22, 26, 27y 28.

2.3.10 Debido a un grave accidente, el gerente de una compañía consultora perdió información de un estudio de mercado que realizó a una importante compañía a nivel nacional de gaseosas. Solo se conoce algunos datos parciales sobre una entrevista que se elaboró a 150 personas.

Nc	Lm	Ls	f	F	h	H
1	0,0	2,1	24			
2	2,1	4,1			0,25	
3	4,1	6,1				0,71
4	6,1	8,1		123		
5	8,1	10,1	8			
6	10,1	12,1			0,11	
7	12,1	14,0				1,00
		TOTAL	**150**		**1,00**	

Reconstruya la tabla de frecuencia.

¿Cuantas personas toman menos de 4 gaseosas por semana?
¿Cuantas personas toman al menos 3 gaseosas por semana?

2.4 CASO: LA GESTIÓN DEL GOBIERNO

El alcalde de un pequeño pueblo, el Dr. Fernández, se siente preocupado por ciertos comentarios que rondan en la calle, en los cuales, lo critican de haberse desempeñado mal en el cargo. El asesor de imagen cree que estas conjeturas son falsas, y propone al Centro de Planeación que realice una encuesta sobre algunas familias (según el DANE, el pueblo cuenta con 1.500 familias) con el propósito de obtener cierta información de la gestión del gobierno actual.

La encuesta presentada por planeación fue la siguiente:

1. Califique de 1 a 5 la gestión del gobierno municipal (siendo 1 el menor valor y 5 el máximo).

_____

2. Marque con una X. En cuál de las siguientes áreas el gobierno presentó la mejor gestión:

a. Economía
b. Obras civiles
c. Servicios Públicos
d. Eventos culturales

3. ¿Aprobaría usted la reelección del alcalde?

Si _____ No_____

4. ¿Cuántos empleos cree usted que generó la Alcaldía municipal en el periodo actual de mandato?

_____

La información recolectada se muestra en la siguiente tabla:

Tarea

El Departamento de planeación lo contrata a usted para que resuma la anterior información en tablas de frecuencias, e interprete los resultados sobre las 30 familias encuestadas. Además, conteste las siguientes preguntas:

1. ¿Que tipo de variables puede identificar en la encuesta?

2. ¿Que tipo de tablas y por que, recomendaría utilizar para el resumen de los datos en cada pregunta?

3. ¿Defina cual es la población, la muestra y el fenómeno estudiado por el departamento de planeación?

4. ¿Que puede concluir de los resultados de la encuesta?

5. ¿Cree usted que la encuesta permite resolver todas las dudas sobre la gestión del gobierno del pueblo? Sustente.

Familia	Valoración	Áreas	Reelección	Empleos
1	3	Economía	Si	150
2	3	Eventos culturales	No	0
3	4	Eventos culturales	Si	200
4	5	Economía	Si	350
5	5	Servicios Públicos	Si	250
6	4	Eventos culturales	No	100
7	2	Economía	No	100
8	1	Servicios Públicos	No	20
9	5	Obras civiles	Si	240
10	1	Economía	No	25
11	3	Eventos culturales	No	100
12	4	Servicios Públicos	No	0
13	5	Economía	Si	250
14	5	Servicios Públicos	Si	450
15	4	Economía	Si	100
16	3	Servicios Públicos	No	0
17	5	Eventos culturales	Si	150
18	5	Economía	Si	50
19	3	Eventos culturales	No	20
20	4	Servicios Públicos	Si	300
21	2	Economía	No	120
22	2	Obras civiles	No	10
23	5	Eventos culturales	Si	410
24	1	Eventos culturales	No	50
25	2	Eventos culturales	No	0
26	5	Economía	Si	150
27	4	Eventos culturales	Si	20
28	4	Eventos culturales	Si	300
29	5	Economía	Si	400
30	3	Eventos culturales	No	250

2.5 CUESTIONARIO DE REPASO

Selección Múltiple con Única Respuesta: Marque con una X la respuesta correcta.

1. Las Tablas de Frecuencia Tipo A se caracterizan por:

A. Trabajan solo con datos cualitativos.
B. Agrupan datos cuyo Rango es bajo.
C. Agrupan datos cuyo Rango es alto.
D. Presentan más intervalos que en las Tablas Tipo B.

2. Las frecuencias relativas se diferencian de las frecuencias absolutas porque:

A. Las frecuencias relativas se establecen de acuerdo a una base.
B. Las frecuencias relativas se expresan como porcentaje.
C. La suma de las frecuencias relativas es igual a 1.
D. La A y B.

3. En una tabla de frecuencia, F4 - F2 es lo mismo que:

A. F4 - f3
B. F2 + f3
C. F3
D. f3 + f4

4. En una tabla de frecuencia, h3 + H2 es lo mismo que:

A. H3
B. H4
C. h3
D. h3 - h4

5. En el proceso de elaboración de las Tablas Tipo B, suele determinarse un nuevo Rango (R') para luego hacer los intervalos de clase. ¿Cuando es necesario hallar R'?

A. Cuando se desea agregar nuevos intervalos.
B. En el momento en que aproximamos el Ancho de los Intervalos.
C. Cuando se aumenta el tamaño de la muestra.
D. Ninguna de las anteriores.

CAPITULO 3

GRÁFICOS ESTADÍSTICOS

Este capitulo lo dedicaremos al estudio de los gráficos estadísticos más usados que parten de resúmenes o tablas de frecuencias.

La ventaja de los gráficos con respecto a las tablas estudiadas en el capitulo anterior, es que permite una fácil interpretación y análisis de los datos, al mostrar las frecuencias mediante símbolos, barras, polígonos y sectores.

3. GRÁFICOS ESTADÍSTICOS

Gráficos Estadísticos: Son representaciones visuales que emplean símbolos, barras, polígonos y sectores, de los datos contenidos en tablas de frecuencias.

Trataremos siete tipos de gráficos estadísticos:

- Gráfico de sectores
- Gráficos de columnas
- Histograma
- Polígonos de frecuencias
- Curvas suavizadas o curvas de frecuencias
- Ojivas
- Pictogramas

3.1 GRÁFICOS DE SECTORES

Este tipo de diagramas consideran una figura geométrica en que la distribución de frecuencias se reparte dentro de la figura como puede ser una dona, pastel, círculo o anillo, en el que cada porción dentro de la figura representa la información porcentual del total de datos.

3.1.1 Ejemplo de gráficos de sectores

Realizar un diagrama de sectores a partir de la siguiente tabla de frecuencia que resume las preferencias de un grupo de encuestados hacia cinco candidatos a elecciones locales:

Clase	Frecuencia
Candidato 1	25
Candidato 2	30
Candidato 3	45
Candidato 4	20
Candidato 5	20
Total	**140**

SOLUCIÓN

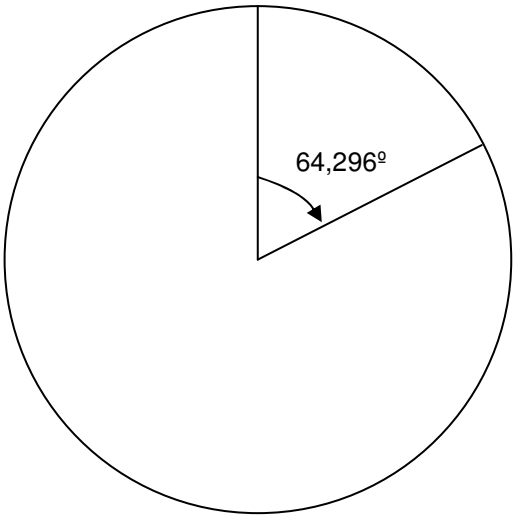
Para crear un gráfico de sectores, hay que tener en cuenta los siguientes pasos:

PASO 1: Determinar las frecuencias relativas para cada clase.

Clase	f	h
Candidato 1	25	0,1786
Candidato 2	30	0,2142
Candidato 3	45	0,3214
Candidato 4	20	0,1429
Candidato 5	20	0,1429
Total	**140**	**1,0000**

PASO 2: Determinar los ángulos que representan las porciones dentro de la figura para cada clase.

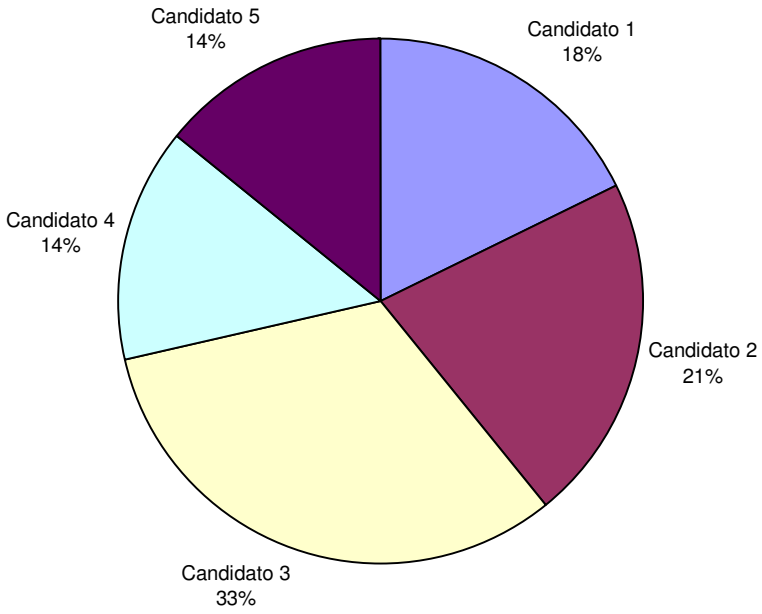
Un círculo esta formado por un ángulo de 360º. La porción correspondiente al Candidato 1 equivale a un 17,86% de esos 360º, es decir, 64,296º. Gráficamente tendríamos (se parte desde el eje vertical superior, y se comienza a graficar cada clase en sentido de las manecillas del reloj):



La tabla final con los ángulos repartidos para cada clase quedaría:

Clase	f	h	Ángulo
Candidato 1	25	0,1786	64,296º
Candidato 2	30	0,2142	77,112º
Candidato 3	45	0,3214	115,704º
Candidato 4	20	0,1429	51,444º
Candidato 5	20	0,1429	51,444º
Total	**140**	**1,0000**	**360º**

El gráfico definitivo se muestra a continuación (nótese que cada sector se identifica con un color diferente):



3.1.2 Características de los gráficos de sectores

- No muestran frecuencias acumuladas.
- Se prefiere para el tratamiento de datos cualitativos o cuasicualitativos.
- La mayor área (o porción de la figura) representa la mayor frecuencia.
- Son muy fáciles de elaborar.
- Suelen utilizarse para representar tablas tipo A.
- La figura completa equivale al 100% de los datos (360º).

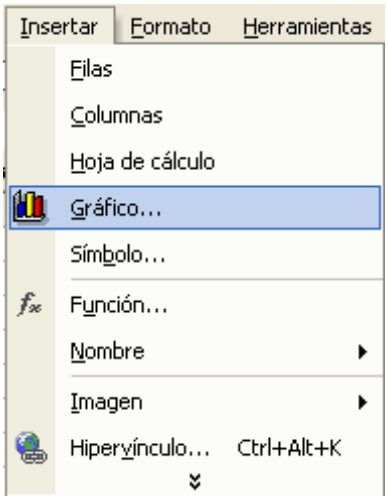
3.1.3 Construcción de los gráficos de sectores en Excel

Vamos a explicar la creación de gráfico de sectores a partir del ejemplo anterior. Los primero es trasladar la tabla de datos (a partir de la celda B2) en una hoja vacía:

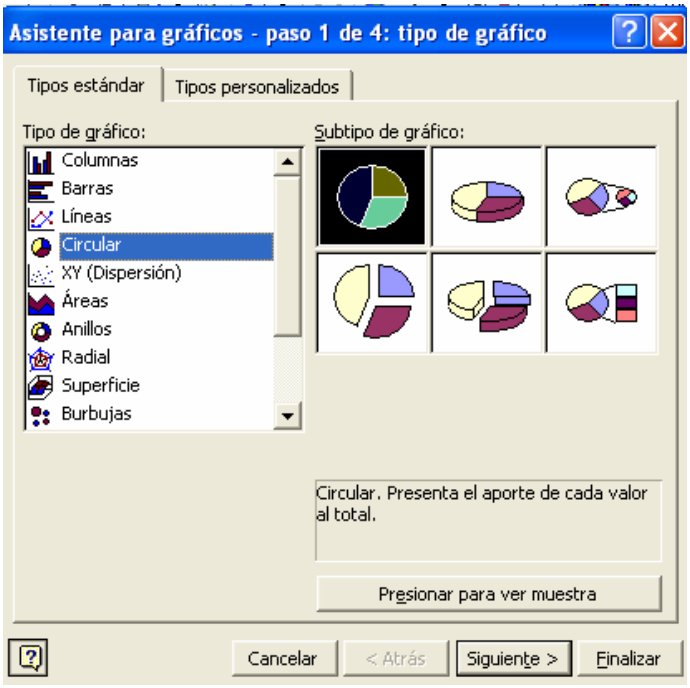
	A	B	C	D	E
1					
2		**Clase**	**f**	**h**	
3		Candidato 1	25	0,1786	
4		Candidato 2	30	0,2142	
5		Candidato 3	45	0,3214	
6		Candidato 4	20	0,1429	
7		Candidato 5	20	0,1429	
8		**Total**	**140**	**1,0000**	

Utilice la función SUMA para calcular los totales

En el caso de Excel, solo es necesario trabajar con una sola frecuencia. Ubiquémoslos en una celda fuera de la tabla que acabamos de crear y pulsemos el botón de Asistentes para Gráficos o elija la opción Gráfico en el menú Insertar. El asistente cuenta de cuatro pasos sencillos para la creación de varios tipos de gráficos

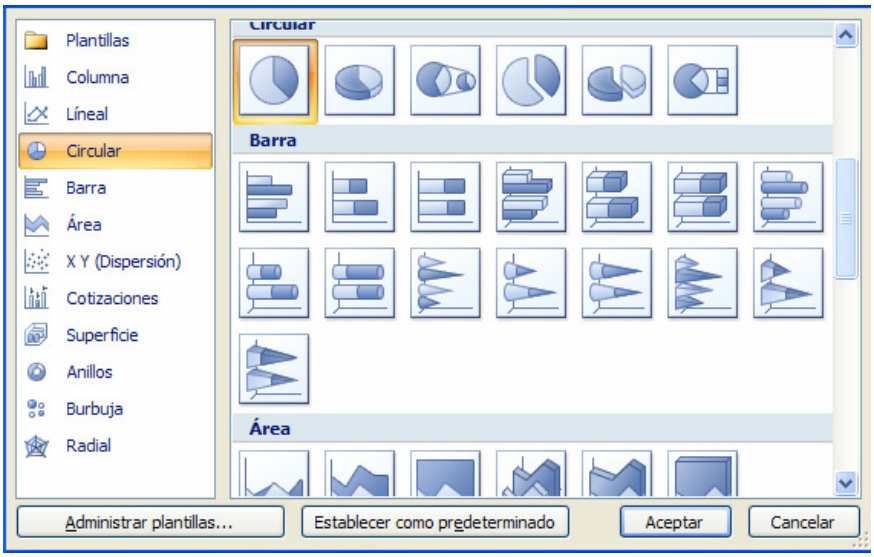


Aparecerá una nueva ventana con diferentes tipos de gráficos. Seleccionemos el correspondiente al gráfico de sectores (Circular en Excel). En el tenemos la opción de 6 gráficos de sectores.

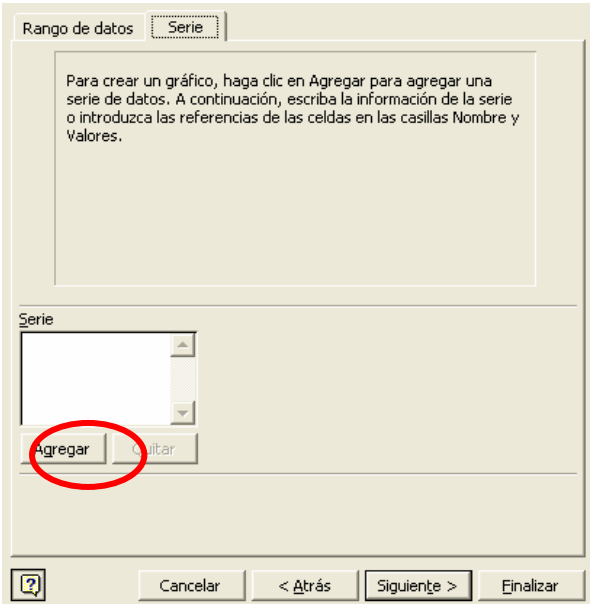


Para observar esta misma ventana en Microsoft Excel 2007 deberá situarse en la ficha **INSERTAR** y pulsar sobre el botón **GRÁFICO**S, luego pulsar el botón que amplía la ventana.

Se selecciona el tipo de gráfico **CIRCULAR**.



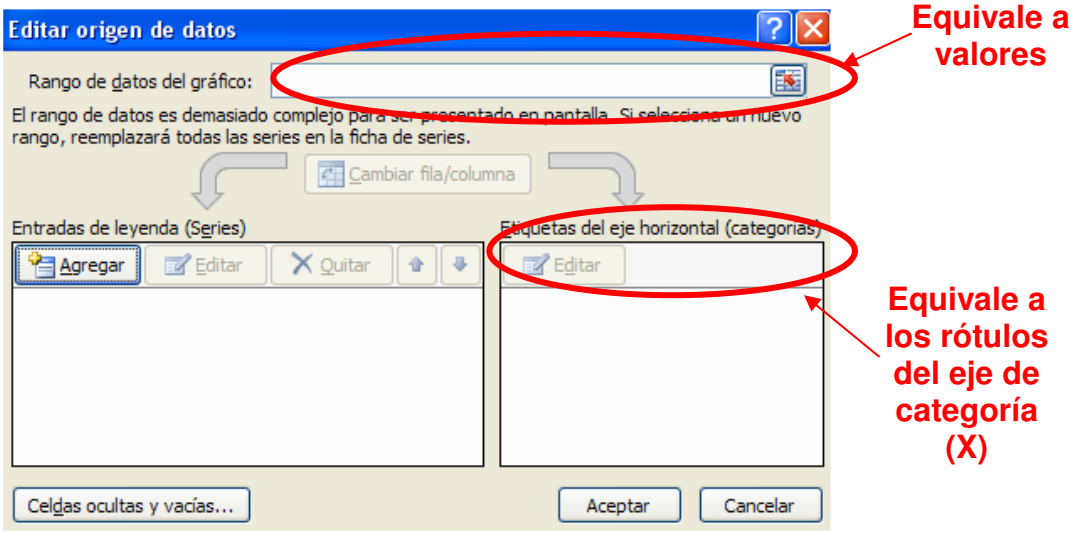
Al pulsar en el botón **SIGUIENTE**, pasaremos al pasos dos del asistente, en donde especificaremos los datos de origen para crear el gráfico. Pulsemos el botón **AGREGAR** que se encuentra en la ficha **SERIE**:



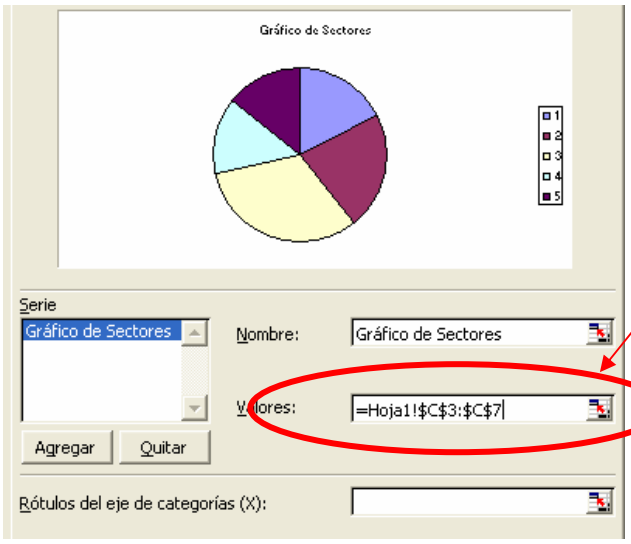
Este botón permite introducir distintas series de datos (provenientes de distintas tablas de frecuencias, variables o poblaciones). Excel pide tres campos para construir el gráfico:

- **Nombre:** Título o encabezado del gráfico.
- **Valores:** Las frecuencias que están en la tabla (puede ser relativa o absoluta, ambas no acumuladas).
- **Rótulos del eje de categorías (X):** Representa las clases de la tabla de frecuencias.

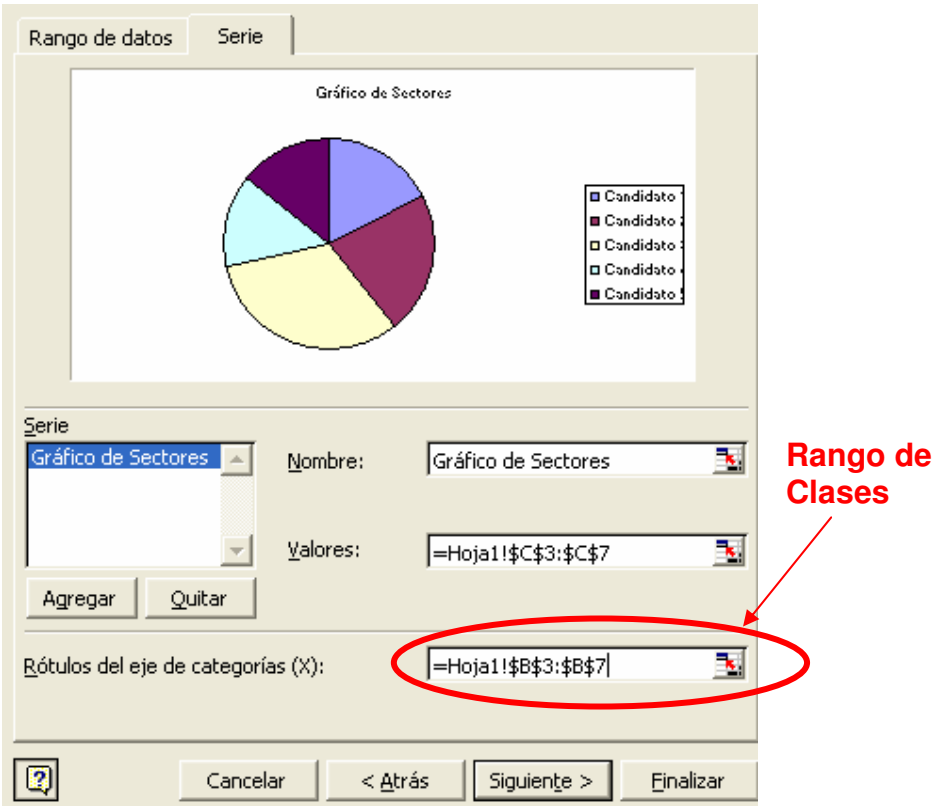
Este formato se incluye en la versión 2007 en la opción EDITAR DATOS DE ORIGEN ("Edit Data Source").



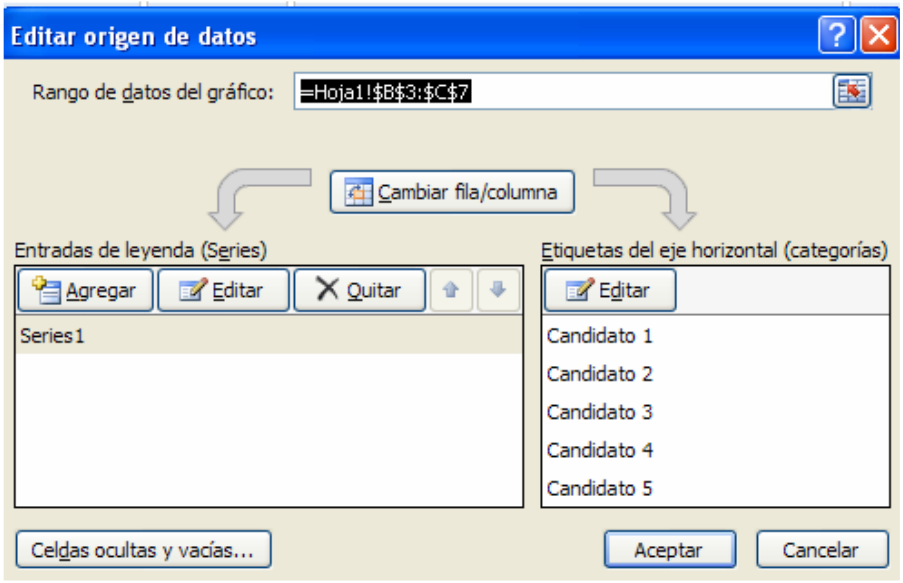
En el título escribamos "Gráfico de Sectores". En valores señalemos las frecuencias absolutas (f):



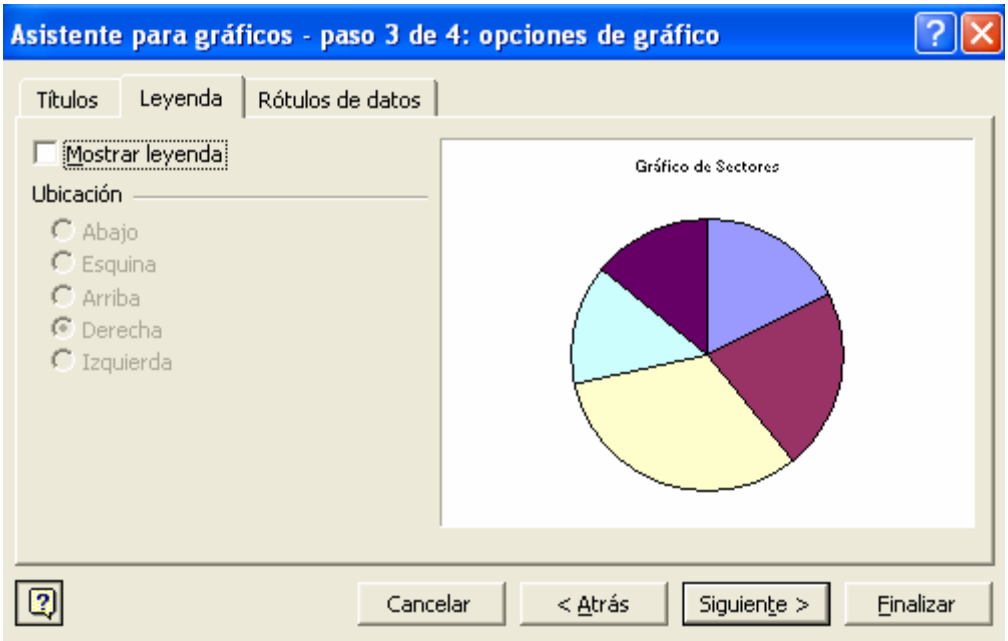
En la casilla rótulos del eje de categoría corresponde al rango de las celdas que muestran las clases:



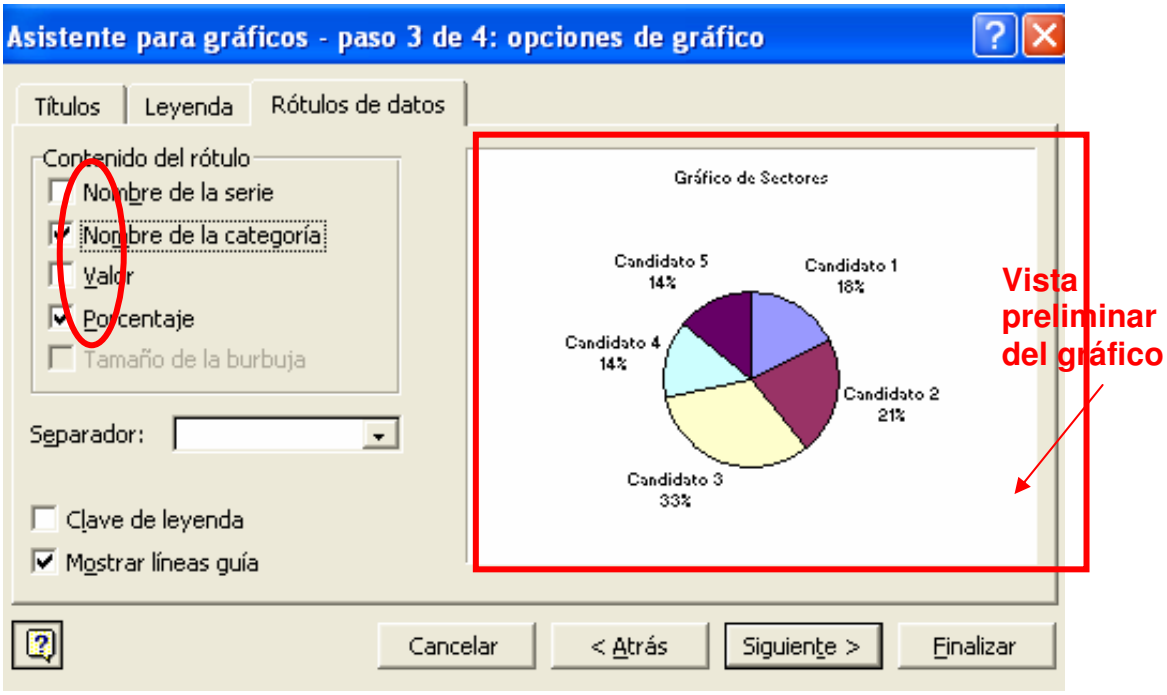
En MS Excel 2007 la ventana para introducir los datos de origen debe quedar como sigue:



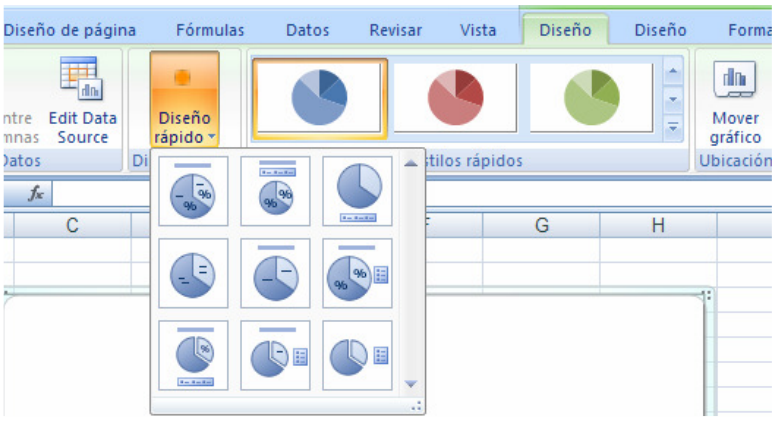
Pulse nuevamente siguiente para dirigirnos al tercer paso del asistente. En el podremos editar tres fichas: Títulos, Leyendas y Rótulos de datos. En la ficha Leyenda (intermedia) desactivemos la casilla Mostrar Leyenda para ampliar un poco el gráfico (es opcional).



En la ficha Rótulos de datos activaremos Nombre de la categoría y Porcentajes

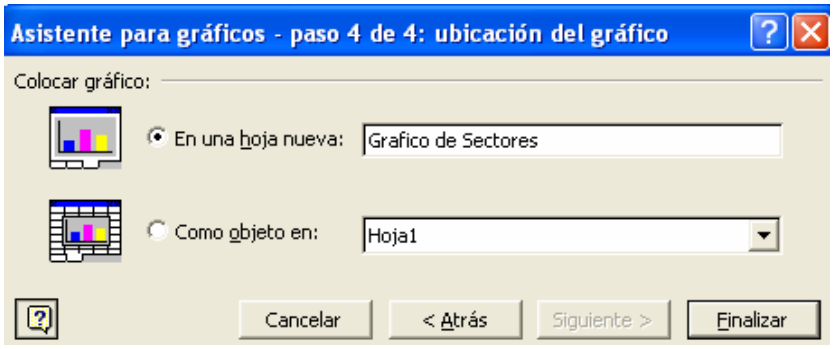


Microsoft Excel 2007 permite editar las opciones del gráfico directamente sobre el gráfico a partir de un DISEÑO RÁPIDO seleccionado.

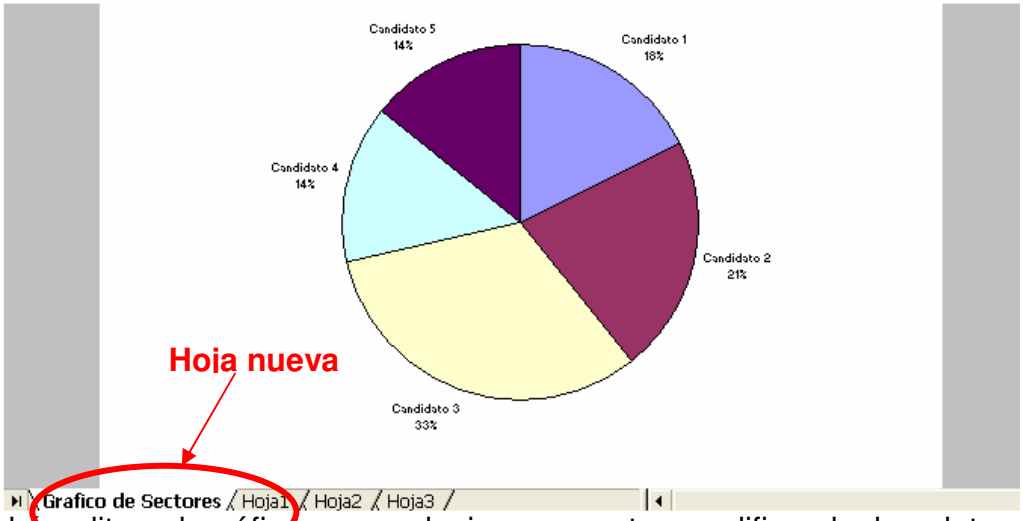


Por último (paso 4), esta la opción de visualización (este paso se aplica en la versión 2007 al presionar sobre el botón **MOVER GRÁFICO**):

- **En una hoja nueva:** El gráfico aparece en una nueva hoja, abarcando todo el espacio.
- **Como objeto en:** Se crea el gráfico como un objeto editable en una hoja existente.

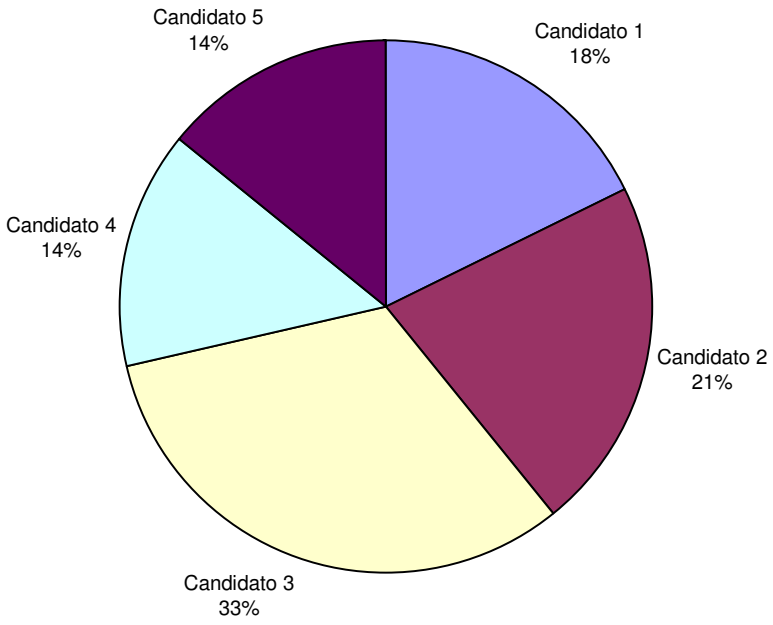


Optemos por activar en una hoja nueva, modificando el nombre de la hoja por "Gráfico de Sectores" y luego pulse en el botón Finalizar.



Podrá editar el gráfico en cualquier momento, modificando los datos de origen, formatos y tipo:

Gráfico de Sectores



3.2 GRÁFICOS DE COLUMNAS

Los gráficos de barras representan las frecuencias mediante columnas (o barras), a través de la altura de las mismas en un plano cartesiano.

3.2.1 Ejemplo de gráficos de columnas

Realizar un gráfico de barras a partir de la siguiente tabla de frecuencia:

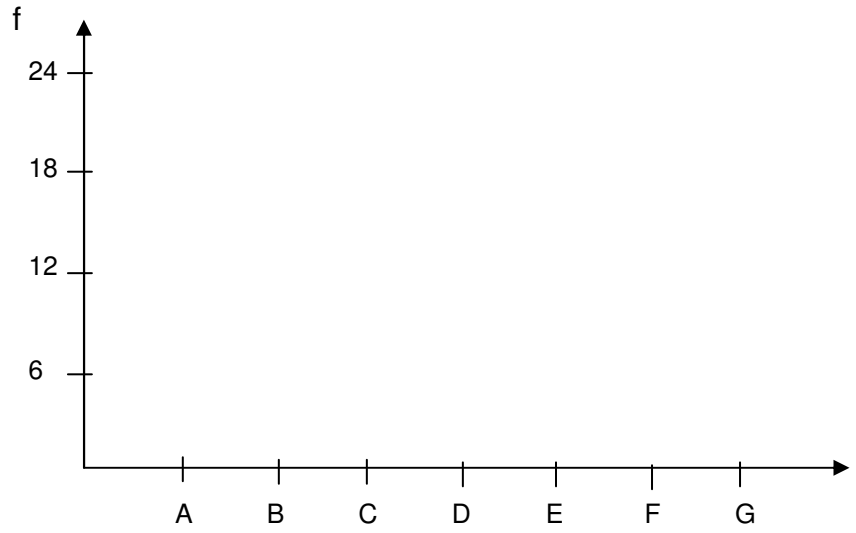
Clase	Frecuencia
A	5
B	11
C	11
D	4
E	15
F	18
G	24
Total	**88**

SOLUCIÓN

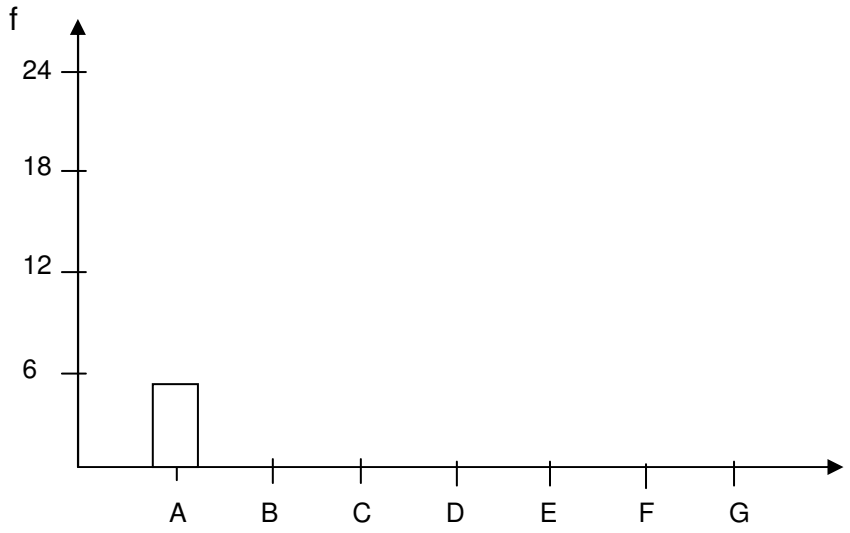
Para crear un gráfico de barras, seguiremos 2 sencillos pasos:

PASO 1: Representar las escalas en los ejes horizontal y vertical del primer cuadrante de un plano de cartesiano. En el eje vertical colocaremos las frecuencias y en el eje horizontal las clases.

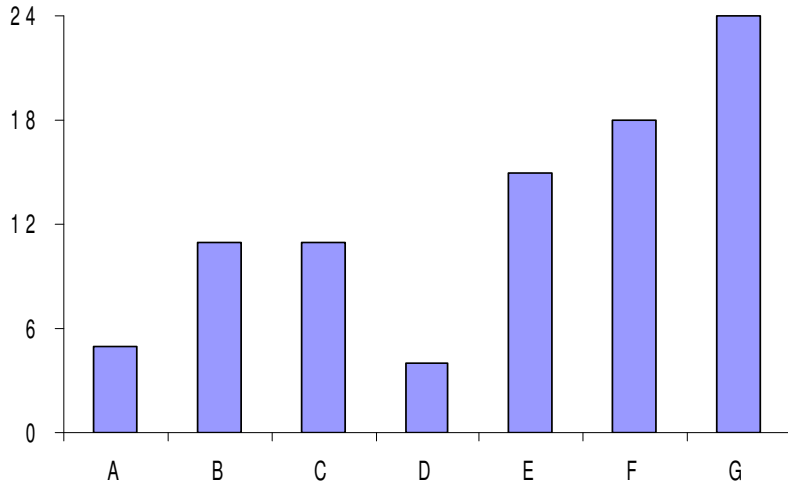
Para establecer la escala en eje vertical, nos guiaremos por la frecuencia máxima, siendo ese el punto más elevado. Puede trabajarse también con frecuencias relativas. El plano resultante quedaría:



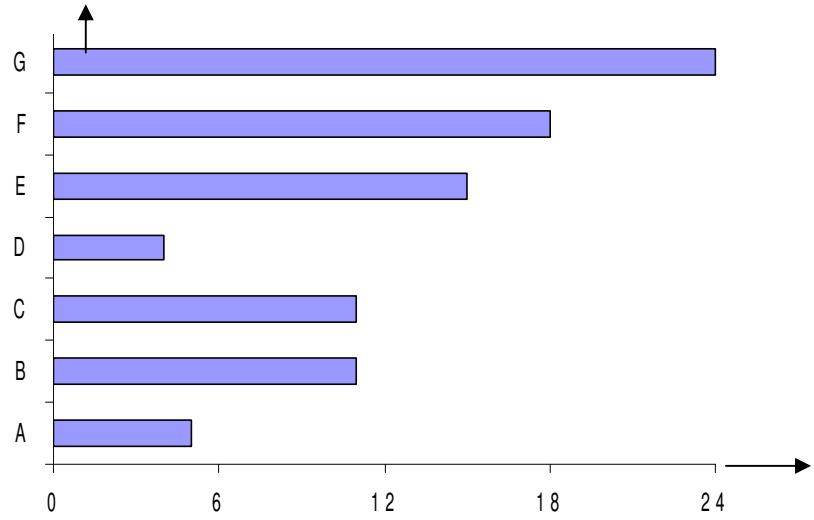
PASO 2: A cada clase se representa con una columna (o barra) cuya altura concuerda con su frecuencia expuesta en el eje vertical. Para la clase A con frecuencia 5, tenemos:



El gráfico final se muestra a continuación:



Puede observar que las columnas se encuentran separadas una de otras. También podríamos realizar el gráfico de forma horizontal (conocido como gráfico de barras):



3.2.2 Características de los gráficos de columnas

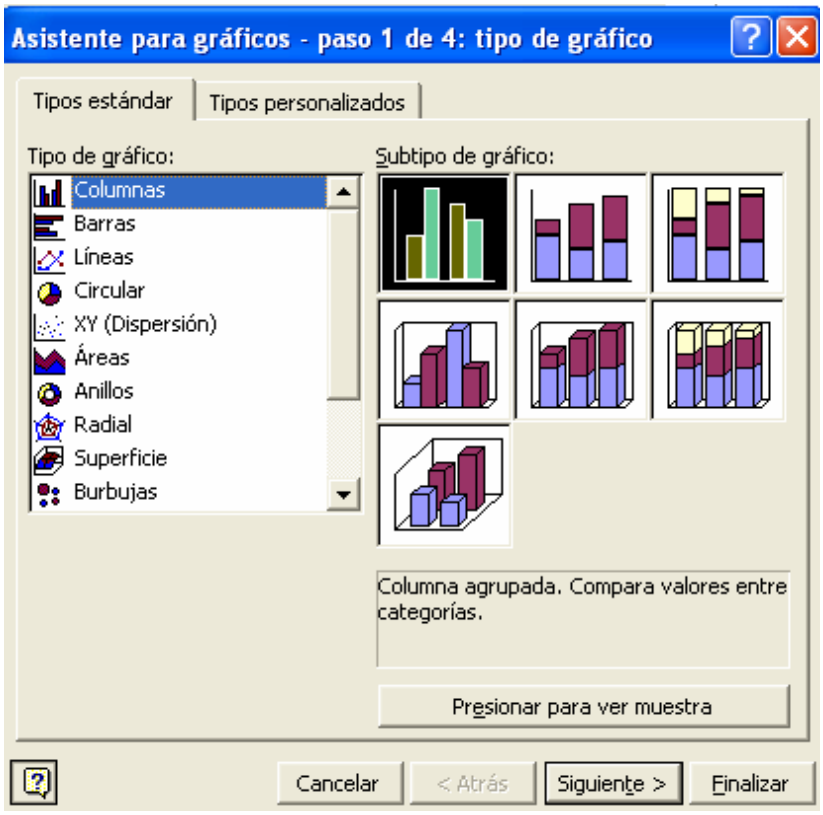
- No muestran frecuencias acumuladas.
- Se prefiere para el tratamiento de datos cualitativos o cuasicualitativos.
- La columna (o barra) con mayor altura representa la mayor frecuencia.
- Son fáciles de elaborar.
- Suelen utilizarse para representar tablas tipo A.
- La sumatoria de las alturas de las columnas equivalen al 100% de los datos.

3.2.3 Construcción de los gráficos de columnas en Excel

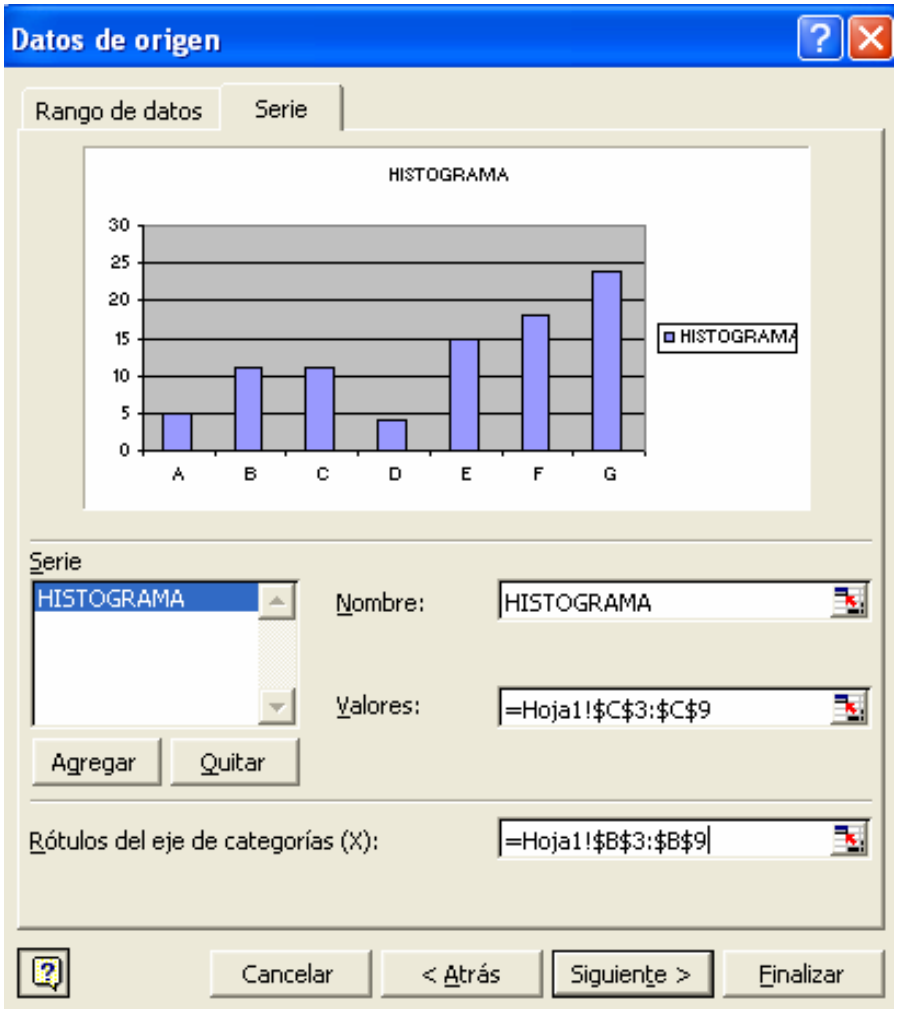
En una hoja nueva copiamos a partir de la celda B2, la tabla de frecuencia del ejemplo anterior y en una celda fuera de la tabla, iniciamos el asistente de gráficos de Excel:

	A	B	C	D
1				
2		**Clase**	**Frecuencia**	
3		A	5	
4		B	11	
5		C	11	
6		D	4	
7		E	15	
8		F	18	
9		G	24	
10		**Total**	**88**	

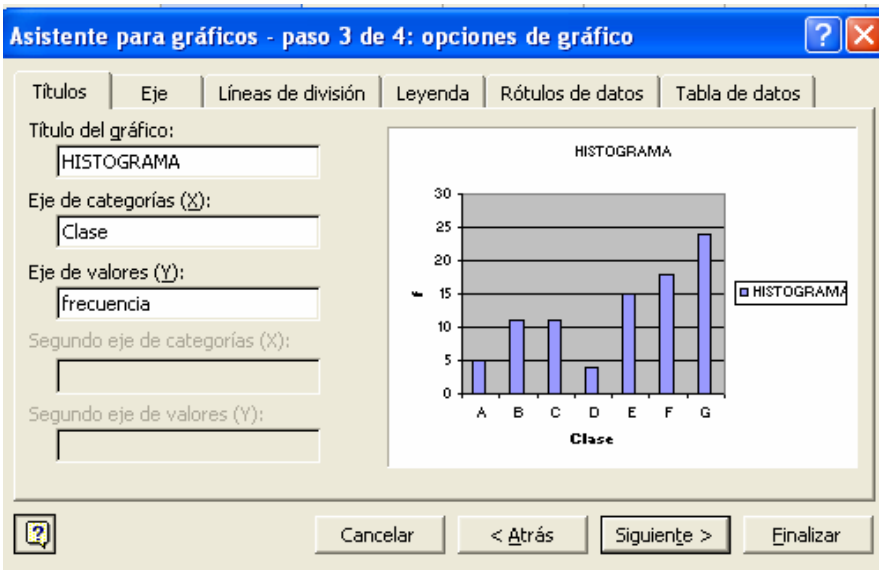
En el asistente seleccionamos el tipo de gráfico columnas y pulsamos el botón Siguiente:

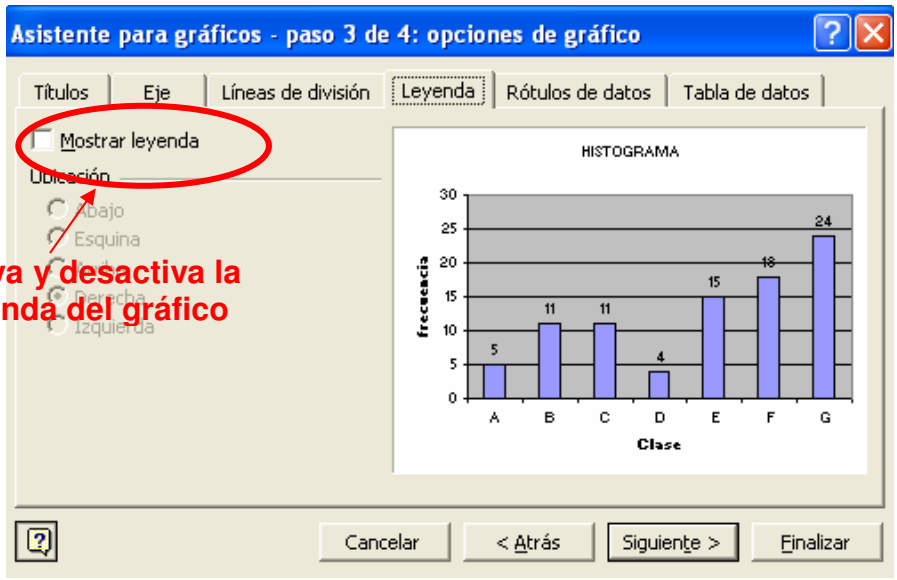


En la ficha Serie pulsamos en el botón Agregar (en el caso de que exista una o varias series en esta ventana, pulsemos en el botón Quitar hasta eliminar todas las series). Llenamos los datos Nombre, Valores y Rótulos del eje de categorías, tal cual como se explico en el gráfico de sectores.

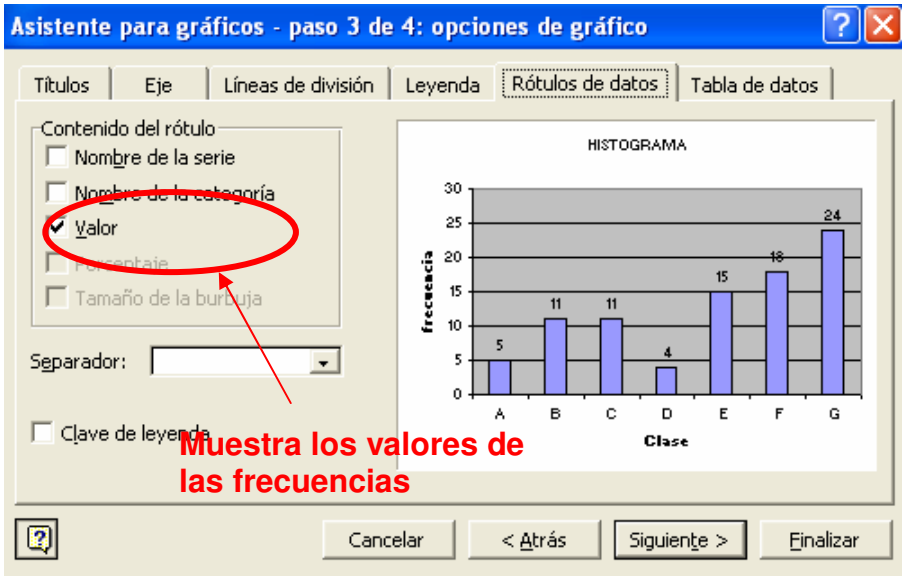


En la ventana siguiente podremos especificar el nombre del eje vertical (eje de valores Y) y horizontal (eje de categorías X) en la ficha Títulos. Para ampliar el gráfico, desactivemos la Leyenda. Podremos también, mostrar los valores de las frecuencias para cada clase en la ficha Rótulos de datos pulsando sobre la opción valor.



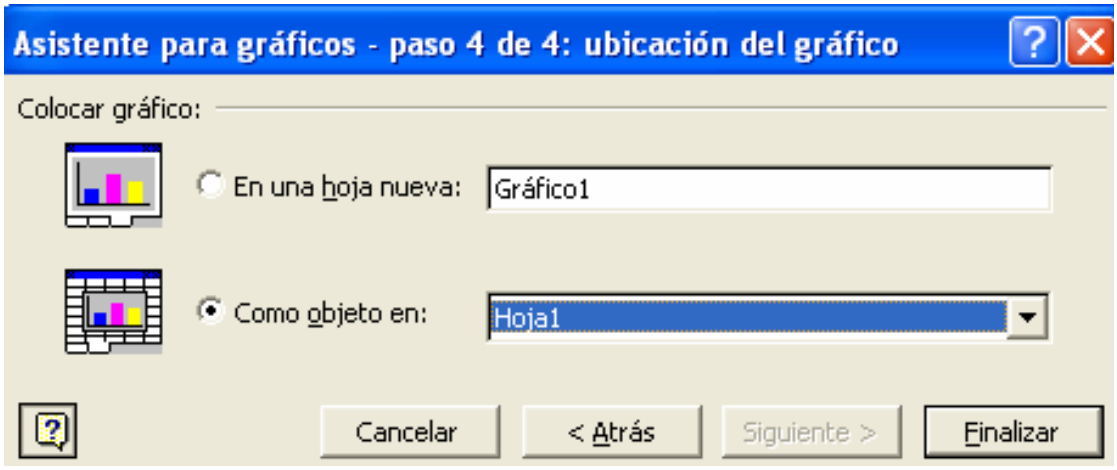


Activa y desactiva la Leyenda del gráfico

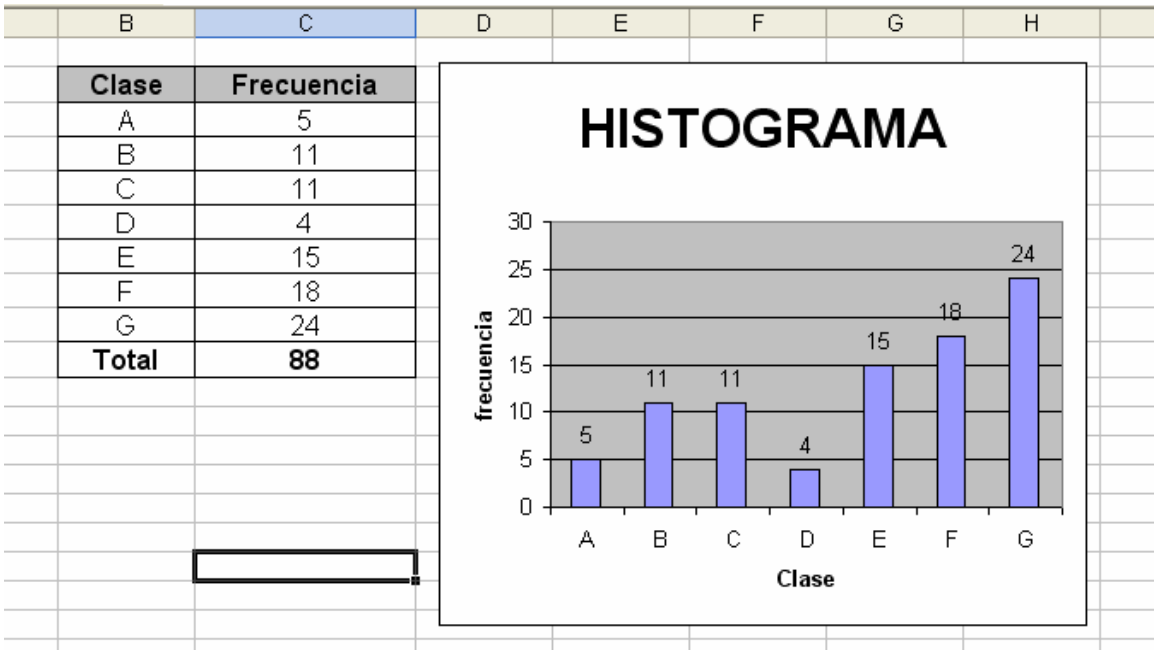


Muestra los valores de las frecuencias

Para terminar, el último paso permite crear el gráfico como objeto en una hoja existente o en una hoja nueva. Esta vez pulsemos sobre Como objeto en y luego en el botón Finalizar.

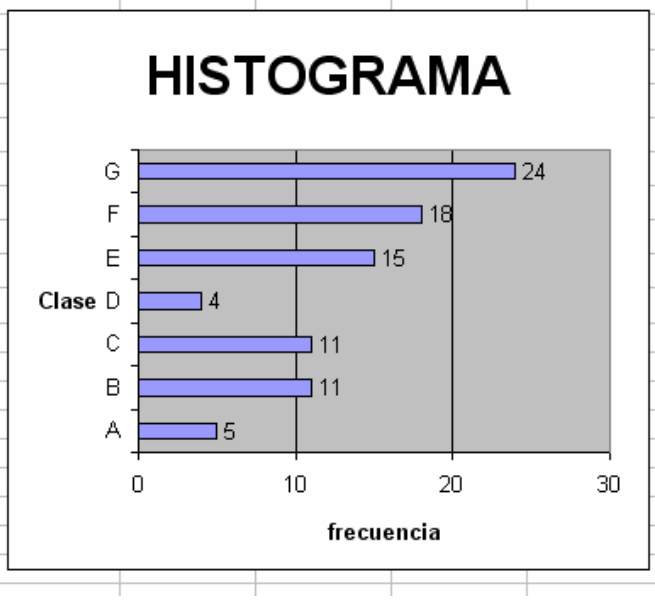


El histograma definitivo (con algunos cambios en el formato y lugar) se muestra a continuación:



Los mismos pasos se emplean para el gráfico de barras:

Clase	Frecuencia
A	5
B	11
C	11
D	4
E	15
F	18
G	24
Total	**88**



3.3 HISTOGRAMA

Se puede considerar como un gráfico de columnas especial. Se realiza sobre el primer cuadrante del plano cartesiano. La diferencia radica en que el histograma se utiliza más a menudo para representar tablas tipo B, donde el ancho de la columna equivale al ancho del intervalo de clase.

Las frecuencias absolutas se colocan en el eje vertical y también puede emplearse las frecuencias relativas. Otra diferencia importante es que no existe espacio entre las barras.

3.3.1 Ejemplo de histogramas

Realizar un histograma a partir de la siguiente tabla de frecuencia:

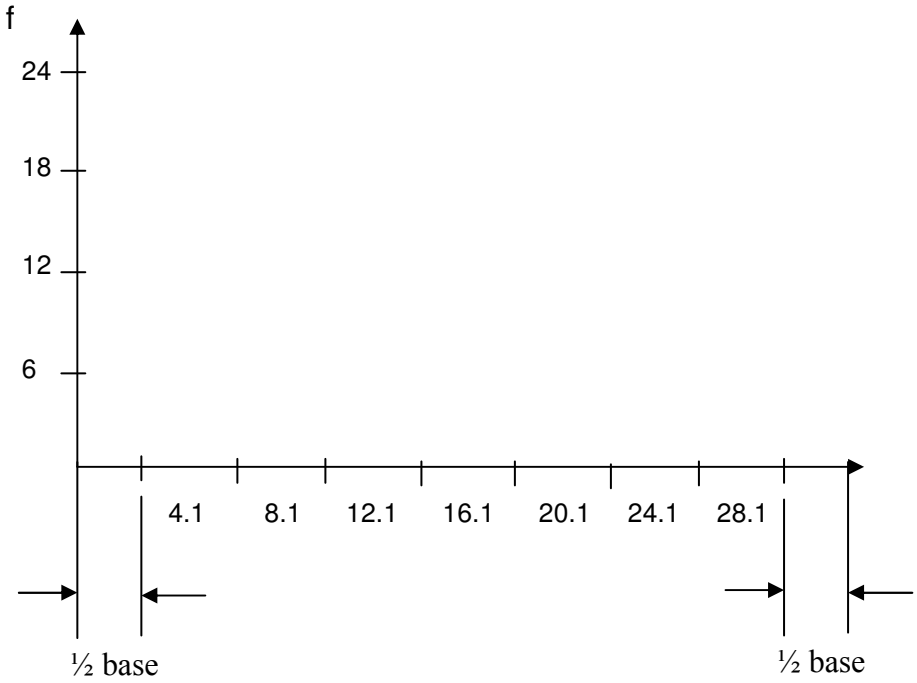
Lm	Ls	Frecuencia	MC
2.0	6.1	12	4.1
6.1	10.1	15	8.1
10.1	14.1	21	12.1
14.1	18.1	24	16.1
18.1	22.1	21	20.1
22.1	26.1	12	24.1
26.1	28.0	8	28.1
Total		**92**	

SOLUCIÓN

Al igual que en gráfico de sectores y el gráfico de columnas, seguiremos 2 pasos para la construcción del histograma.

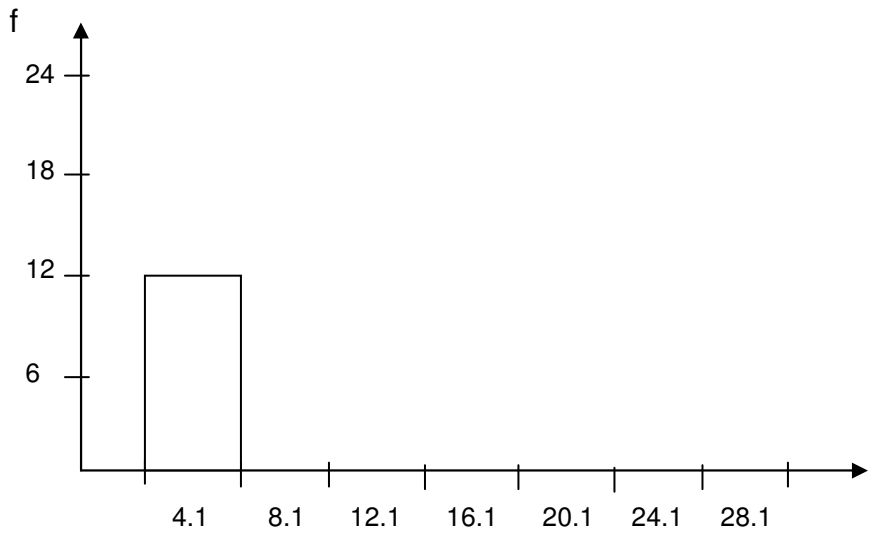
PASO 1: Representar las escalas en los ejes horizontal y vertical del primer cuadrante de un plano de cartesiano. En el eje vertical colocaremos las frecuencias y en el eje horizontal las marcas de clases.

Para diferenciar este paso del anterior, trabajaremos con un ancho de columna único, y dejaremos la mitad de ese espacio entre el vertical y la primera columna, y el final del eje horizontal y la última columna.

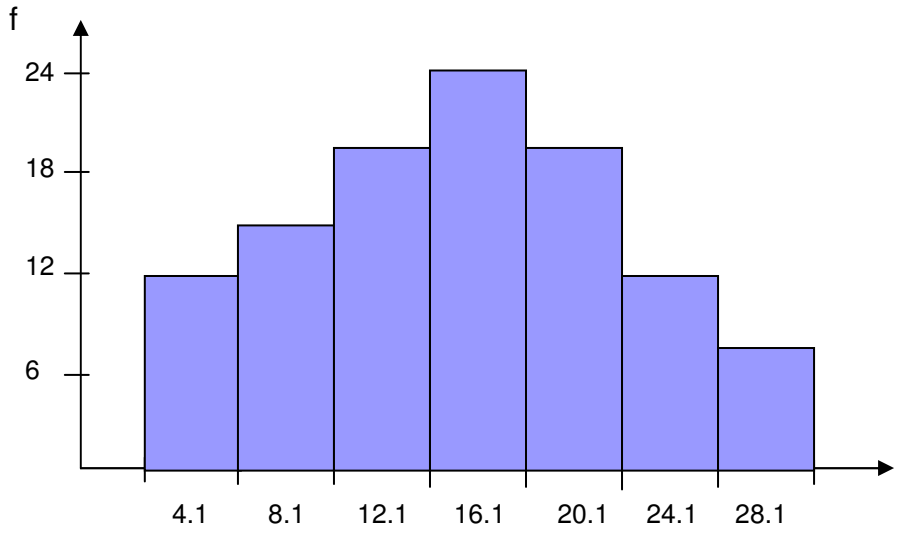


PASO 2: A cada clase se representa con una columna cuya altura concuerda con su frecuencia expuesta en el eje vertical. Podemos decir que la marca de clase 4.1 es representada por una frecuencia 12:

Las barras estarán pegadas una junto a la otra, es decir, ocupando el total del ancho de cada intervalo de clase dispuesto en el gráfico.



El gráfico final quedaría:



3.3.2 Características de los histogramas

- No muestran frecuencias acumuladas.
- Se prefiere para el tratamiento de datos cuantitativos.
- La columna (o barra) con mayor altura representa la mayor frecuencia.
- Suelen utilizarse para representar tablas tipo B.
- La sumatoria de las alturas de las columnas equivalen al 100% de los datos.

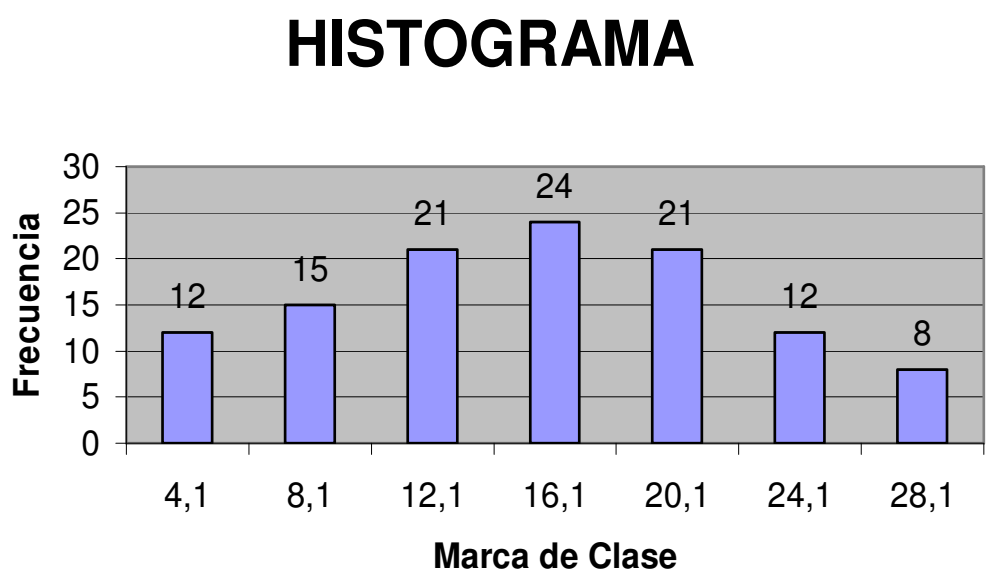
3.3.3 Construcción de histogramas en Excel

Excel no posee un módulo independiente para la creación de histogramas gráficos. Con unos ajustes podremos adaptar un gráfico de columna en algo muy parecido a un histograma, solo que no dejaremos espacios al inicio y fin.

En una hoja nueva copiamos a partir de la celda B2, la tabla de frecuencia tipo B del ejemplo anterior y creamos un gráfico de columnas como lo vimos en la sección anterior:

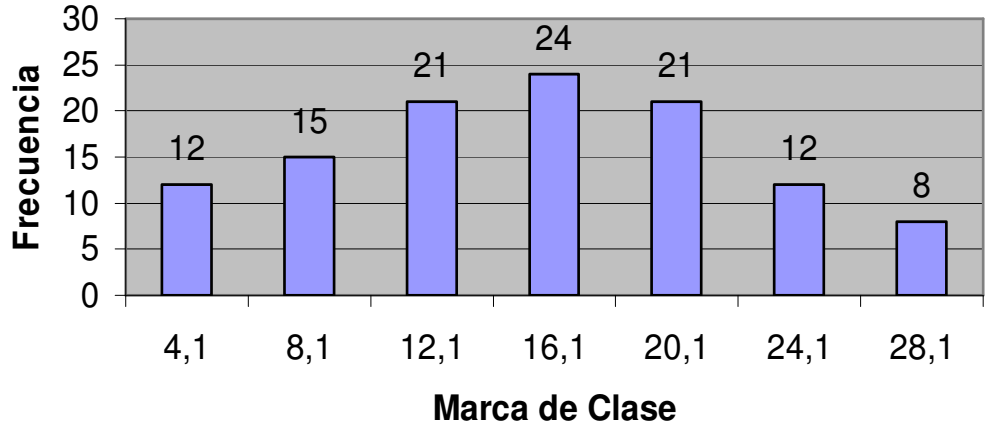
	A	B	C	D	E	F
1						
2		**Lm**	**Ls**	**Frecuencia**	**MC**	
3		2	6,1	12	4,1	
4		6,1	10,1	15	8,1	
5		10,1	14,1	21	12,1	
6		14,1	18,1	24	16,1	
7		18,1	22,1	21	20,1	
8		22,1	26,1	12	24,1	
9		26,1	28	8	28,1	
10		**Total**		**92**		

El gráfico se vería como sigue:

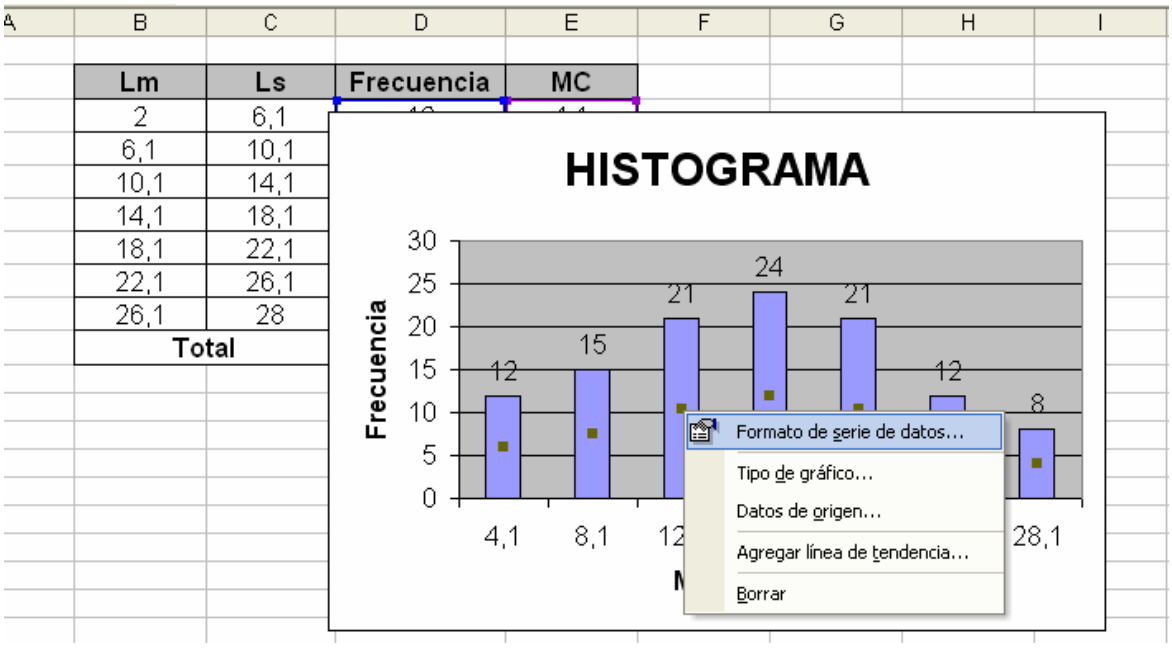


Ubicando el puntero del Mouse sobre una de las columnas, pulsamos el botón derecho y en el menú flotante que aparece, se selecciona Formato de series de datos:

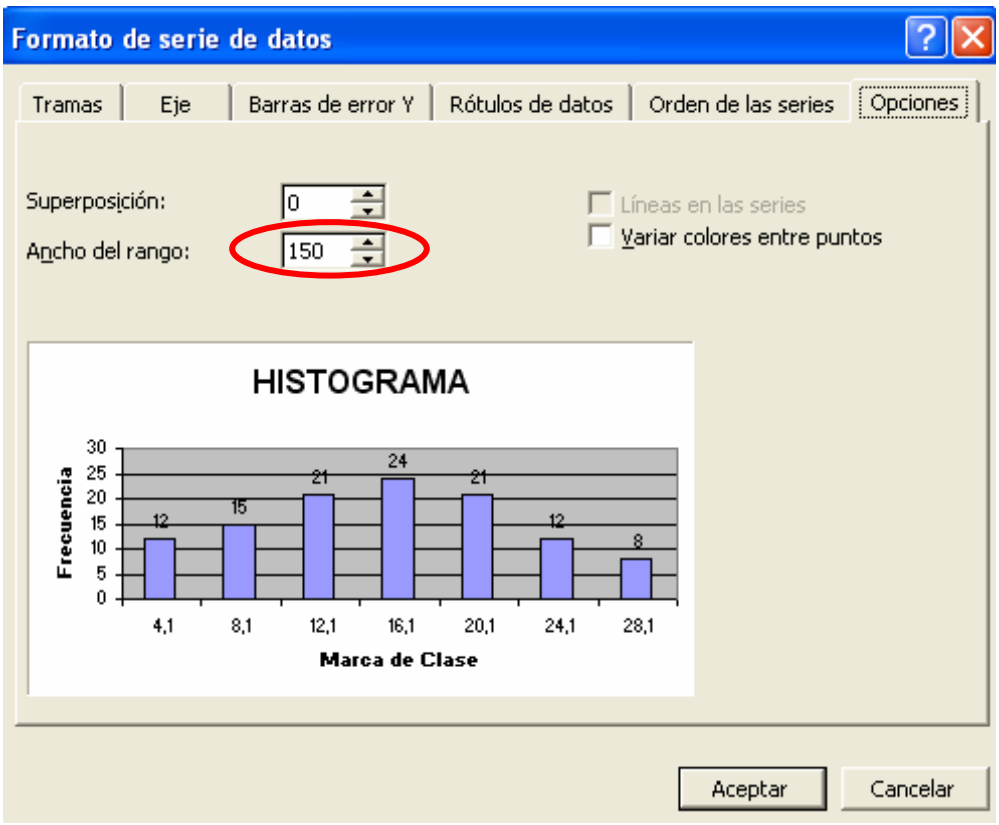
HISTOGRAMA



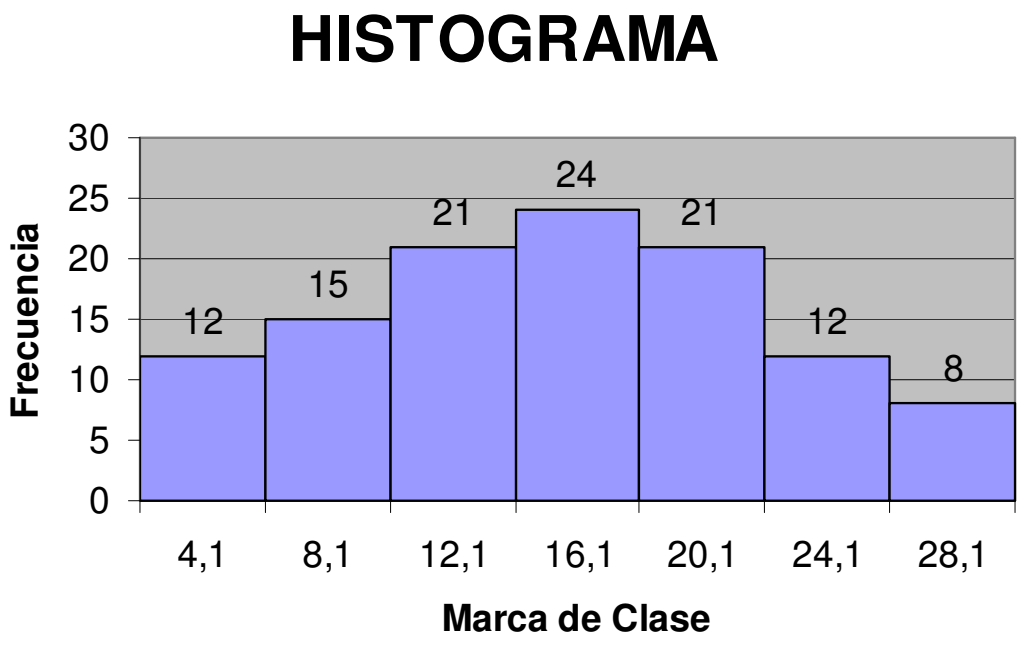
Ubicando el puntero del Mouse sobre una de las columnas, pulsamos el botón derecho y en el menú flotante que aparece, se selecciona Formato de series de datos:



En la ventana generada pulsaremos sobre la ficha opciones:



Disminuimos la casilla Ancho de rango a cero para juntar las barras y pulsamos en el botón Aceptar:



3.4 POLÍGONOS DE FRECUENCIAS

Este gráfico se utiliza para el caso de variables cuantitativas, tanto discretas como continuas, partiendo del diagrama de columnas, barras o histograma, según el tipo de tabla de frecuencia manejada.

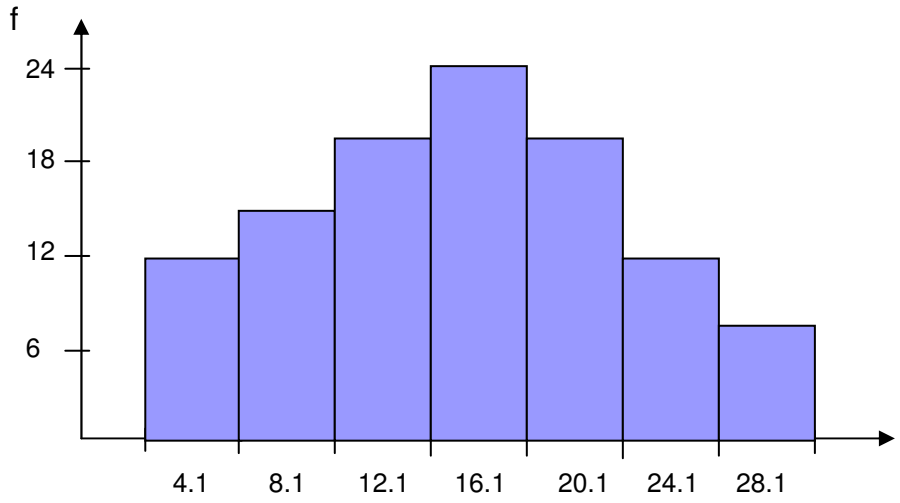
3.4.1 Ejemplo de polígonos de frecuencias

Realizar un polígono de frecuencia a partir de la tabla de frecuencia dada en el ejemplo anterior:

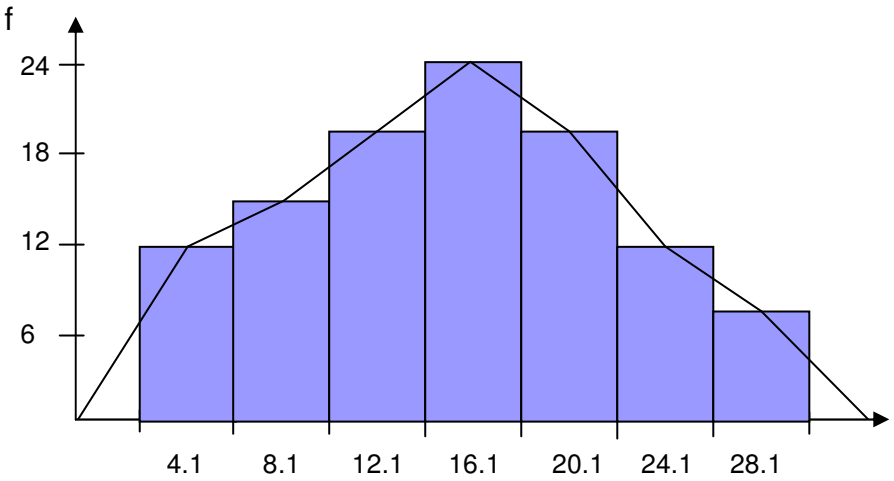
Lm	Ls	Frecuencia	MC
2.0	6.1	12	4.1
6.1	10.1	15	8.1
10.1	14.1	21	12.1
14.1	18.1	24	16.1
18.1	22.1	21	20.1
22.1	26.1	12	24.1
26.1	28.0	8	28.1
Total		**92**	

SOLUCIÓN

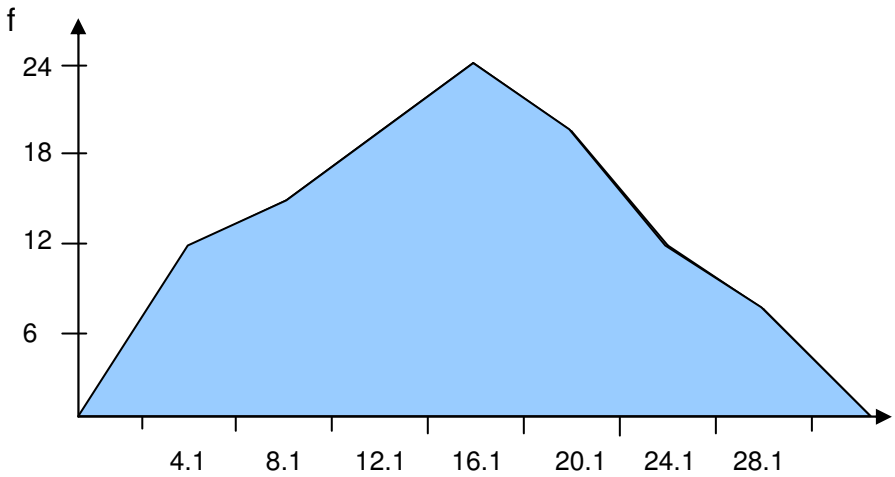
PASO 1: Crear un histograma (tabla tipo B) o gráfico de columnas (tabla tipo A).



PASO 2: Trazar líneas rectas entre los puntos medios de los techos de columnas contiguas, partiendo desde el punto de origen (0,0) hasta el punto final definido en el eje horizontal.

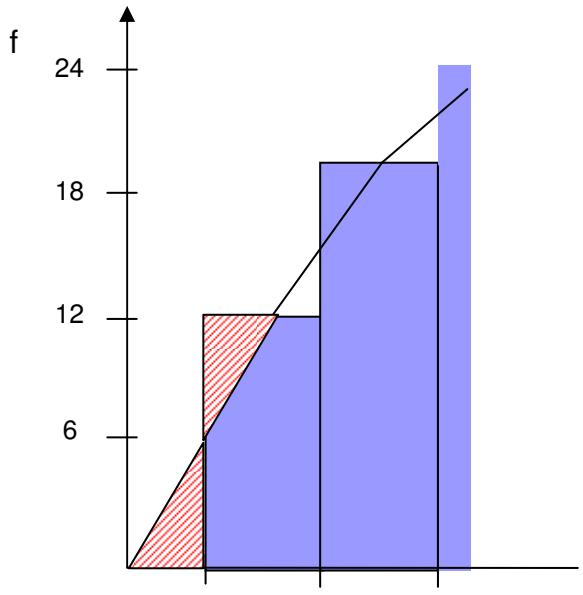


Nuestro polígono de frecuencias sin el histograma quedaría de la siguiente forma:

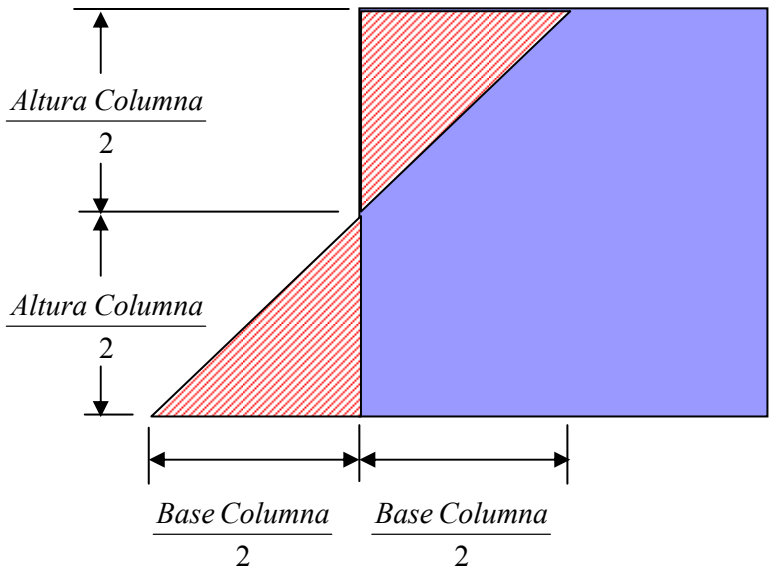


3.4.2 Características de los polígonos de frecuencias

- No muestran frecuencias acumuladas.
- Se prefiere para el tratamiento de datos cuantitativos.
- El punto con mayor altura representa la mayor frecuencia.
- Suelen utilizarse para representar tablas tipo B.
- El área bajo la curva representa el 100% de los datos. El polígono de frecuencia esta diseñado para mantener la misma área de las columnas. Analicemos una porción de nuestro gráfico para probar esta afirmación:

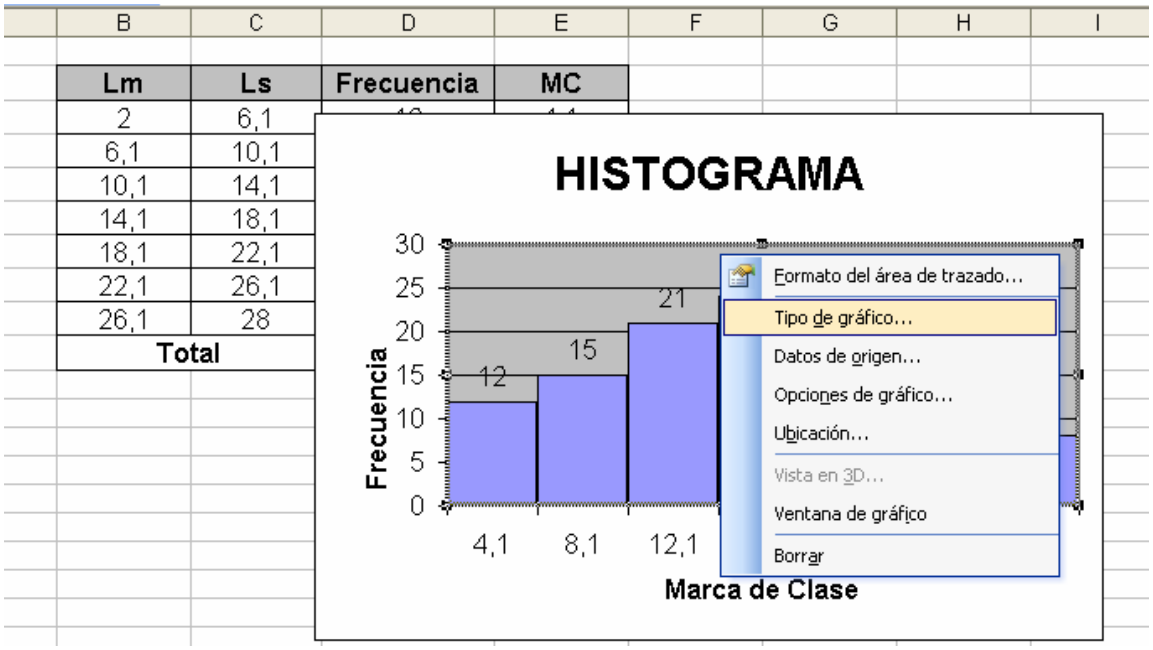


Observe que cada línea corta una porción de la columna, pero a su vez, agrega una porción adicional. Ambas porciones son iguales (triangulo rectángulos iguales), manteniendo el área global en el gráfico.

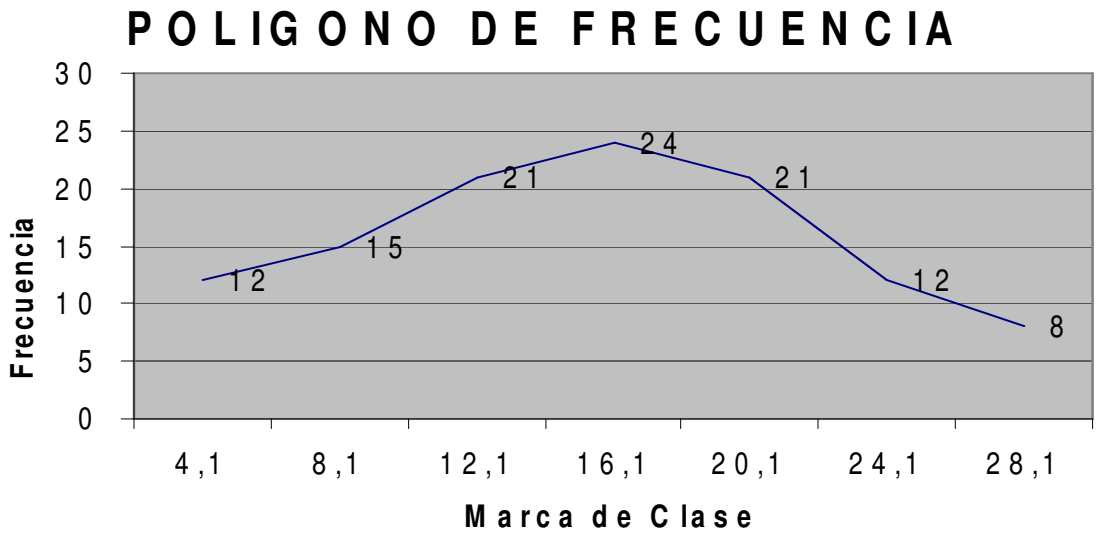


3.4.3 Construcción de los polígonos de frecuencias en Excel

A partir de la construcción de un histograma en Excel, procedemos a cambiar el tipo de gráfico pulsando con el botón derecho del Mouse sobre el gráfico y eligiendo la opción tipo de gráfico.



Cambiemos el gráfico a líneas y pulsemos el botón Aceptar. Cambiemos también el titulo de HISTOGRAMA por POLIGONO DE FRECUENCIA.

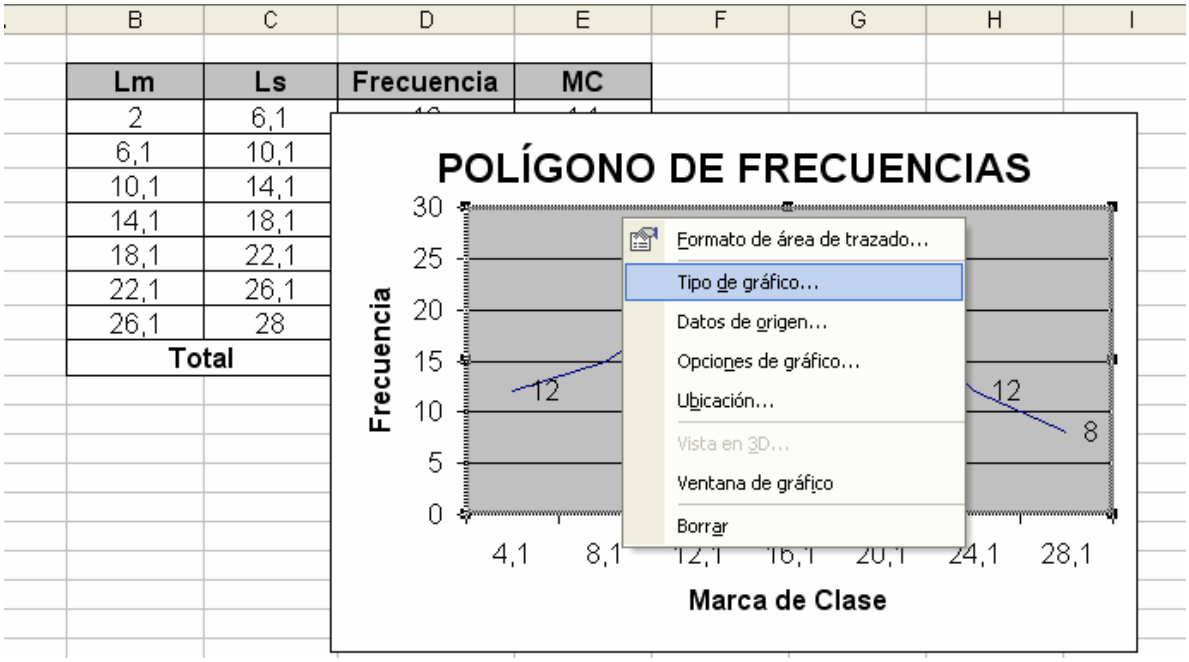


3.5 CURVAS SUAVIZADAS O CURVAS DE FRECUENCIAS

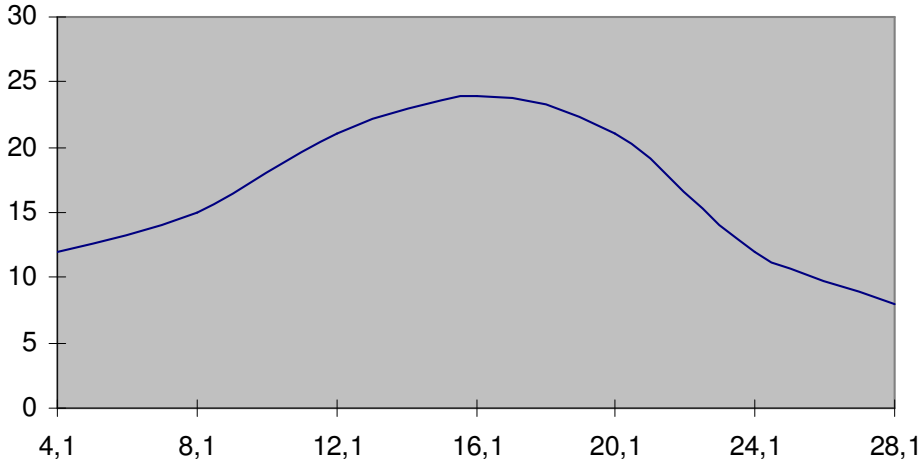
Son gráficos representados por una sola línea curva (el polígono de frecuencia esta conformado por varias líneas rectas consecutivas).

3.5.1 Construcción de las curvas suavizadas en Excel

Sobre el gráfico anterior, pulsemos el botón derecho del Mouse y en la opción Tipo de gráfico.



En la ventana seleccionamos la ficha Tipos personalizados y elegimos línea suavizada. Eliminemos la leyenda que aparece para ampliar el gráfico y cambiamos el título a LÍNEA SUAVIZADA.



3.5.2 Características de las curvas suavizadas

- No muestran frecuencias acumuladas.
- Se prefiere para el tratamiento de datos cuantitativos.
- El punto con mayor altura representa la mayor frecuencia.

- Suelen utilizarse para representar tablas tipo B.
- Son más complicadas de elaborar que los gráficos anteriores.
- El área bajo la curva representa el 100% de los datos.

3.6 OJIVAS

En este gráfico se emplea un polígono de frecuencia o curva suavizada con una característica muy particular: muestra las frecuencias absolutas o relativas acumuladas.

3.6.1 Ejemplo de ojivas

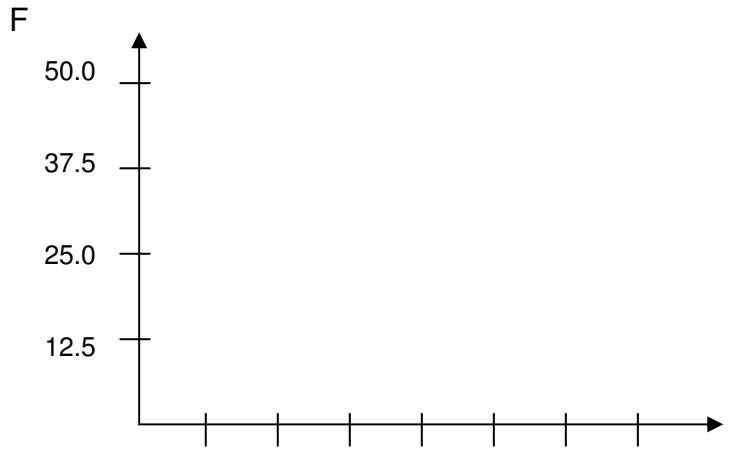
Los ingresos de 50 trabajadores de una empresa se resumen en la siguiente tabla de frecuencia:

Ni	Lm	Ls	f	F	h	H	MC
1	[100000	150000)	3	3	6,00%	6,00%	125000
2	[150000	200000)	2	5	4,00%	10,00%	175000
3	[200000	250000)	5	10	10,00%	20,00%	225000
4	[250000	300000)	10	20	20,00%	40,00%	275000
5	[300000	350000)	4	24	8,00%	48,00%	325000
6	[350000	400000)	15	39	30,00%	78,00%	375000
7	[400000	450000]	11	50	22,00%	100,00%	425000
Total			**50**		**100,00%**		

SOLUCIÓN

PASO 1: En un plano cartesiano (primer cuadrante), procedemos a establecer la escala de ambos ejes:

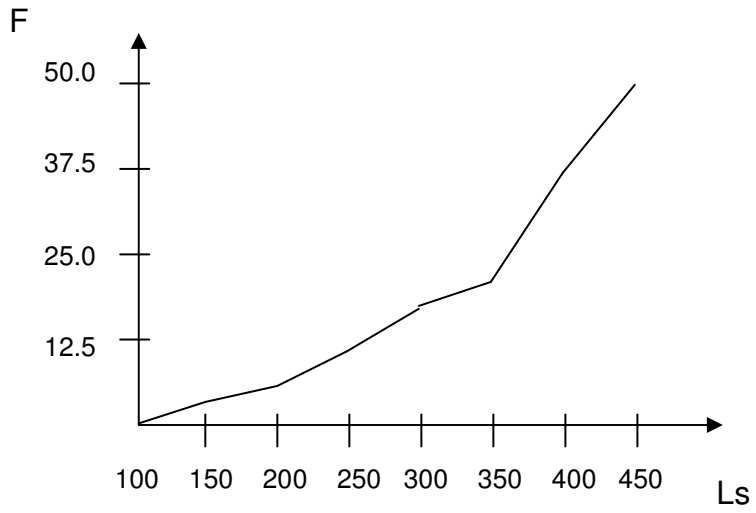
- En el eje vertical se colocaran las frecuencias, partiendo desde 0 hasta la última frecuencia acumulada (absoluta o relativa).
- En el eje horizontal se ubicaran los límites superiores, partiendo desde el primer límite menor.



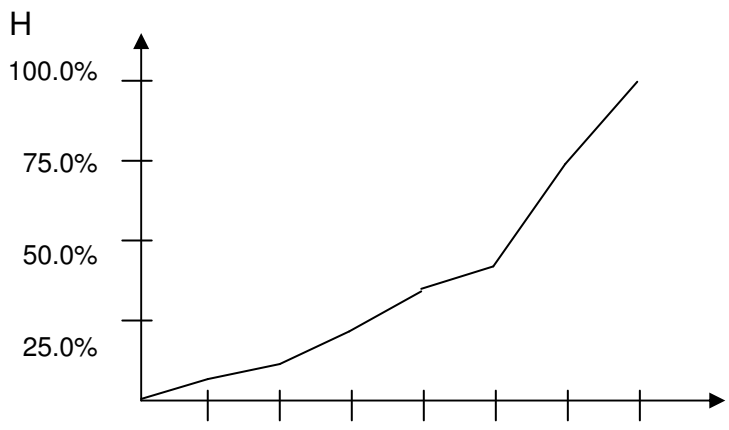
PASO 2: A cada límite superior le corresponde su frecuencia acumulada. El punto inicial será 0 (no existen datos por debajo de 100). Por ejemplo:

- Hasta un valor cercano a 150.000 existen acumulado 3 datos.
- Hasta un valor cercano a 200.000 existen acumulado 5 datos.

Gráficamente tenemos:



Podremos cambiar la escala del eje vertical para que trabaje con las frecuencias relativas acumuladas sin afectar el gráfico:



3.6.2 Características de las ojivas

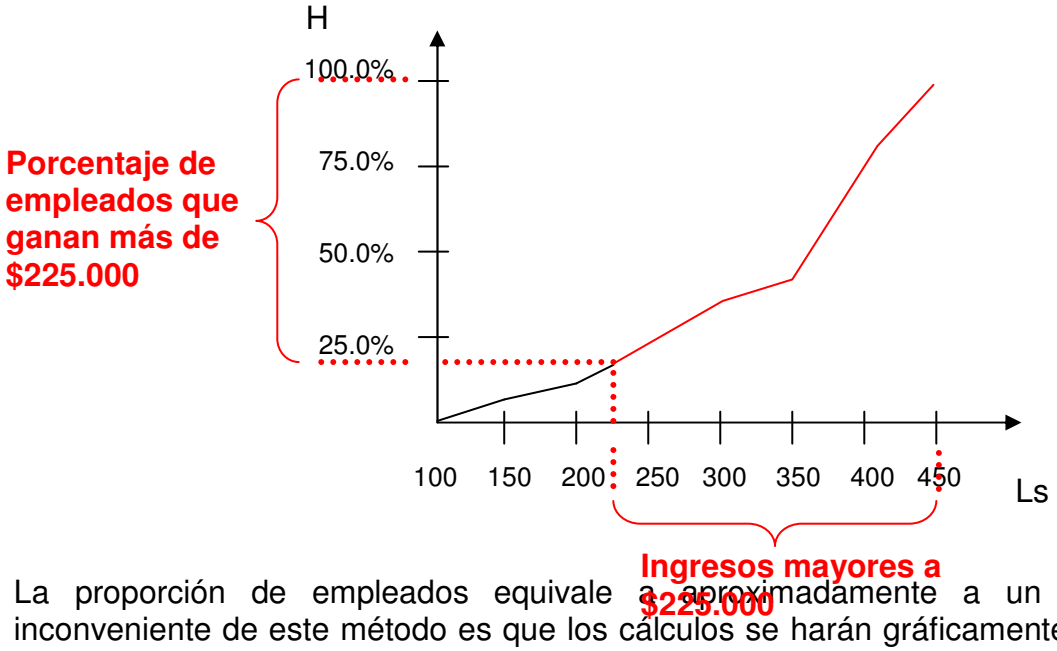
- Muestran frecuencias acumuladas.
- Se prefiere para el tratamiento de datos cuantitativos.
- El punto de inicio equivale a una frecuencia de 0.
- Suelen utilizarse para representar tablas tipo B.
- El punto final equivale al 100% de los datos.

3.6.3 Interpretando la información en las ojivas

Dada su ventaja de representar frecuencias acumuladas, las ojivas se convierten en una herramienta vital para el análisis estadístico.

Partiendo de la información del ejemplo, se desea conocer que porcentaje de los trabajadores ganan más $225.000.

Para conocer esta información, debemos ubicar primero el valor de $225.000 en el gráfico y luego, proyectar este punto en el eje vertical:



La proporción de empleados equivale a aproximadamente a un 80%. El inconveniente de este método es que los cálculos se harán gráficamente (valores aproximados) y no de forma aritmética (valores exactos).

3.6.4 Construcción de ojivas en Excel

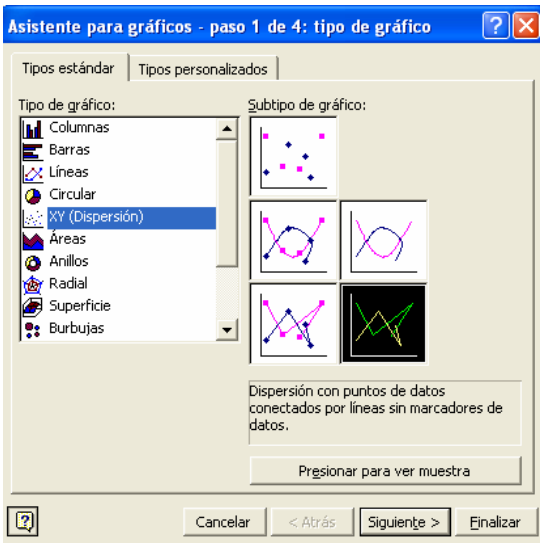
A continuación construiremos una ojiva, empleando la tabla de frecuencia vista en el ejemplo. Modificaremos los límites superiore e inferiores para poder trabajar en Excel:

	Lm	Ls	f	F	h	H	MC
3	100000.0	150000.1	3	3	0,06	0,06	125000
4	150000.1	200000.1	2	5	0,04	0,1	175000
5	200000.1	250000.1	5	10	0,1	0,2	225000
6	250000.1	300000.1	10	20	0,2	0,4	275000
7	300000.1	350000.1	4	24	0,08	0,48	325000
8	350000.1	400000.1	15	39	0,3	0,78	375000
9	400000.1	450000.0	11	50	0,22	1	425000
10	**Total**		**50**		**1**		

Antes del iniciar el asistiente de gráficos, debemos indicar que para un valor menor a $100.000, no existen datos:

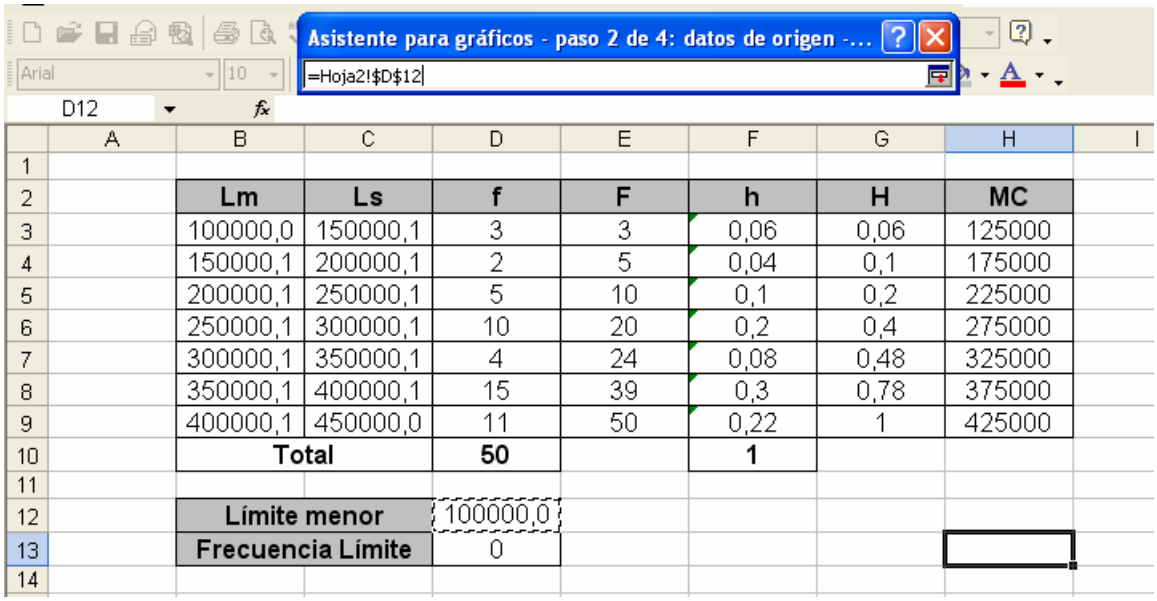
12	**Límite menor**	100000,0
13	**Frecuencia Límite**	0

En el asistente selecionamos el tipo de gráfico XY dispersión, optando por un gráfico por líneas rectas:



En la ventana siguiente agregamos el nombre OJIVA, y procedemos a asignar los valores para el eje horizontal (X) y eje vertical (Y). Habíamos recalcado que la

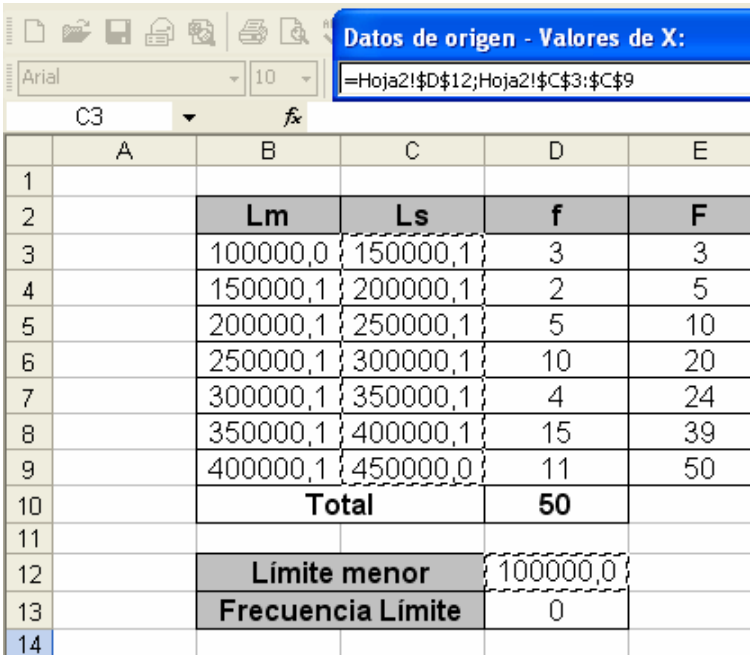
ojiva comienza con el primer límite inferior, por tanto, debemos pulsar sobre la celda C12 equivalente a los $100.000.



	Lm	Ls	f	F	h	H	MC
3	100000,0	150000,1	3	3	0,06	0,06	125000
4	150000,1	200000,1	2	5	0,04	0,1	175000
5	200000,1	250000,1	5	10	0,1	0,2	225000
6	250000,1	300000,1	10	20	0,2	0,4	275000
7	300000,1	350000,1	4	24	0,08	0,48	325000
8	350000,1	400000,1	15	39	0,3	0,78	375000
9	400000,1	450000,0	11	50	0,22	1	425000
10	Total		50		1		
12	Límite menor		100000,0				
13	Frecuencia Límite		0				

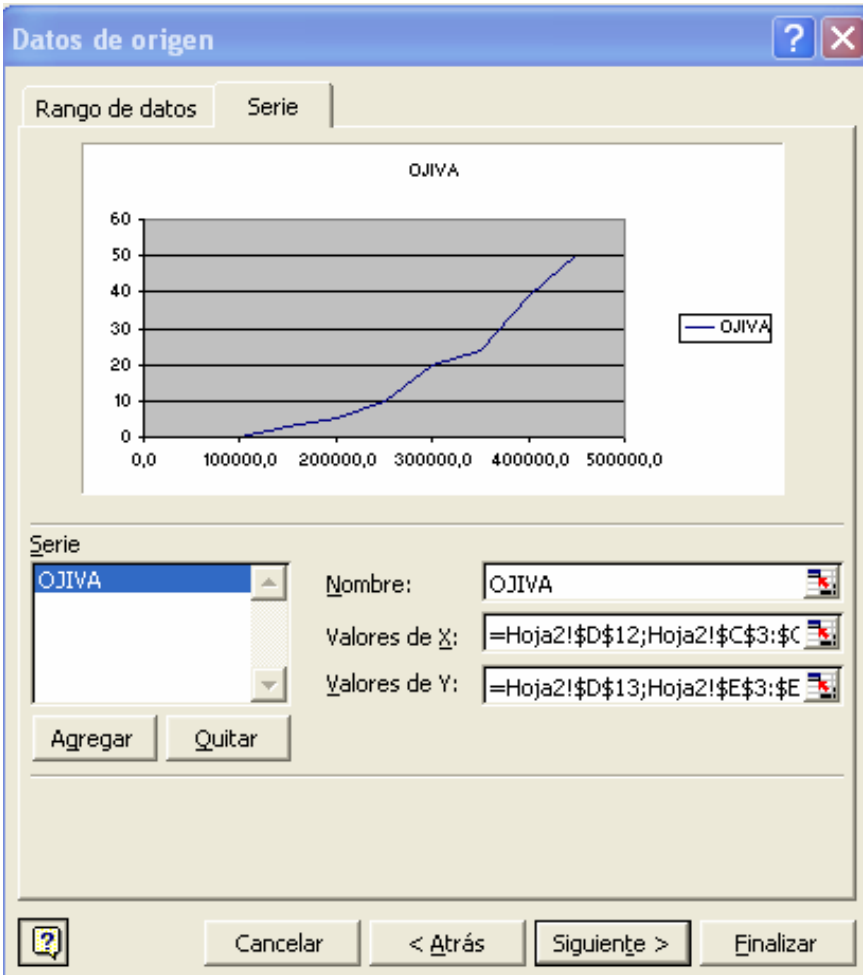
Siguen los límites superiores, que se seleccionaran manteniendo la tecla Ctrl pulsada:

+

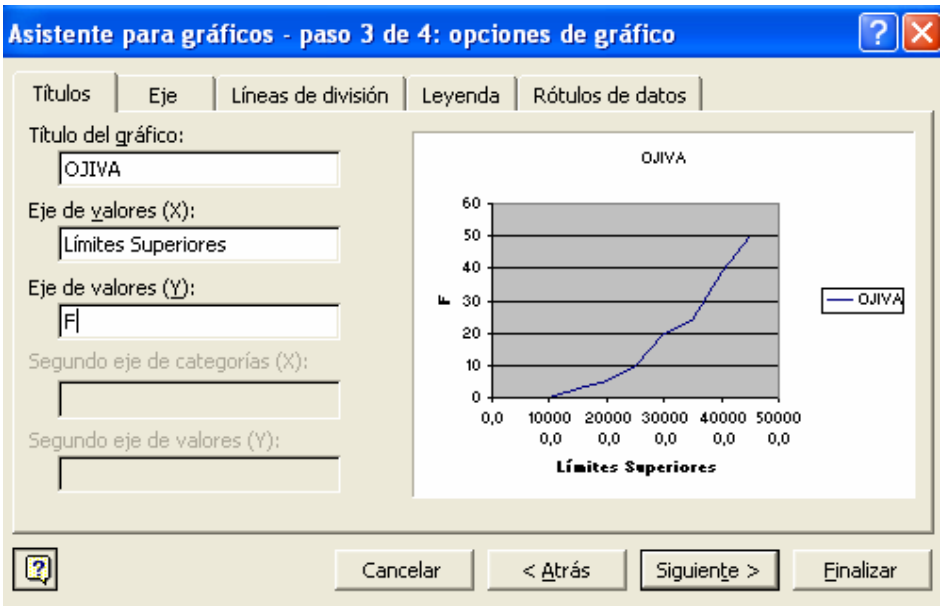


	Lm	Ls	f	F
3	100000,0	150000,1	3	3
4	150000,1	200000,1	2	5
5	200000,1	250000,1	5	10
6	250000,1	300000,1	10	20
7	300000,1	350000,1	4	24
8	350000,1	400000,1	15	39
9	400000,1	450000,0	11	50
10	Total		50	
12	Límite menor		100000,0	
13	Frecuencia Límite		0	

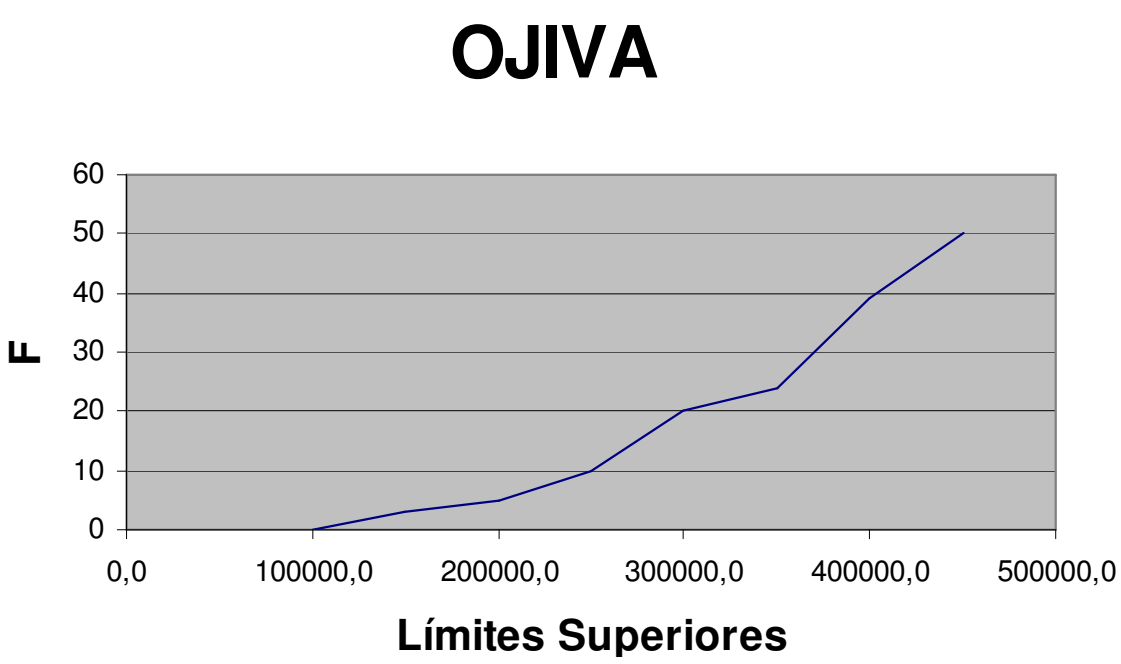
De igual forma aplicaremos la misma operación a los datos ligados al eje horizontal. La ventana Datos de origen quedaría:



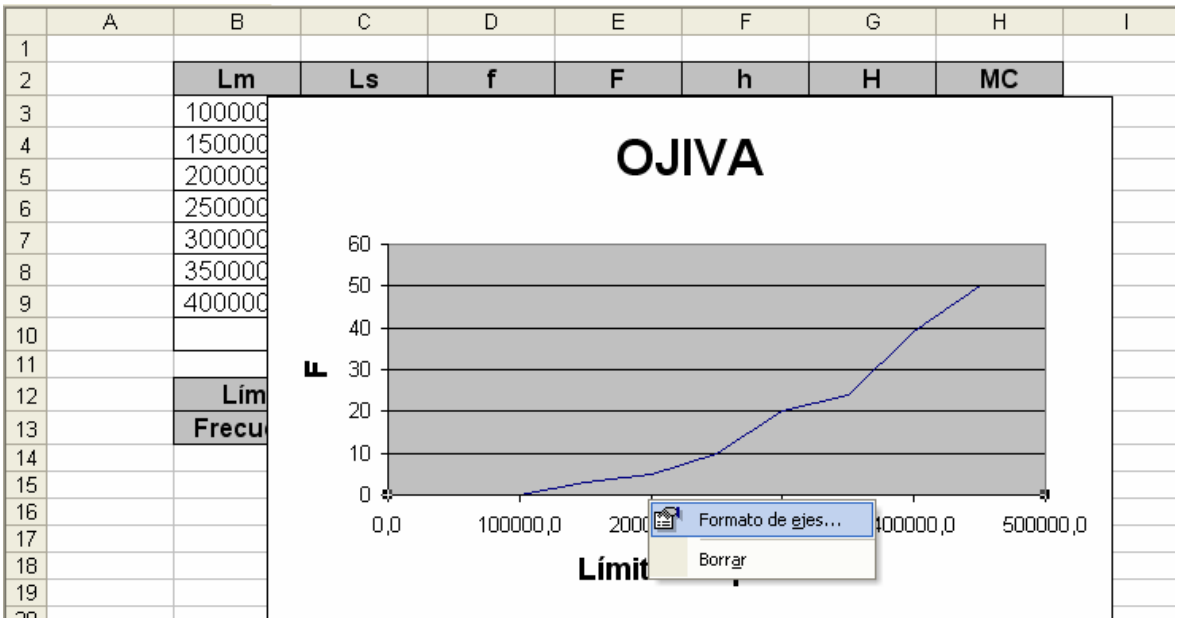
En la tercera ventana agregamos información adicional al gráfico y desactivar o activar la leyenda:



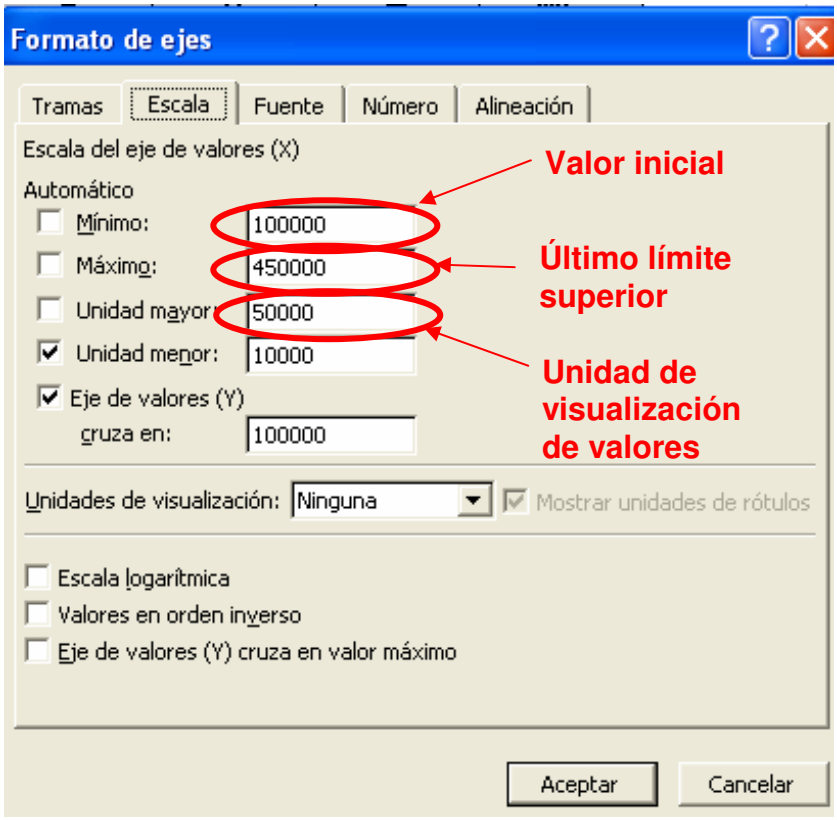
La gráfica resultante debería quedar como sigue (con algunos ajustes en el formato):



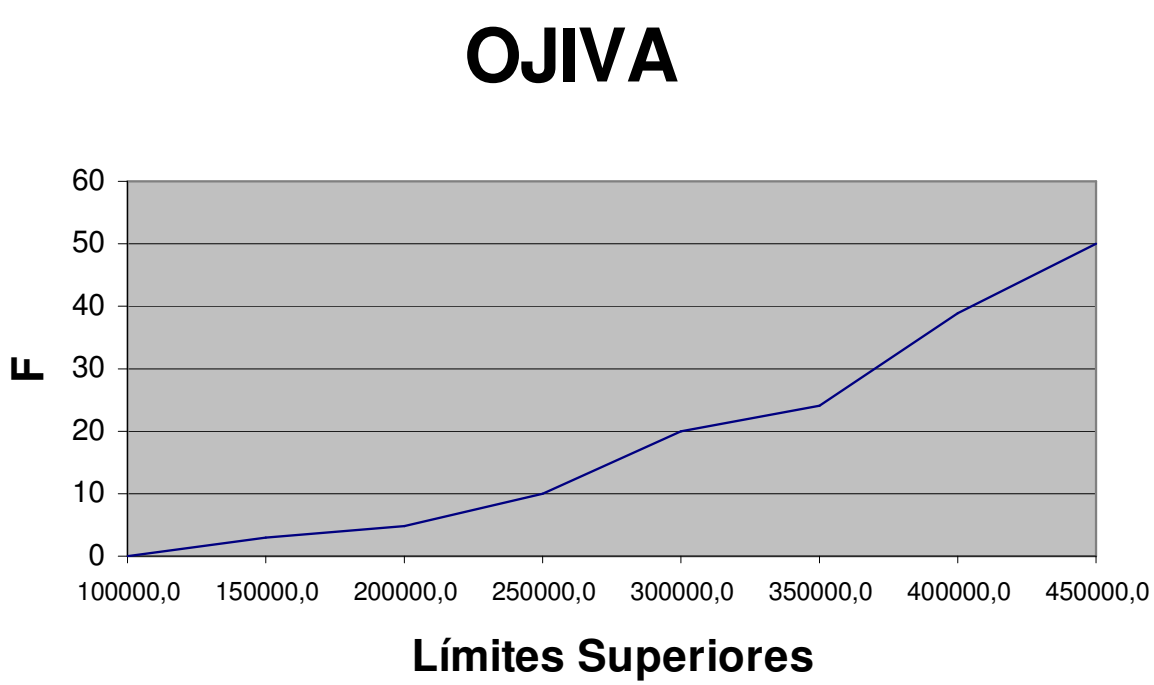
Corrijamos la escala del eje horizontal para que empiece con $100.000. Esto se realiza pulsando sobre la escala con el botón derecho del Mouse y marcando la opción Formato de ejes.



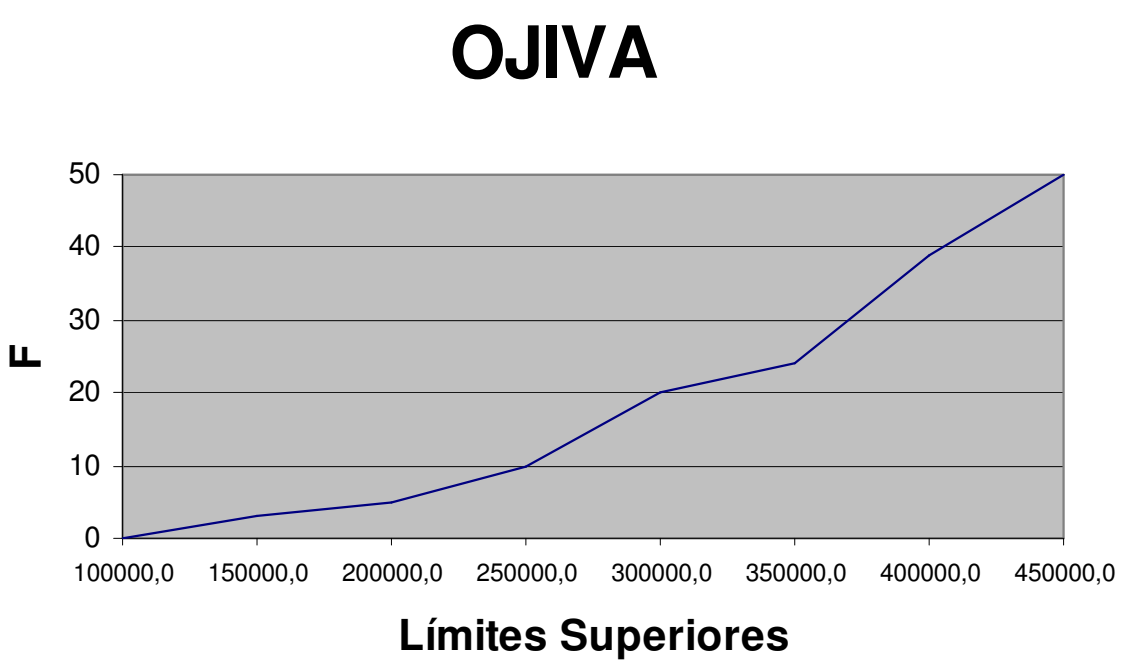
En la ficha Escala realizamos los siguientes cambios:



La escala modificada se mostrará así:



Modificamos también el valor máximo del eje vertical para que termine en 50.



3.7 PICTOGRAMAS

Los pictogramas utilizan símbolos para representar un conjunto de datos. La mayor frecuencia se identifica por la mayor acumulación de símbolos.

Los pictogramas se emplean sobre todo, para hacer más amigables e entendibles los informes estadísticos.

3.7.1 Ejemplo de pictogramas

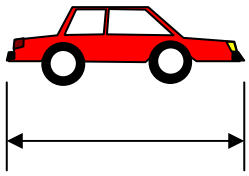
La demanda anual de un tipo particular de vehículos en algunos países de Suramérica se muestra a continuación:

País	Demanda
Colombia	20.000
Venezuela	40.000
Argentina	120.000
Chile	150.000
Brasil	160.000

Realizar un pictograma para la tabla anterior.

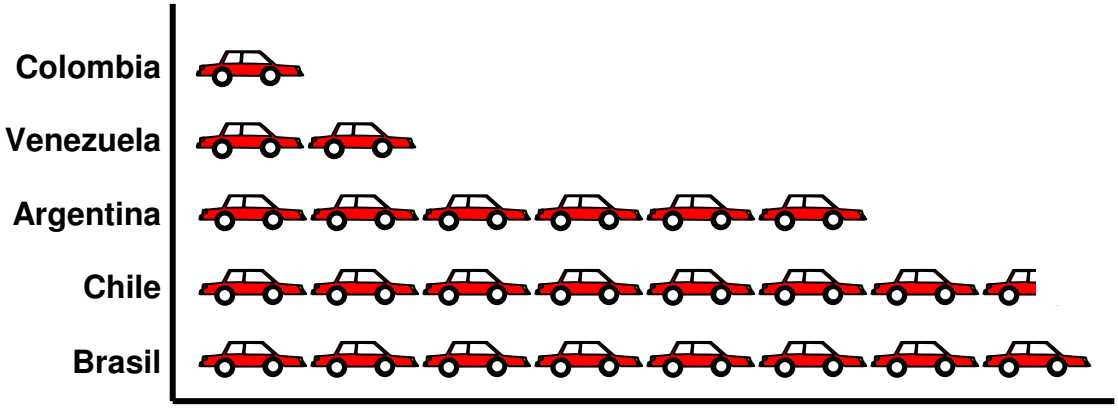
SOLUCIÓN

El símbolo que emplearemos tendrá forma de vehículo, asiendo referencia al tema del informe. Cada símbolo tendrá una equivalencia de 20.000 unidades demandadas.



20.000
unidades

En un eje cartesiano colocamos los países en el eje vertical y las demandas en el horizontal.



En el caso de Chile, la demanda equivale a 7 y ½ vehículos (140.000 + 10.000 unidades). Este tipo de pictogramas tiene forma a un gráfico de barras.

3.7.2 Características de los pictogramas

- Su formato es libre.
- Emplean una secuencia de símbolos para representar frecuencias.
- Se emplean para el tratamiento de datos tanto cualitativos como cuantitativos.

3.8 EJERCICIOS PROPUESTOS

3.8.1 Realice un gráfico de sectores a la tabla de frecuencia que aparece en el ***ejercicio 2.3.2***.

3.8.2 Realice un gráfico de columnas a la tabla de frecuencia que aparece en el ***ejercicio 2.3.3***.

3.8.3 Realice un histograma a la tabla de frecuencia que aparece en el ***ejercicio 2.3.5***.

3.8.4 Una muestra de 100 estudiantes del programa de ingeniería de una universidad, mostraron sus preferencias respecto a la creación de un nuevo laboratorio en una encuesta para el diario estudiantil:

Tipo de laboratorio	Número de alumnos a favor
Estadística	25
Control de calidad	10
Neumática	15
Hidráulica	20
Simulación	30

Muestre los datos gráficamente empleando:

a. Un gráfico de columna
b. Un gráfico de barras
c. Un gráfico de sectores

3.8.5 A partir de los siguientes datos, cree la correspondiente tabla de frecuencia y grafique:

6,42	66,49	72,71
92,64	49,55	37,33
64,86	9,8	36,33
14,97	42,92	19,6
13,22	5,32	85,45
66,85	77,37	93,43

a. Un histograma
b. Un polígono de frecuencia
c. Una OJIVA

3.8.6 A partir del gráfico de ojiva, responda las siguientes preguntas (Tamaño de la muestra es 500):

a. ¿Que cantidad de datos hay acumulado hasta 260.5?
b. ¿Sirve este tipo de gráfico para mostrar la frecuencia absoluta (f)?
c. Diseñe la tabla de frecuencia respectiva

3.8.7 Cree una tabla de frecuencia que contenga 7 intervalos de clase, para los siguientes datos:

31,2	44,3	31,8
19,0	59,9	87,9
66,1	5,4	47,9
96,6	36,5	74,0
42,7	10,6	56,0
87,7	11,7	30,1
5,3	11,7	31,4
51,2	67,0	46,8
60,7	29,6	55,6
67,0	32,1	82,2
81,2	75,5	91,0
40,4	42,4	31,8
26,6	70,1	30,4
6,4	19,1	77,6
57,3	62,1	40,9

Construya el histograma respectivo.

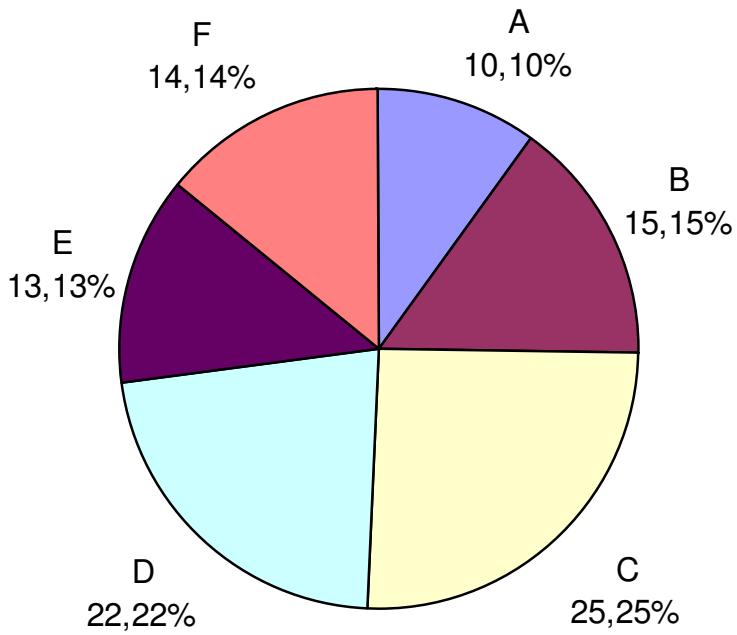
3.8.8 El cuadro que figura más abajo da los caudales mensuales del río Magdalena observados durante los meses del abril y mayo, desde 1988 a 2005 (la unidad de medida no se precisa).

AÑO	***ABRIL***	***MAYO***
1988	600	512
1989	227	211
1990	487	469
1991	560	370
1992	521	363
1993	423	272
1994	307	241
1995	390	253
1996	364	408
1997	284	233
1998	415	245
1999	255	199
2000	209	215
2001	230	297
2002	424	309
2003	528	303
2004	258	196
2005	242	166

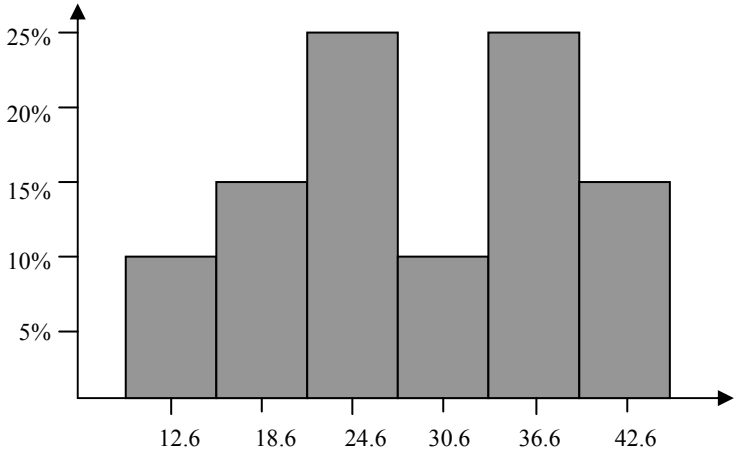
Se desea ordenar estos datos y efectuar el análisis siguiente:

a. Dar una representación global de los caudales de abril y mayo. Graficar mediante dos histogramas los datos resumidos.
b. ¿Qué comportamiento puede observar en las frecuencias en ambos meses?

3.8.9 Determine los ángulos de las porciones y complete la tabla de frecuencia TIPO A, a partir del siguiente gráfico de sectores, si el total de datos es de 99:



3.8.10 Complete la tabla de frecuencia a partir del siguiente histograma, si el total de datos es de 200:



3.9 CASO: EL PROVEEDOR DE TUBOS DE ACERO

Una importante empresa desea contratar el suministro de tubos de acero. Para la licitación se presentaron tres empresas (llámense A, B y C), las cuales venden la unidad al mismo precio y con las mismas especificaciones del material. La empresa solicita que el proveedor mantenga un diámetro promedio por cada 30 tubos entregados de 200 mm; para lo cual solicitó a cada empresa una muestra de este tamaño, obteniendo los siguientes diámetros (las unidades están en milímetros):

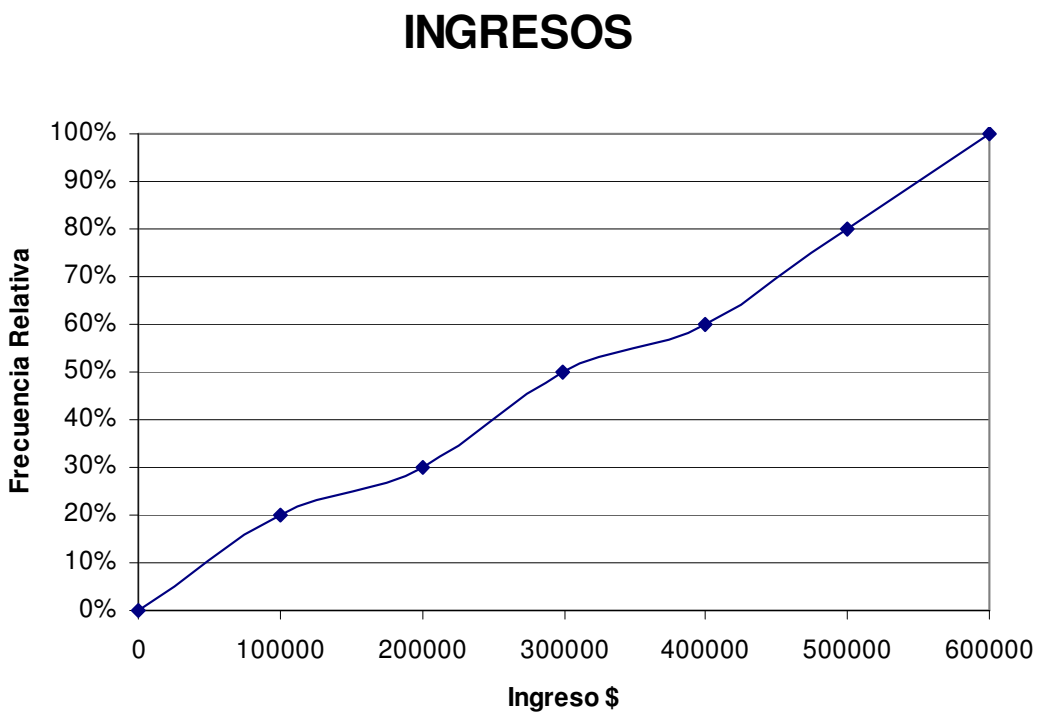
COMPAÑÍA A			**COMPAÑÍA B**			**COMPAÑÍA C**		
196	193	196	199	197	199	203	209	152
187	208	196	197	194	199	191	180	172
202	188	213	200	205	204	260	201	200
221	190	212	203	197	200	179	191	171
211	192	198	203	197	200	228	187	207
217	178	194	205	195	198	200	194	172
178	194	219	194	198	206	200	211	214
197	195	208	199	202	203	169	245	278
215	201	223	203	200	208	190	229	170
189	196	193	197	199	198	193	243	160

¿Cuál de los tres proveedores escogería usted? Justifique su respuesta mediante un análisis gráfico de los histogramas resultantes para cada compañía **(RECOMENDACIÓN:** Cree los histogramas a partir de tablas de frecuencias que empleen los mismos intervalos de clases).

3.10 CUESTIONARIO DE REPASO

Selección Múltiple con Única Respuesta: Marque con una X la respuesta correcta.

Para las preguntas 1 y 2: A partir del siguiente gráfico de Ojiva.



1. Se puede concluir:

   a. El 60% de la población gana más de $400.000
   b. El 50% equivale a un ingreso de $200.000
   c. El 80% de la población gana más de $100.000
   d. Todas las anteriores

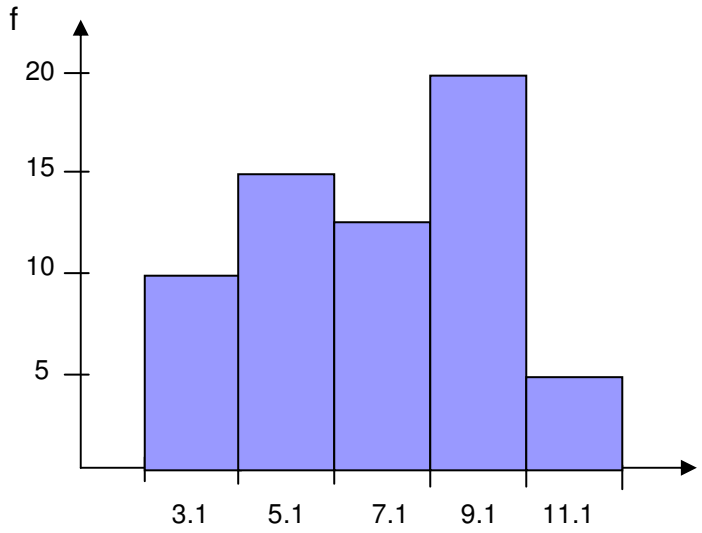
2. Se puede concluir:

   a. El 50% de la población gana más de $300.000
   b. El 30% de la población gana menos de $200.000
   c. El sueldo máximo registrado fue de $600.000
   d. Todas las anteriores

3. Los gráficos de sectores son usados para:

   a. Mostrar frecuencias acumuladas
   b. Mostrar las marcas de clase de una tabla de frecuencia tipo B
   c. Mostrar solo las frecuencias absolutas
   d. Mostrar frecuencias no acumuladas

4. Cual de las tablas de frecuencia corresponde al siguiente histograma:



a.

Ni	Lm	Ls	f	F	h	H	MC
1	2,0	4,1	10	10	15,87%	15,87%	3,1
2	4,1	6,1	15	25	23,81%	39,68%	5,1
3	6,1	8,1	13	38	20,63%	60,32%	7,1
4	8,1	10,1	20	58	31,75%	92,06%	9,1
5	10,1	12,0	5	63	7,94%	100,00%	11,1
Total			**63**		**100,00%**		

b.

Ni	Lm	Ls	f	F	h	H	MC
1	2,0	4,1	5	5	7,94%	7,94%	3,1
2	4,1	6,1	20	25	31,75%	39,68%	5,1
3	6,1	8,1	13	38	20,63%	60,32%	7,1
4	8,1	10,1	15	53	23,81%	84,13%	9,1
5	10,1	12,0	10	63	15,87%	100,00%	11,1
Total			**63**		**100,00%**		

c.

Ni	Lm	Ls	f	F	h	H	MC
1	2,0	4,1	10	10	15,63%	15,63%	3,1
2	4,1	6,1	15	25	23,44%	39,06%	5,1
3	6,1	8,1	13	38	20,31%	59,38%	7,1
4	8,1	10,1	20	58	31,25%	90,63%	9,1
5	10,1	12,1	5	63	7,81%	98,44%	11,1
6	12,1	14,0	1	64	1,56%	100,00%	13,1
Total			**64**		**100,00%**		

d. Ninguna de las anteriores

CAPITULO 4

MEDIDAS DE TENDENCIA CENTRAL

La estadística descriptiva en su función básica de reducir datos, propone una serie de indicadores que permiten tener una percepción rápida de lo que ocurre en un fenómeno.

La primera gama de indicadores corresponde a las “Medidas de Tendencia Central”. Existen varios procedimientos para expresar matemáticamente las medidas de tendencia central, de los cuales, los más conocidos son: la media aritmética, la moda y la mediana.

CAPITULO 4: MEDIDAS DE TENDENCIA CENTRAL

Medidas de tendencia central: Son indicadores estadísticos que muestran hacia que valor (o valores) se agrupan los datos.

Esta primera parte la dedicaremos a analizar tres medidas de tendencia central:

- La media aritmética
- La moda
- La mediana

En el suplemento de este capitulo incluiremos otras medidas de tendencia central.

4.1 LA MEDIA ARITMÉTICA

Equivale al cálculo del promedio simple de un conjunto de datos. Para diferenciar datos muestrales de datos poblacionales, la media aritmética se representa con un símbolo para cada uno de ellos: si trabajamos con la población, este indicador será μ; en el caso de que estemos trabajando con una muestra, el símbolo será $\overline{X}$.

Media aritmética (μ o $\overline{X}$): Es el valor resultante que se obtiene al dividir la sumatoria de un conjunto de datos sobre el número total de datos. Solo es aplicable para el tratamiento de datos cuantitativos.

Hay que entender que existen dos formas distintas de trabajar con los datos tanto poblacionales como muestrales: sin agruparlos o agrupándolos en tablas de frecuencias. Esta apreciación nos sugiere dos formas de representar la media aritmética.

4.1.1 Media aritmética para datos no agrupados

Podemos diferenciar la fórmula del promedio simple para datos poblaciones y muestrales:

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

Población

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

Muestra

Observe que la variación de ambas fórmulas radica en el tamaño de los datos (***N*** identifica el tamaño de la población, mientras que ***n*** el de la muestra).

4.1.2 Ejemplo: la media aritmética para datos no agrupados

El profesor de la materia de estadística desea conocer el promedio de las notas finales de los 10 alumnos de la clase. Las notas de los alumnos son:

3,2	3,1	2,4	4,0	3,5
3,0	3,5	3,8	4,2	4,0

¿Cuál es el promedio de notas de los alumnos de la clase?

SOLUCIÓN

Aplicando la fórmula para datos no agrupados tenemos:

$$\mu = \frac{3,2 + 3,1 + 2,4 + 4,0 + 3,5 + 3,0 + 3,5 + 3,8 +,4,2 + 4,0}{10} = \frac{34,7}{10}$$

$$\mu = 3,47$$

Cabe anotar que en el ejemplo estamos hablando de una población correspondiente a todos los alumnos de la clase (10 alumnos en total). El promedio de las notas es de 3,47.

Modifiquemos la primera nota por 0,0 y calculemos nuevamente la media aritmética.

$$\mu = \frac{0,0 + 3,1 + 2,4 + 4,0 + 3,5 + 3,0 + 3,5 + 3,8 +,4,2 + 4,0}{10} = \frac{31,5}{10}$$

$$\mu = 3,15$$

En este caso la media pasa de 3,47 a 3,15. Esta variación notoria se debió a que la media aritmética es sensible a los valores extremos cuando tratamos con pocos datos. El 0,0 es una nota atípica comparada con las demás, que están ubicadas entre 3,0 y 4,2.

4.1.3 Media aritmética para datos agrupados

En el capitulo 2 explicábamos dos tipos de tablas de frecuencias (A y B). Cuando los datos se agrupan en tablas tipo A, la media aritmética es igual a la división de la sumatoria del producto de las clases por la frecuencia sobre el número de datos.

$$\mu = \frac{\sum_{i=1}^{Nc} X_i f_i}{N}$$

$$\overline{X} = \frac{\sum_{i=1}^{Nc} X_i f_i}{n}$$

La sumatoria parte desde el primer intervalo de clase (i = 1) hasta el último (Nc), siendo X_i la clase del intervalo i.

Cuando los datos se agrupan en tablas de frecuencias tipo B, el cálculo de la media varía un poco, ya que existe una pérdida de información en el momento en que se trabaja con intervalos de frecuencia y no con los datos directamente (los datos se agrupan por intervalo, desconociendo el valor exacto de cada uno de ellos).

$$\mu = \frac{\sum_{i=1}^{Nc} Mc_i f_i}{N}$$

Población

$$\overline{X} = \frac{\sum_{i=1}^{Nc} Mc_i f_i}{n}$$

Muestra

Las marcas de clases (Mc) cumple la función de representar los intervalos de clase.

4.1.4 Ejemplo: media aritmética para datos agrupados en tablas tipo A

La siguiente tabla de frecuencia muestra el número de preguntas de 81 encuestados sobre un Test que consta de solo seis preguntas.

Preguntas Buenas	Personas
1	15
2	13
3	8
4	19
5	21
6	5

SOLUCIÓN

PASO 1: Realizar la sumatoria del producto resultante de las clases por su frecuencia absoluta. Para efectos del cálculo de la media, deberíamos sumar 15 veces el valor 1, 13 veces el valor 2, 8 veces el valor 3, hasta llegar a la última clase:

$$\sum_{i=1}^{Nc} X_i f_i = 1x15 + 2x13 + 3x8 + 4x19 + 5x21 + 6x5 = 276$$

PASO 2: Dividir la sumatoria sobre el número total de datos.

$$\overline{X} = \frac{\sum_{i=1}^{Nc} X_i f_i}{n} = \frac{276}{81}$$

$$\overline{X} = 3{,}41$$

En promedio los encuestados contestaron aproximadamente 3 (el valor exacto es 3,41) preguntas buenas.

4.1.5 Ejemplo: media aritmética para datos agrupados en tablas tipo B

Calcular la media para los datos distribuidos en la siguiente tabla de frecuencia:

Ni	Lm	Ls	f	Mc
1	40,0	48,1	3	44,1
2	48,1	56,1	8	52,1
3	56,1	64,1	11	60,1
4	64,1	72,1	32	68,1
5	72,1	80,1	21	76,1
6	80,1	88,1	18	84,1
7	88,1	96,1	14	92,1
8	96,1	104,0	1	100,1

SOLUCIÓN

Las marcas de clase representan a los intervalos de clase, por ejemplo, suponemos que la marca de clase para el primer intervalo (44,1) se repite 3 veces, al desconocer los 3 valores exactos que están dentro de dicho intervalo.

PASO 1: Realizar la sumatoria del producto resultante entre las marcas de clase por su frecuencia absoluta.

- $\sum_{i=1}^{Nc} Mc_i f_i = 44{,}1x3 + 52{,}1x8 + 60{,}1x11 + 68{,}1x32 + 76{,}1x21 + 84{,}1x18 + 92{,}1x14 + 100{,}1x1$

$$\sum_{i=1}^{Nc} Mc_i f_i = 7890{,}6$$

PASO 2: Dividir la sumatoria sobre el número total de datos.

$$\overline{X} = \frac{\sum_{i=1}^{Nc} Mc_i f_i}{n} = \frac{7890{,}6}{108}$$

$$\overline{X} = 73{,}1$$

4.1.6 Ejemplo: comparativa entre el cálculo de la media aritmética para datos no agrupados y datos agrupados en tablas tipo B

Calcular la media aritmética a los siguientes datos sin agrupar y agrupándolos en una tabla de frecuencia tipo B (suponga que los datos son poblacionales):

47,8	23,1	12,4	35,4	44,0	26,2
18,6	11,0	32,0	12,4	49,4	41,4
18,6	21,0	26,3	11,1	21,4	30,6
12,8	43,1	18,1	38,1	16,8	12,4
33,6	40,9	15,2	33,2	48,2	37,0

SOLUCIÓN

Calculemos la media para los datos sin agrupar:

$$\mu = \frac{47{,}8 + 23{,}1 + 12{,}4 + 35{,}4 + 44{,}0 + 26{,}2 + ... + 37{,}0}{30} = \frac{832{,}1}{30}$$

$$\mu = 27{,}74$$

Luego construyamos la tabla tipo B y calculemos su media aritmética con el fin de comparar ambos resultados:

Ni	Lm	Ls	f	Mc
1	11,00	17,41	8	14,21
2	17,41	23,81	6	20,61
3	23,81	30,21	2	27,01
4	30,21	36,61	5	33,41
5	36,61	43,01	4	39,81
6	43,01	49,40	5	46,21
Total			**30**	

PASO 1: Realizar la sumatoria del producto resultante entre las marcas de clase por su frecuencia absoluta.

$$\sum_{i=1}^{Nc} Mc_i f_i = 14{,}21x8 + 20{,}61x6 + 27{,}01x2 + 33{,}41x5 + 39{,}81x4 + 46{,}21x5 = 848{,}70$$

PASO 2: Dividir la sumatoria sobre el número total de datos.

$$\overline{X} = \frac{\sum_{i=1}^{Nc} Mc_i f_i}{n} = \frac{848{,}70}{30}$$

$$\overline{X} = 28{,}29$$

Podemos ver claramente una diferencia entre ambas medias: 27,74 para los datos no agrupados y 28,29 para los datos agrupados. Esta diferencia radica que en la tabla tipo B existe una perdida de información, al agrupar los datos en los intervalos de clase. El valor de la media exacta es el calculado para los datos no agrupados, pero dada la proximidad de la media para los datos agrupados, se tomar esta última como cierta.

4.1.7 Cálculo de la media aritmética en Excel

Excel presenta la función **PROMEDIO** para el cálculo de la media aritmética:

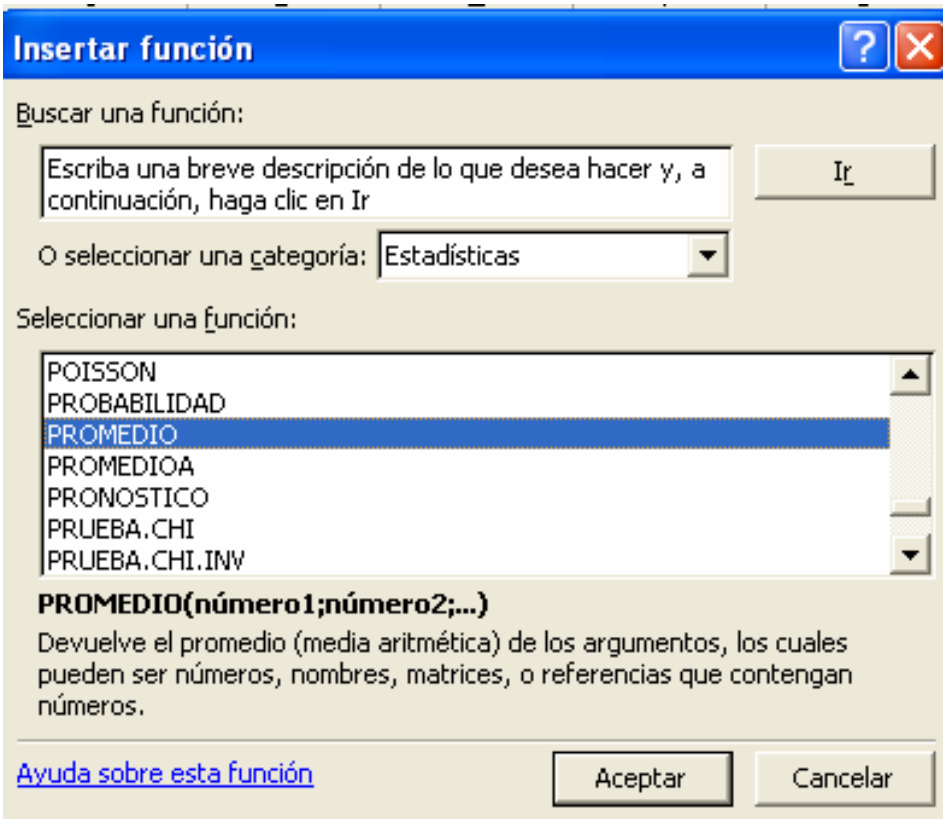
PROMEDIO: Permite calcular la media aritmética (o promedio simple) de un conjunto de datos.

Formato: PROMEDIO(número1;número2;…)
Categoría: Estadísticas

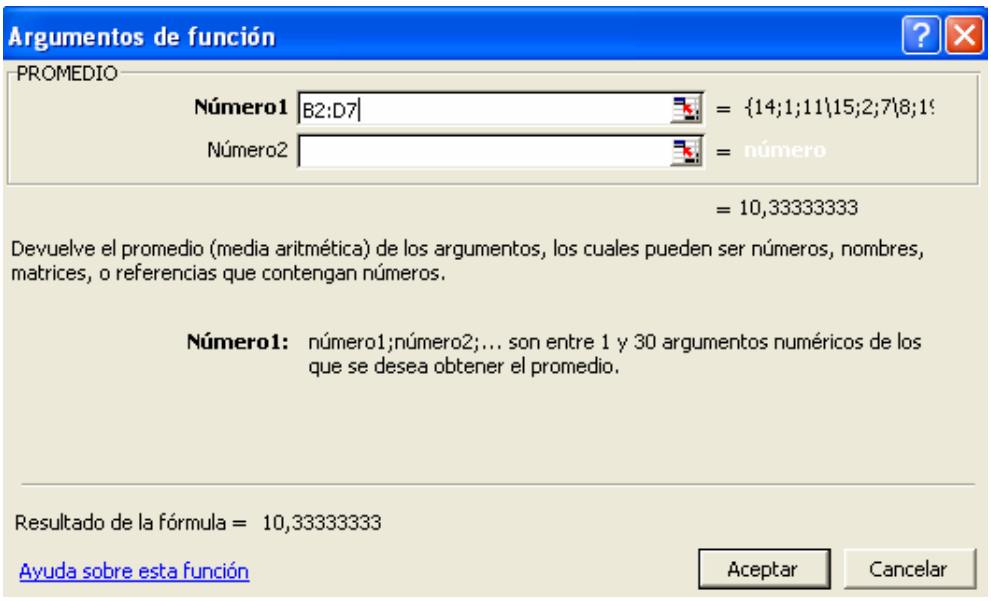
En una hoja nueva, copie los siguientes datos a partir de la celda B2:

	A	B	C	D	E
1					
2		14	1	11	
3		15	2	7	
4		8	19	8	
5		11	14	15	
6		6	13	13	
7		3	12	14	
8					
9					

Ubiquémonos en la celda B9 y activemos la venta de funciones, seleccionando la función **PROMEDIO**:



En la primera casilla (número 1), seleccionamos el conjunto de datos:



Pulsemos en el botón Aceptar para mostrar el resultado en la celda B9.

	A	B	C	D	E
1					
2		14	1	11	
3		15	2	7	
4		8	19	8	
5		11	14	15	
6		6	13	13	
7		3	12	14	
8					
9		10,3333333	←	=PROMEDIO(B2:D7)	
10					

El procedimiento varía cuando tenemos tablas de frecuencia. Copie la siguiente tabla en una hoja nueva a partir de la celda B2:

	A	B	C	D
1				
2		**Clase**	**Frecuencia**	
3		1	12	
4		2	15	
5		3	23	
6		4	21	
7		5	40	
8		6	5	
9		**Total**	**116**	
10				
11				

Recordemos que el primer paso es calcular la sumatoria del producto entre clase y frecuencia, empleando la siguiente función:

SUMAPRODUCTO: Calcula la suma de los productos entre datos.

Formato: SUMAPRODUCTO(matriz1;matriz2;matriz3;…)
Categoría: Matemáticas y trigonométricas

Activemos esta función desde la celda B11, considerando al campo matriz 1 como las clases y matriz 2 como las frecuencias.

Al pulsar en Aceptar, tendremos el valor de la sumatoria.

	A	B	C	D
1				
2		**Clase**	**Frecuencia**	
3		1	12	
4		2	15	
5		3	23	
6		4	21	
7		5	40	
8		6	5	
9		**Total**	**116**	
10				
11		425		
12				

=SUMAPRODUCTO(B3:B8;C3:C8)

Necesitamos ahora dividir el resultado de la sumatoria sobre los 116 datos incluidos en el ejercicio. Modifiquemos la fórmula actual y agreguemos:

f_x =SUMAPRODUCTO(B3:B8;C3:C8)/C9

Donde C9 es la celda que muestra el total de los datos. El resultado final es 3,6637931.

	A	B	C	D
1				
2		**Clase**	**Frecuencia**	
3		1	12	
4		2	15	
5		3	23	
6		4	21	
7		5	40	
8		6	5	
9		**Total**	**116**	
10				
11		3,6637931		
12				

4.1.8 Ventajas

- Es la medida de tendencia central más usada.
- El promedio es estable en el muestreo.
- Es sensible a cualquier cambio en los datos (puede ser usado como un detector de variaciones en los datos).
- Se emplea a menudo en cálculos estadísticos posteriores.
- Presenta rigor matemático.
- En la gráfica de frecuencia representa el centro de gravedad.

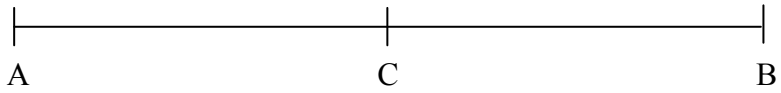
4.1.9 Desventajas

- Es sensible a los valores extremos.
- No es recomendable emplearla en distribuciones muy asimétricas.
- Si se emplean variables discretas o cuasi-cualitativas, la media aritmética puede no pertenecer al conjunto de valores de la variable.

4.2 LA MEDIANA

Mediana (Me): Valor que divide una serie de datos en dos partes iguales. La cantidad de datos que queda por debajo y por arriba de la mediana son iguales.

La definición de geométrica se refiere al punto que divide en dos partes a un segmento. Por ejemplo, la mediana del segmento $\overline{AB}$ es el punto C.



Existen entonces dos segmentos iguales:

$$\overline{AC} = \overline{CB}$$

4.2.1 Ejemplo: mediana para datos no agrupados (cantidad de datos impar)

Encontrar la mediana para los siguientes datos:

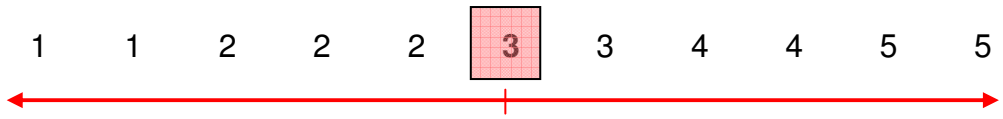
4 1 2 3 4 2 2 1 5 5 3

SOLUCIÓN

PASO 1: Ordenar los datos.

1 1 2 2 2 3 3 4 4 5 5

PASO 2: Localizar el valor que divide en dos parte iguales el número de datos.



La mediana es 3, dejando 5 datos a cada lado.

$$Me = 3$$

4.2.2 Ejemplo: mediana para datos no agrupados (cantidad de datos par)

Modifiquemos el ejemplo anterior, eliminando el último dato. Encontrar la mediana:

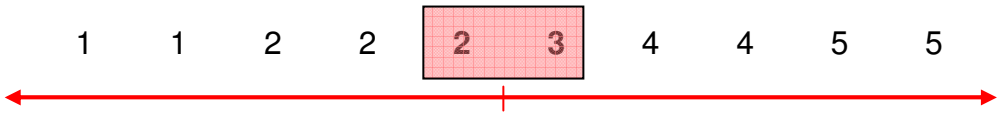
4 1 2 3 4 2 2 1 5 5

SOLUCIÓN

PASO 1: Ordenar los datos.

1 1 2 2 2 3 4 4 5 5

PASO 2: Localizar el valor que divide en dos parte iguales el número de datos.



El punto medio se encuentra entre dos valores: 2 y 3, por tanto, el valor de la mediana será 2,5.

$$Me = 2{,}5$$

4.2.3 Ejemplo: mediana para datos agrupados en tablas tipo A

Calcular la mediana a partir de la siguiente tabla de frecuencia:

Ni	Clase	f	F	h	H
1	10	5	5	10,4%	10,4%
2	20	7	12	14,6%	25,0%
3	30	10	22	20,8%	45,8%
4	40	13	35	27,1%	72,9%
5	50	10	45	20,8%	93,8%
6	60	2	47	4,2%	97,9%
7	70	1	48	2,1%	100,0%
Total		**48**		**100,0%**	

SOLUCIÓN

PASO 1: Localizar entre que clases se encuentra la mediana. Observe que la mediana se encuentra entre las clases 3 y 4, donde podremos encontrar una frecuencia relativa acumulada del 50%.

Ni	Clase	f	F	h	H
1	10	5	5	10,4%	10,4%
2	20	7	12	14,6%	25,0%
3	30	10	22	20,8%	45,8%
4	40	13	35	27,1%	72,9%
5	50	10	45	20,8%	93,8%
6	60	2			
7	70	1			
Total		**48**			

Entre las clases 3 y 4 se encuentra el punto que divide en dos partes iguales la cantidad de datos.

PASO 2: Interpolar los datos para encontrar la mediana.

En el paso anterior habíamos dicho que el punto que divide el 2 parte iguales se encuentra entre 30 y 40.

	Clase	H
	40	72,9%
	30	45,8%
Diferencia	**10**	**27,1%**

La diferencia entre las frecuencias relativas nos indica que existe entre las clases 27,1% de los datos. Para llegar al 50% de los datos, debemos incrementar en 4,2% datos partiendo desde la clase 30.

$$50{,}0\% = 45{,}8\% + 4{,}2\%$$

Con una regla de tres sencilla hallaremos el incremento en unidades dada en la clase para ese 4,2%.

$$\begin{array}{ccc} 10 & \longrightarrow & 27{,}1\% \\ \text{Incremento} & \longrightarrow & 4{,}2\% \end{array}$$

$$Incremento = \frac{4{,}2\% x 10}{27{,}1\%} = 1{,}55$$

Para llegar al 50% de los datos, a la clase 30 debemos incrementarle 1,55.

$$Me = 31{,}55$$

4.2.4 Ejemplo: mediana para datos agrupados en tablas tipo B

Determinar la mediana de la siguiente tabla de frecuencia:

Ni	Lm	Ls	f	F	h	H	Mc
1	21,20	29,21	5	5	12,50%	12,50%	25,21
2	29,21	37,21	2	7	5,00%	17,50%	33,21
3	37,21	45,21	10	17	25,00%	42,50%	41,21
4	45,21	53,21	7	24	17,50%	60,00%	49,21
5	53,21	61,21	12	36	30,00%	90,00%	57,21
6	61,21	69,21	3	39	7,50%	97,50%	65,21
7	69,21	77,20	1	40	2,50%	100,00%	73,21
Total			**40**		**100,00%**		

SOLUCIÓN

PASO 1: Localizar entre que intervalos de clase se encuentra la mediana.

Podemos observar que el punto que divide el 50% de los datos esta entre el intervalo de clase 3 y 5, para ser más preciso, entre los valores 45,21 y 53,21 (hasta 45,21 hay agrupados el 42,50% de los datos, y hasta 53,21 se resume el 60,00% de los datos).

Ni	Lm	Ls	f	F	h	H	Mc
1	21,20	29,21	5	5	12,50%	12,50%	25,21
2	29,21	37,21	2	7	5,00%	17,50%	33,21
3	37,21	45,21	10	17	25,00%	42,50%	41,21
4	45,21	53,21	7	24	17,50%	60,00%	49,21
5	53,21	61,21	12	36	30,00%	90,00%	57,21
6	61,21	69,21				97,50%	65,21
7	69,21	77,20				100,00%	73,21
Total							

En el intervalo 4 se encuentra el punto que divide en dos partes iguales el total de los datos.

PASO 2: Interpolar los datos para encontrar la mediana. En resumen tenemos que:

	Límite Superior	H
	53,21 (Ls_4)	60,00% (H_4)
	45,21 (Ls_3)	42,50% (H_3)
Diferencia	**8,00**	**17.50%**

Entre los dos límites superiores abarcan un total de 17,50% de los datos. Se debe aumentar en 7,50% los datos desde límite superior del tercer intervalo de clase.

$$\begin{array}{ccc} 8{,}00 & \longrightarrow & 17{,}50\% \\ \text{Incremento} & \longrightarrow & 7{,}50\% \end{array}$$

$$Incremento = \frac{7{,}50\% x 8{,}00}{17{,}50\%} = 3{,}43$$

Para llegar al 50% de los datos, 45,21 se aumenta en 3,43 unidades.

$$Me = 45{,}21 + 3{,}43$$

$$Me = 48{,}64$$

4.2.5 La fórmula para calcular la mediana

De este último ejemplo podemos determinar la fórmula para calcular la mediana. Observe que la mediana parte del límite superior del intervalo de clase anterior, la cual simbolizaremos por Ls_{i-1}, siendo i igual a 4 (cuarto intervalo de clase). A este valor se le suma el incremento para llegar al 50% de los datos:

$$Me = Ls_{i-1} + Incremento$$

El incremento resulta de multiplicar el incremento para llevar la frecuencia al 50% (50% - H_{i-1}) por el ancho de la clase (A) sobre la diferencia porcentual entre los límites superiores ($H_i - H_{i-1}$):

$$Me = Ls_{i-1} + A.\frac{(50\% - H_{i-1})}{(H_i - H_{i-1})}$$

Simplificando aún más la fórmula, recordemos que $H_i - H_{i-1}$ es lo mismo la frecuencia relativa del intervalo de clase i (h_i).

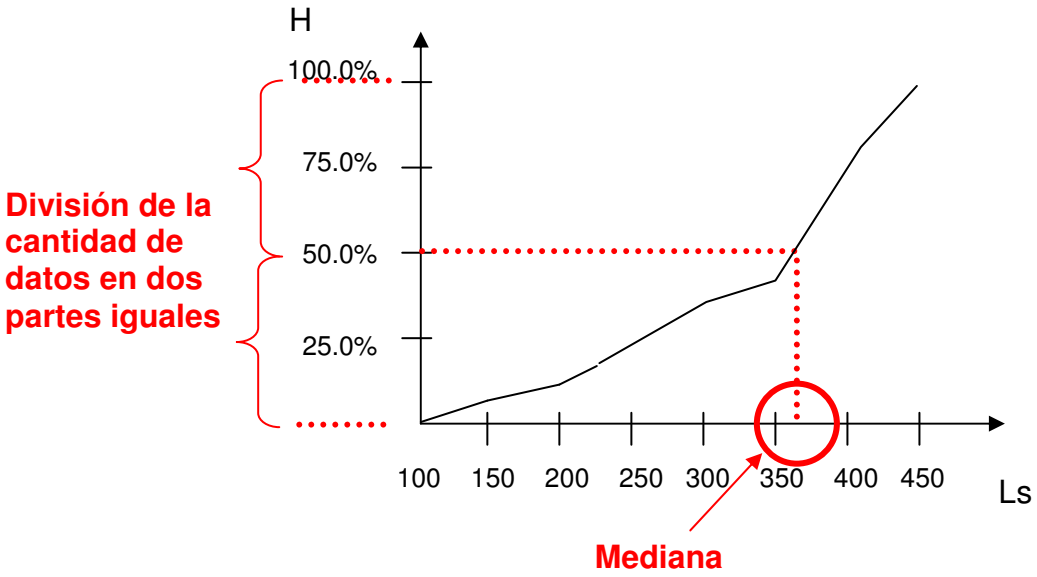
$$Me = Ls_{i-1} + A.\frac{(50\% - H_{i-1})}{h_i}$$

Para expresar la fórmula en frecuencias absolutas tenemos que:

$$Me = Ls_{i-1} + A.\frac{(\frac{n}{2} - F_{i-1})}{f_i}$$

4.2.6 Ubicando la mediana en el gráfico de ojiva

En un gráfico de ojiva, la mediana corresponde a la proyección del punto en eje horizontal que equivale al 50% de los datos. En la el gráfico de ojiva del ejemplo 3.6.1, la mediana estaría ubicada en el sexto intervalo, entre 350 y 400:



4.2.7 Calculo de la mediana en Excel

Excel posee la función **MEDIANA** para el cálculo de la mediana en datos no agrupados.

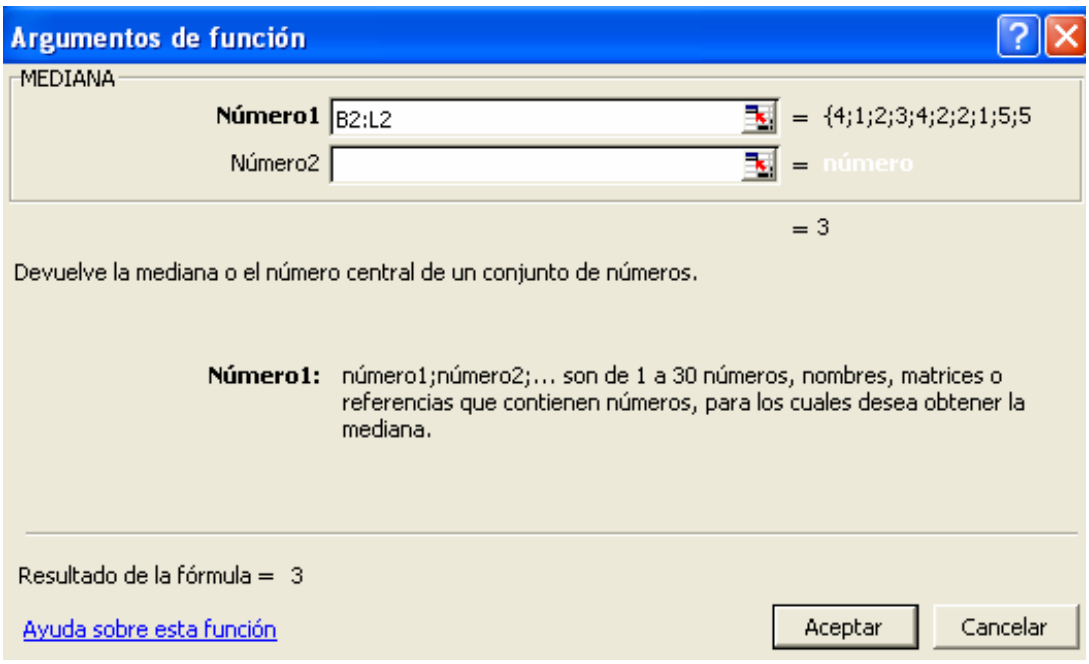
MEDIANA: Calcula la mediana para una serie de datos.

Formato: MEDIANA(número1;número2;…)
Categoría: Estadísticas

Copie los datos dados en el ejemplo 4.2.1 a partir de la celda B2:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2		4	1	2	3	4	2	2	1	5	5	3	
3													

Active la función MEDIANA desde la celda B4 y en el campo número1 selecciones los datos del ejercicio.



La mediana en este caso es 3:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2		4	1	2	3	4	2	2	1	5	5	3	
3													
4		3	←	=MEDIANA(B2:L2)									
5													

4.2.8 Ventajas

- Es estable a los valores extremos.
- Es recomendable para distribuciones muy asimétricas.

4.2.9 Desventajas

- No presenta todo el rigor matemático.
- Se emplea solo en variables cuantitativas.

4.3 LA MODA

Moda (Mo): indica el valor que más se repite, o la clase que posee mayor frecuencia.

En el caso de que dos valores presenten la misma frecuencia, decimos que existe un conjunto de datos bimodal. Para más de dos modas hablaremos de un conjunto de datos multimodal.

4.3.1 Ejemplo: moda para datos no agrupados

Los siguientes datos provienen del resultado de entrevistar a 30 personas sobre la marca de gaseosa que más consume a la semana:

Marca 1	Marca 2	Marca 1	Marca 1	Marca 1	Marca 3
Marca 1	Marca 3	Marca 1	Marca 2	Marca 1	Marca 1
Marca 2	Marca 1	Marca 3	Marca 3	Marca 2	Marca 1
Marca 1	Marca 1	Marca 1	Marca 3	Marca 1	Marca 2
Marca 3	Marca 1	Marca 3	Marca 3	Marca 2	Marca 3

SOLUCIÓN

PASO 1: Determinar las frecuencias de cada valor de la variable.

La marca 1 se repite 15 veces
La marca 2 se repite 6 veces
La marca 3 se repite 9 veces

PASO 2: la moda representa el valor que más se repite. En este caso es la marca 1.

$$Mo = \text{Marca 1}$$

4.3.2 Ejemplo: moda para datos agrupados

Calcular la moda a partir de la siguiente tabla de frecuencia:

Ni	Lm	Ls	f	Mc
1	[4	6)	2	5
2	[6	8)	4	7
3	[8	10)	4	9
4	[10	12)	5	11
5	[12	14]	5	13
	Total		**20**	

SOLUCIÓN

Las marcas de clase que más frecuencias tienen son 11 y 13, por tanto decimos que es un caso donde aparecen dos modas (bimodal).

$$Mo_1 = 11$$

$$Mo_2 = 13$$

4.3.3 Calculo de la moda mediante fórmula

Algunos autores suelen aplicar una fórmula para determinar la moda para tablas de frecuencia.

$$Mo = L_{S-1} + A.\frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i-1})}$$

Donde L_{S-1} equivale al límite superior del intervalo anterior donde se encuentra la moda.

4.3.4 Calculo de la mediana en Excel

Con la función MODA que provee Excel, podremos calcular el valor que posee mayor frecuencia en datos no agrupados.

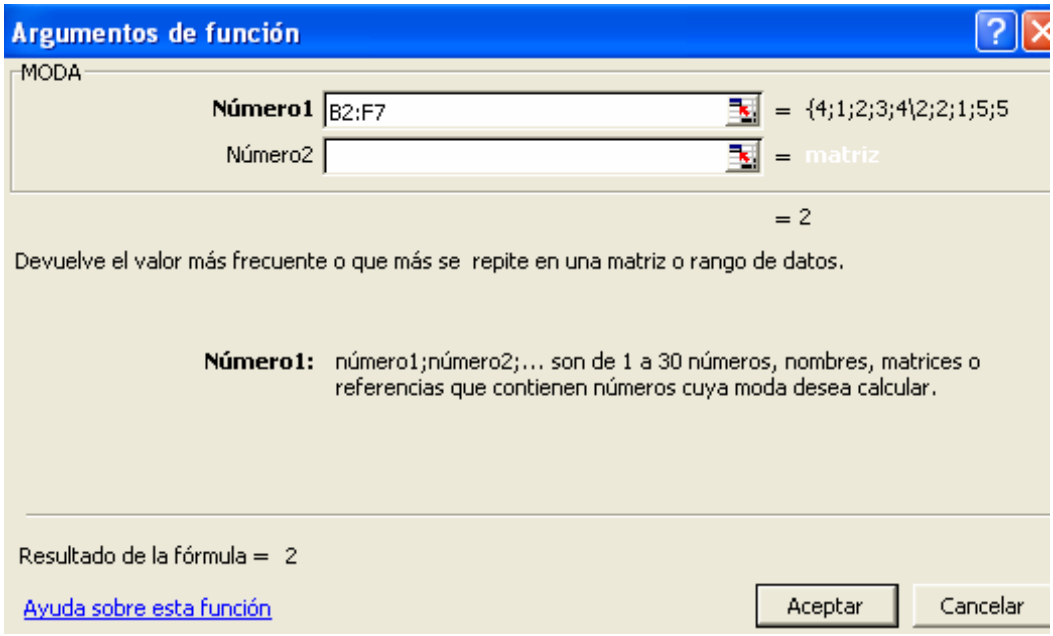
MODA: Determina el valor que más se repite en un conjunto de datos.

Formato: MODA(número1;número2;…)
Categoría: Estadísticas

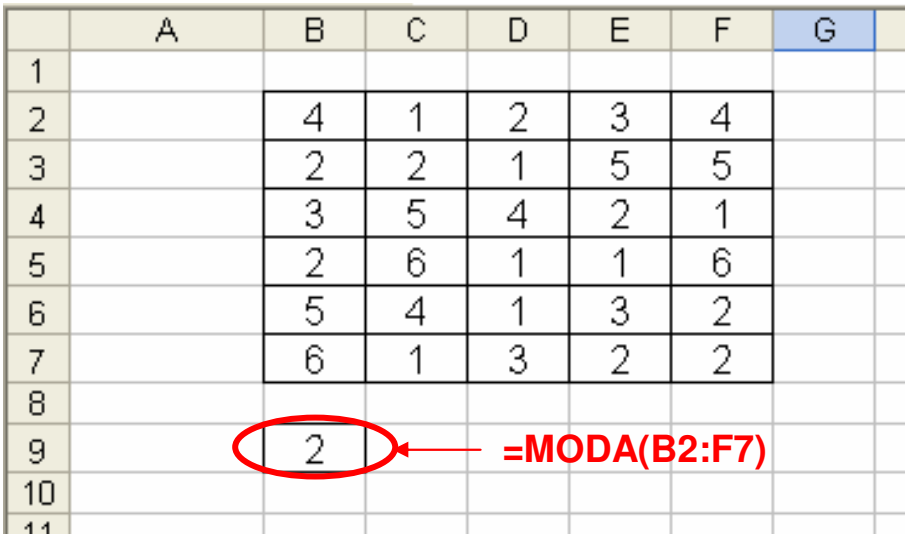
Calcule la moda a partir de los siguientes datos copiados en una hoja nueva de Excel:

	A	B	C	D	E	F	G
1							
2		4	1	2	3	4	
3		2	2	1	5	5	
4		3	5	4	2	1	
5		2	6	1	1	6	
6		5	4	1	3	2	
7		6	1	3	2	2	
8							

Active la función MODA en la celda B9 y en el campo número1 selecciones los datos del ejercicio.



La moda del ejercicio es 2.



Esta fórmula solo muestra una moda, correspondiente a la de menor valor. En el caso de que no exista la moda aparecen los símbolos **#N/A**.

4.3.5 Ventajas

- Es estable a los valores extremos.
- Es recomendable para el tratamiento de variables cualitativas.

4.3.6 Desventajas

- Pueda que no se presente.
- Puede existir más de una moda.
- En distribuciones muy asimétricas suele ser un dato muy poco representativo.
- Carece de rigor matemático.

4.4 EJERCICIOS PROPUESTOS

4.4.1 Calcular la media, mediana y moda para los siguientes datos:

11	5	4
8	9	8
6	11	3
7	10	2
7	3	8

4.4.2 Determinar la media, mediana y moda a la siguiente tabla de frecuencia:

Ni	Lm	Ls	f
1	100,0	150,1	1
2	150,1	200,1	2
3	200,1	250,1	15
4	250,1	300,1	16
5	300,1	350,1	21
6	350,1	400,1	14
7	400,1	450,1	11
8	450,1	500,0	7
Total			**87**

4.4.3 Para que un producto sea aceptado por su cliente principal, debe cumplir con ciertas especificaciones de calidad. Una de ellas, radica en que el promedio de longitud de los 20 primeros productos este entre 20,0 y 20,9 centímetros. Si las medidas son:

22,3	20,4	19,8	19,9	20,1	20,8	21,6	19,8	20,5	23,4
19,6	21,5	18,5	18,7	20,9	21,1	20,1	21,5	22,3	17,9

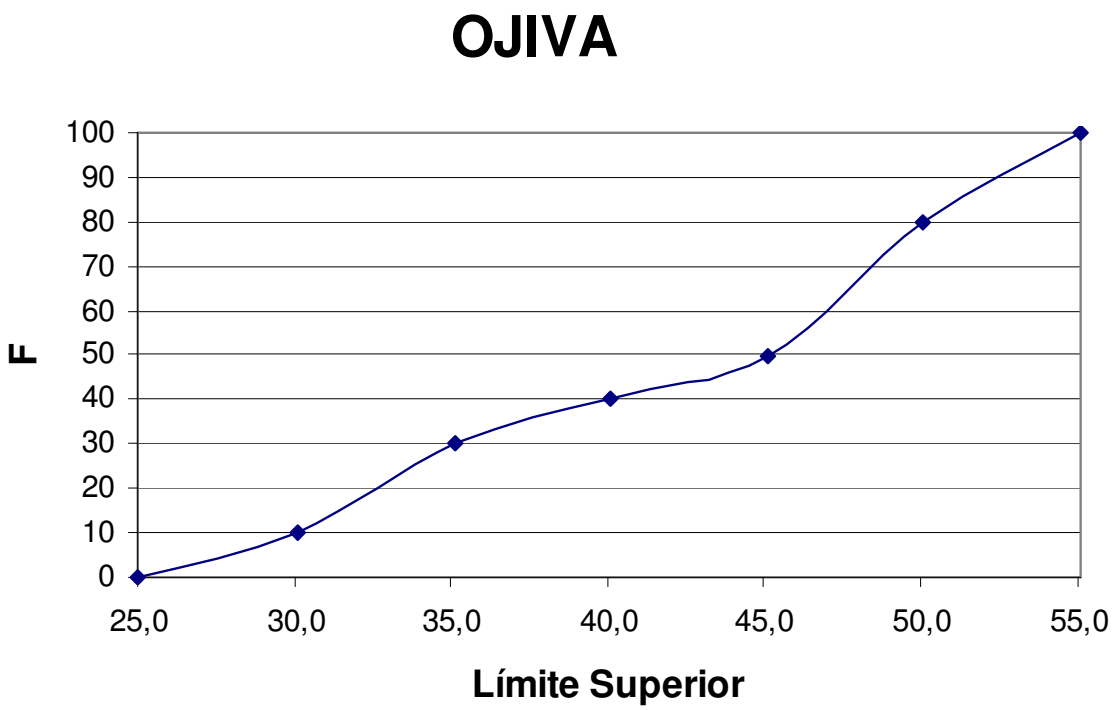
¿Cumple en el proveedor con las especificaciones del cliente?

4.4.4 Calcular la media, mediana y moda para los siguientes datos (agrúpelos en una tabla de frecuencia):

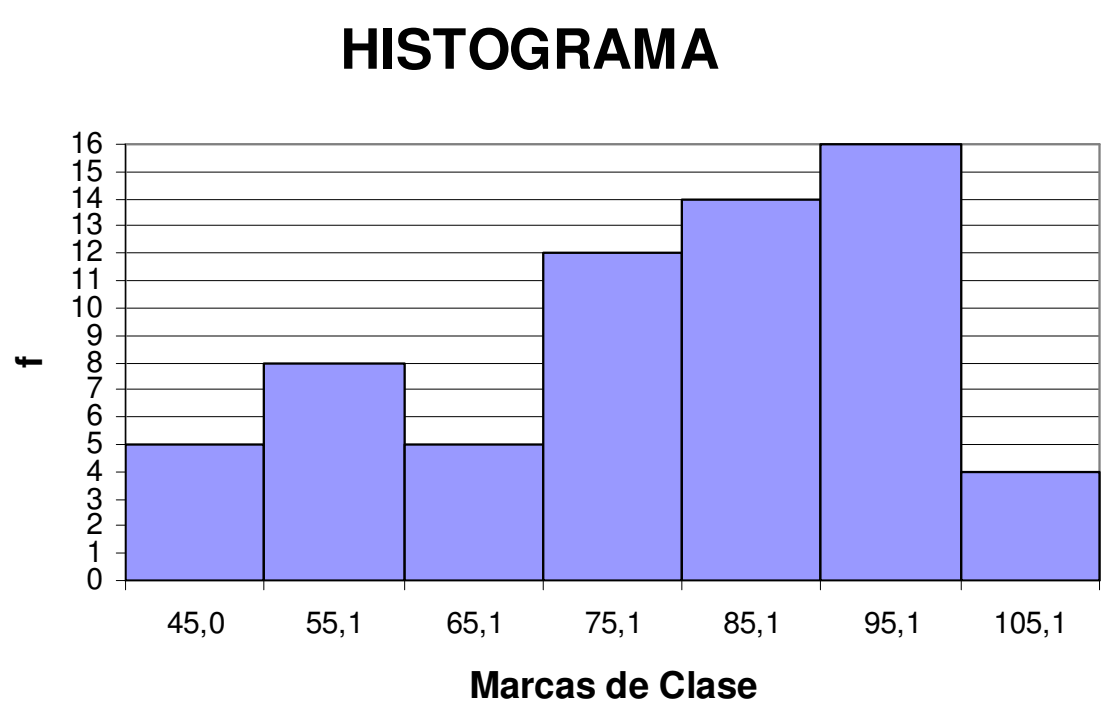
22,1	44,4	32,1
56,0	29,4	37,7
32,3	29,0	30,5
45,3	20,7	15,6
41,1	41,2	39,5
20,8	34,1	31,8
21,9	47,0	25,6

4.4.5 Calcular la media, mediana y moda de la tabla de frecuencia dada en el **ejercicio 2.3.10**.

4.4.6 Calcule y ubique la media, mediana y moda en el siguiente gráfico de ojiva:



4.4.7 Calcule la media, mediana y moda a partir del siguiente histograma:



4.5 CASO: POBLACIÓN Y MUESTRA

Los ingresos en dólares de 30 hombres elegidos al azar (entre un total de 1000) se muestran a continuación:

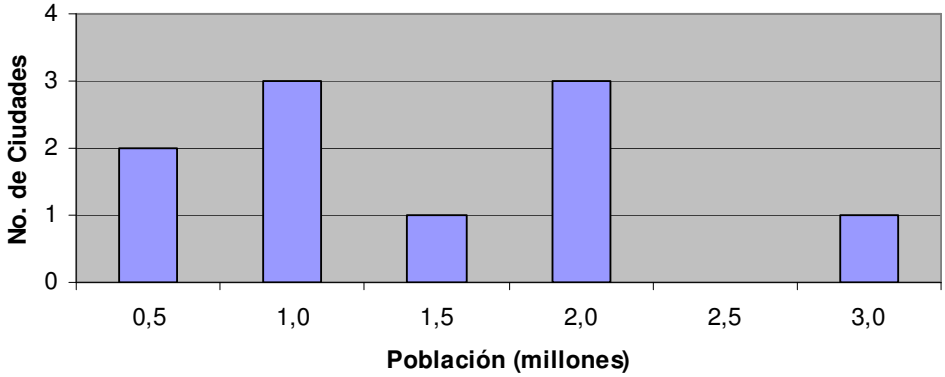
45,16	79,85	76,91
88,91	62,59	88,61
68,89	54,33	16,60
19,92	19,48	6,37
58,42	56,70	37,25
83,61	22,07	65,73
99,49	34,20	41,50
92,22	53,20	62,59
58,00	77,41	47,10
42,16	91,46	45,40

a. Calcule la media aritmética para todos los datos sin agruparlos.
b. Calcule la media aritmética empleando la tabla de frecuencias.
c. ¿Cuál cree usted son las razones de las diferencias entre ambas medias?
d. ¿Explique mediante este ejemplo, la diferencia entre media, mediana y moda?
e. ¿Qué representa para usted la moda y mediana (en termino de pesos)?
f. ¿Se puede considera que la población de 1000 personas tendrán la misma media que la muestra de 30 personas?

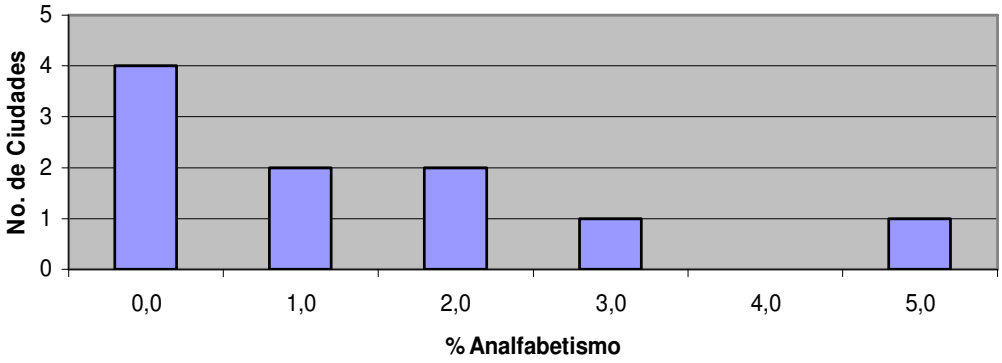
4.6 CUESTIONARIO DE REPASO

Para las preguntas 1 a 4: Se muestran los histogramas como resultado de mediciones realizadas a 10 ciudades de un país. El primer histograma muestra las poblaciones de las ciudades (unidades dadas en millones), indicando que solo una ciudad alcanza los 3'000.000 de habitantes. El histograma siguiente muestra el porcentaje de analfabetismo de las ciudades objeto del estudio.

POBLACIÓN



ANALFABETISMO



1. El total de ciudades consideradas en el estudio es de:

   a. 3
   b. 5
   c. 10
   d. 30

2. Que significado tiene la moda para el estudio del analfabetismo

   a. Cuatro de las ciudades no presentan analfabetismo
   b. La mayoría de las ciudades no presentan analfabetismo
   c. Seis ciudades presentan problemas de analfabetismo
   d. Ninguna de las anteriores

3. El porcentaje promedio de analfabetismo que arroja el estudio es de:

a. 0,4%
b. 1,0%
c. 1,4%
d. 2,0%

4. El estudio arrojado al número de habitante por ciudad indica que:

a. El promedio de habitantes por ciudad es de 0,5 millones.
b. El promedio de habitantes por ciudad es de 1,0 millón.
c. El promedio de habitantes por ciudad es de 1,45 millones.
d. El promedio de habitantes por ciudad es de 2,0 millones.

CAPITULO 5

MEDIDAS DE DISPERSIÓN

No solo basta con determinar las medidas de tendencia central para comprender el comportamiento de una serie de datos, es importante además, conocer que tan alejados están esos datos respecto a ese punto de concentración.

Las medidas de dispersión nos indican la distancia promedio de los datos respecto a las medidas de tendencia central. Así podremos diferenciar dos conjuntos de datos que poseen iguales medias, siendo los datos de uno más dispersos del otro.

CAPITULO 5: MEDIDAS DE DISPERSIÓN

Medidas de dispersión: Son indicadores estadísticos que muestran la distancia promedio que existe entre los datos y la media aritmética.

En el estudio de las medidas de dispersión daremos un vistazo a cuatro indicadores básicos:

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

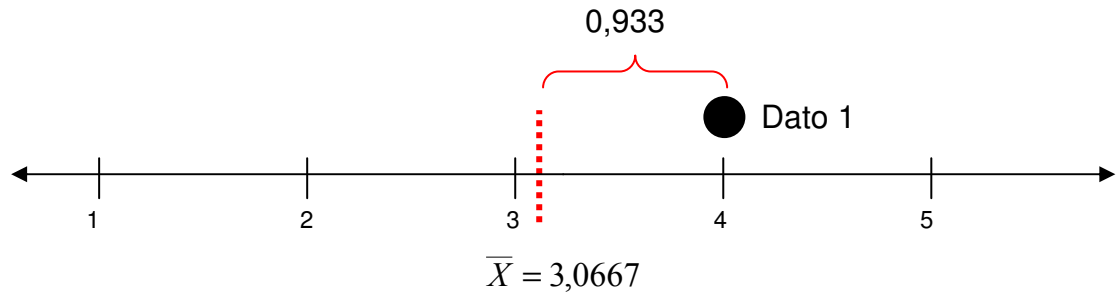
El cálculo de cada uno de ellos se toma basado en la media aritmética.

5.1 DESVIACIÓN MEDIA

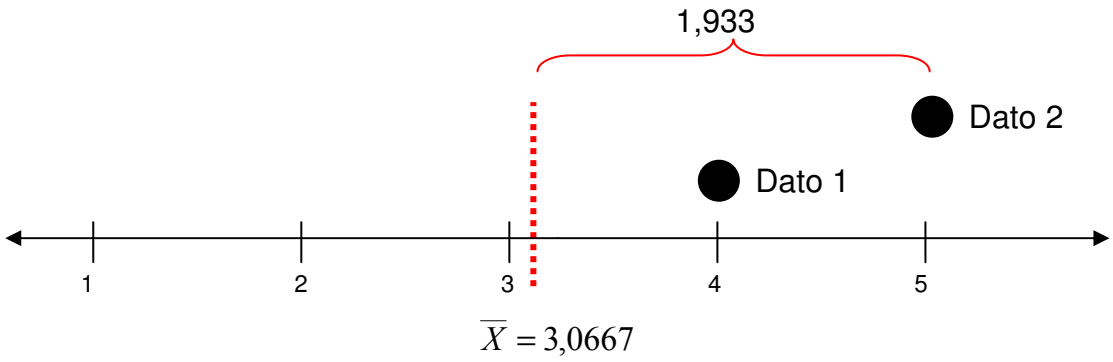
Para conocer con un solo indicador que tan disperso se encuentran un conjunto de datos a un punto de concentración, debemos como primera medida, calcular la distancia de cada dato respecto a una medida de tendencia central. Por ejemplo:

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

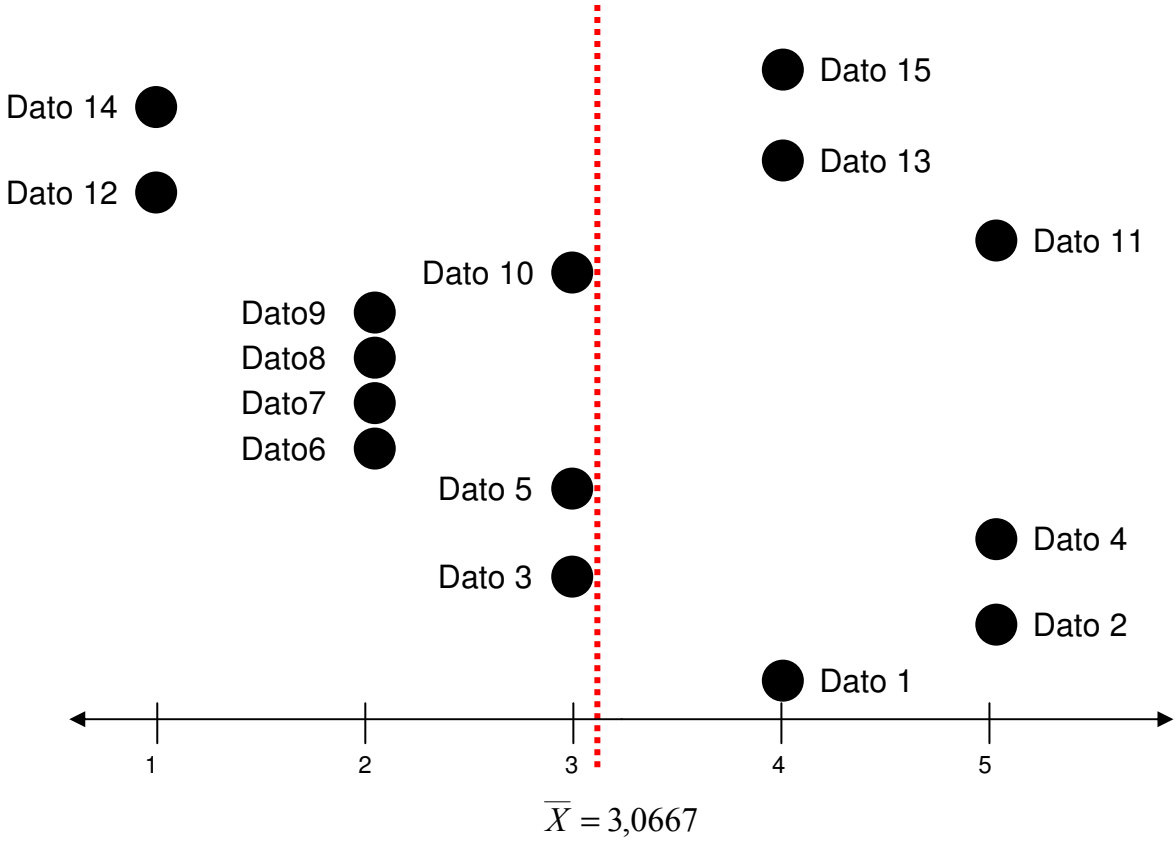
Tenemos que la media aritmética es de aproximadamente 3,0667 (indicador de tendencia central por excelencia). El primer dato (4), se aleja de la media en 0,9333 hacia la derecha. Gráficamente tendríamos:



Para el segundo dato (5) la distancia es de 1,9333 respecto a la media aritmética:



Note que el tercer dato (3) posee una distancia de 0,0667 hacia la izquierda de la media. Para indicar las distancias de estos puntos, agregaremos el signo negativo, por tanto, la distancia del tercer dato sería –0,0667. La representación gráfica de todos los puntos quedaría:



El total de las distancias de los puntos que están a la izquierda respecto a la media es de -8,6 (empleando todos los decimales), que es igual a la sumatoria de las distancias de los puntos que están a la derecha respecto a la media 8,6. Concluimos que la sumatoria de todas las distancias de cada punto respecto a la media aritmética es igual a cero (las distancias se anulan):

$$\sum_{i=1}^{n}\left(X_i - \overline{X}\right) = 0$$

Para responder a la pregunta de ¿qué tan disperso están los datos respecto a la media aritmética?, recurriremos nuevamente al promedio simple. Para llegar a una fórmula básica de dispersión, en que las distancias positivas y negativas no se eliminen, modificaremos la fórmula anterior para trabajar solo con distancias positivas mediante el valor absoluto:

$$\sum_{i=1}^{n}\left|X_i - \overline{X}\right| = 17{,}2$$

La distancia promedio sería de aproximadamente 1,15 (resultado de la división entre la distancia total absoluta y el total de datos). A esta distancia promedio se le conoce con el nombre de desviación media y significa que en promedio, los datos se separan de la media en 1,15.

Desviación media (Dm): Equivale a la división de la sumatoria del valor absoluto de las distancias existentes entre cada dato y su media aritmética y el número total de datos.

$$Dm = \frac{\sum_{i=1}^{n}\left|X_i - \overline{X}\right|}{n}$$

Se debe hacer la distinción que para datos poblacionales (no agrupados), la fórmula quedaría:

$$Dm = \frac{\sum_{i=1}^{N}\left|X_i - \mu\right|}{N}$$

La variación para los datos agrupados en tablas tipo B radica en cambiar el valor de X_i por la marca de clase correspondiente, multiplicando esa distancia por su frecuencia:

$$Dm = \frac{\sum_{i=1}^{Nc} |Mc_i - \mu| . f_i}{N}$$

Población

$$Dm = \frac{\sum_{i=1}^{Nc} |Mc_i - \overline{X}| . f_i}{n}$$

Muestra

Para las tablas tipo A solo cambiaremos la marca de clase por su respectivo valor de clase (representada por X_i):

$$Dm = \frac{\sum_{i=1}^{Nc} |X_i - \mu| . f_i}{N}$$

Población

$$Dm = \frac{\sum_{i=1}^{Nc} |X_i - \overline{X}| . f_i}{n}$$

Muestra

5.1.1 Ejemplo: Desviación media para datos no agrupados

Tres alumnos son sometidos a una competencia para probar sus conocimientos en 10 materias diferentes, cada una sustentada con 10 preguntas. La idea del concurso es encontrar al alumno más idóneo para representar al colegio en un torneo a nivel nacional.

El número de preguntas buenas por materia se muestra a continuación:

Materia	Carlos	Pedro	Juan
1	2	7	5
2	9	2	6
3	10	2	5
4	2	6	5
5	3	6	5
6	1	3	5
7	9	6	4
8	9	7	5
9	1	6	6
10	4	5	4

SOLUCIÓN

Lo primero que analizaremos es la media de los puntajes para cada uno de los alumnos, con el fin de determinar el alumno con mayor promedio de preguntas buenas.

$$Carlos: \overline{X}_c = \frac{50}{10} = 5$$

$$Pedro: \overline{X}_p = \frac{50}{10} = 5$$

$$Juan: \overline{X}_j = \frac{50}{10} = 5$$

Las medias para los resultados de los alumnos coinciden: los tres alumnos tienen responden en promedio 5 preguntas correctas por prueba. ¿Cuál sería entonces el indicador diferenciador entre los alumnos?.

Complementemos el análisis anterior calculando la desviación media:

$$Dm_c = \frac{|2-5|+|9-5|+|10-5|+|2-5|+|3-5|+|1-5|+|9-5|+|9-5|+|1-5|+|4-5|}{10} = \frac{39}{10} = 3{,}9$$

$$Dm_p = \frac{|7-5|+|2-5|+|2-5|+|6-5|+|6-5|+|3-5|+|6-5|+|7-5|+|6-5|+|5-5|}{10} = \frac{21}{10} = 2{,}1$$

$$Dm_j = \frac{|5-5|+|6-5|+|5-5|+|5-5|+|5-5|+|5-5|+|4-5|+|5-5|+|6-5|+|4-5|}{10} = \frac{9}{10} = 0{,}9$$

Carlos muestra una desviación media de 3,9 indicando que los datos se alejan en promedio de la media en 3,9 preguntas buenas. Pedro disminuye su variación (2,9), siendo Juan el que menos variación presenta con 0,9 preguntas tanto por arriba como por debajo de la media aritmética. Se recomienda al colegio elegir como ganador en este caso a Juan, presenta resultados más constantes que los otros dos alumnos, Juan en promedio acierta 5 preguntas buenas con una variación muy baja (rondando entre 4 y 6).

5.1.2 Ejemplo: Desviación media para datos agrupados

Una maquina dispensadora de gaseosas esta programada para llenar un envase con 350 c.c. de un refresco popular. A partir de una muestra de prueba realizada sobre 30 envases se realizó la siguiente tabla de frecuencia:

Ni	Lm	Ls	F	Mc
1	130.0	140.1	2	135.1
2	140.1	150.1	5	145.1
3	150.1	160.1	14	155.1
4	160.1	170.1	4	165.1
5	170.1	180.1	4	175.1
6	180.1	190.0	1	185.1
Total			**30**	

Calcular e interpretar la desviación media.

SOLUCIÓN

PASO 1: Calcular la media aritmética.

$$\overline{X} = \frac{4712,84}{30} = 157,095$$

PASO 2: Calcular la desviación media.

$$Dm = \frac{|135,1-150,095|.2+|145,1-150,095|.5+|155,1-150,095|.14+|165,1-150,095|.4+|175,1-150,095|.4+|185,1-150,095|.1}{30}$$

La desviación media es de aproximadamente 8,8 c.c. Concluimos que con datos suministrados de una muestra, el dispensador llenó los 30 envases con un promedio de 157,095 c.c. con una desviación media de 8,8 c.c.

La desviación media describe un rango de dispersión promedio de llenado del dispensador, ubicándolo entre 148,295 c.c. (equivale a restar la media a la desviación media) y 165,895 c.c. (sumar una desviación media a la media aritmética).

5.1.3 Cálculos de la desviación media en Excel

Presentaremos el cálculo de la desviación media en Excel tanto para datos sin agrupar, como para los datos agrupados en tablas de frecuencias. Copiemos los siguientes datos a partir de la celda B2.

	A	B	C	D	E	F	G
1							
2		14	17	8	9	10	
3		11	12	19	15	9	
4		15	8	21	13	11	
5		16	10	14	13	10	
6		14	11	11	15	7	
7		14	17	8	16	15	
8							
9							

Excel cuenta con la función **DESVPROM** para el cálculo de la desviación media para datos sin agrupar.

DESVPROM: Calcula la desviación media de un conjunto de datos numéricos.

Formato: DESVPROM(número1;número2;…)
Categoría: Estadísticas

Activemos esta función en la celda B9, señalando el rango de celdas B2:F7 en el campo número1.

Al pulsar en el botón Aceptar, se mostrará la desviación media.

	A	B	C	D	E	F	G
1							
2		14	17	8	9	10	
3		11	12	19	15	9	
4		15	8	21	13	11	
5		16	10	14	13	10	
6		14	11	11	15	7	
7		14	17	8	16	15	
8							
9		2.91555556					
10							
11							

=DESVPROM (B2:F7)

Para el cálculo de la desviación media en tablas de frecuencia debemos calcular de antemano la media aritmética y el valor absoluto de las distancias.

Copiemos la siguiente tabla de frecuencia en una hoja nueva en Excel (es la misma utilizada en el ejemplo 5.1.2).

	A	B	C	D	E	F	G
1							
2		**Ni**	**Lm**	**Ls**	**f**	**Mc**	
3		1	130	140.1	2	135.1	
4		2	140.1	150.1	5	145.1	
5		3	150.1	160.1	14	155.1	
6		4	160.1	170.1	4	165.1	
7		5	170.1	180.1	4	175.1	
8		6	180.1	190	1	185.1	
9		**Total**			**30**		
10							
11							

El primer paso es calcular la media aritmética para datos agrupados con ayuda de la función **SUMAPRODUCTO** (ver el ejemplo dado en el punto 4.1.7), aplicado sobre las frecuencias y marcas de clases.

	A	B	C	D	E	F	G
1							
2		**Ni**	**Lm**	**Ls**	**f**	**Mc**	
3		1	130	140.1	2	135.1	
4		2	140.1	150.1	5	145.1	
5		3	150.1	160.1	14	155.1	
6		4	160.1	170.1	4	165.1	
7		5	170.1	180.1	4	175.1	
8		6	180.1	190	1	185.1	
9		**Total**			**30**		
10							
11		157.1	←	**=SUMAPRODUCTO(E3:E8,F3:F8)/E9**			
12							

Luego hallaremos las distancias de cada marca de clase respecto a la media, convirtiéndolas a su valor absoluto con la función **ABS**.

ABS: Devuelve el valor absoluto de un número.

Formato: ABS (número)
Categoría: Matemáticas y trigonométricas

Esta función posee un único campo (número) el cual contendrá, la distancia entre la marca de clase y la media. Para el primer intervalo de clase tendríamos:

```
=ABS(F3-B11)
```

Donde F3 representa la primera marca de clase y B11 la media aritmética. Para completar el cálculo, multiplicaremos esta función por la frecuencia respectiva:

=ABS(F3-B11)*E3

Para poder arrastrar la fórmula, debemos recordar que la celda B11 no varía (la media aritmética es una sola), ubicándonos sobre las letras B11 en modo de edición y luego pulsando la tecla F4.

=ABS(F3-B11)*E3

El resultado final, después de haber arrastrado la fórmula, debería verse como sigue:

	A	B	C	D	E	F	G	H
1								
2		**Ni**	**Lm**	**Ls**	**f**	**Mc**	**Distancia**	
3		1	130	140.1	2	135.1	44	
4		2	140.1	150.1	5	145.1	60	
5		3	150.1	160.1	14	155.1	28	
6		4	160.1	170.1	4	165.1	32	
7		5	170.1	180.1	4	175.1	72	
8		6	180.1	190	1	185.1	28	
9		**Total**			**30**		**264**	
10								
11		157.1						

El total de las distancias se muestra en la celda G9. La desviación (que ubicaremos en la celda B12), es el resulta de la división de la distancia total sobre el número de datos empleados en el ejercicio.

	A	B	C	D	E	F	G	H
1								
2		**Ni**	**Lm**	**Ls**	**f**	**Mc**	**Distancia**	
3		1	130	140.1	2	135.1	44	
4		2	140.1	150.1	5	145.1	60	
5		3	150.1	160.1	14	155.1	28	
6		4	160.1	170.1	4	165.1	32	
7		5	170.1	180.1	4	175.1	72	
8		6	180.1	190	1	185.1	28	
9		**Total**			**30**		**264**	
10								
11		157.1						
12		8.8						
13								

=G9/E9

5.2 LA VARIANZA

Otra forma para asegurar que las diferencias entre la media y los puntos de un valor positivo, es elevándola al cuadrado. Al promedio de estas distancias al cuadrado se le conoce como varianza.

Varianza (S^2 o σ^2): Es el resultado de la división de la sumatoria de las distancias existentes entre cada dato y su media aritmética elevadas al cuadrado, y el número total de datos.

$$S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1} \quad \text{o} \quad \sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$$

Distinguimos dos símbolos para identificar la varianza: S^2 para datos muestrales, y σ^2 para datos poblacionales. Note que la fórmula para la varianza muestral presenta en su denominador al tamaño de la muestra menos uno, tendencia adoptada por los estadísticos para denotar una varianza más conservadora.

Al igual que ocurre con la desviación media, podemos definir las fórmulas para datos agrupados en tablas tipo A y tipo B. Para las tablas tipo A tenemos:

$$\sigma^2 = \frac{\sum_{i=1}^{Nc}(X_i - \mu)^2 . f_i}{N}$$

Población

$$S^2 = \frac{\sum_{i=1}^{Nc}\left(X_i - \overline{X}\right)^2 . f_i}{n}$$

Muestra

Para las tablas tipo B, la clase cambia por la marca de clase del intervalo:

$$\sigma^2 = \frac{\sum_{i=1}^{Nc}(Mc_i - \mu)^2 . f_i}{N}$$

Población

$$S^2 = \frac{\sum_{i=1}^{Nc}\left(Mc_i - \overline{X}\right)^2 . f_i}{n}$$

Muestra

Una advertencia en el uso de esta medida, es que al elevar las distancias al cuadrado, automáticamente se elevan las unidades. Por ejemplo, si unidad trabajada en los datos es centímetros, la varianza da como resultados centímetros al cuadrado.

5.2.1 Ejemplo: Varianza para datos no agrupados

La siguiente muestra representa las edades de 25 personan sometidas a un análisis de preferencias para un estudio de mercado.

25	19	21	35	44
20	27	32	38	33
18	30	19	29	33
26	24	28	39	31
31	18	17	30	27

Determinar la varianza.

SOLUCIÓN

PASO 1: Calcular la media aritmética.

$$\overline{X} = \frac{694}{25} = 27{,}76$$

PASO 2: Calcular la varianza

En este punto, la varianza es identificada por S^2.

$$S^2 = \frac{(25-27{,}76)^2 + (19-27{,}76)^2 + (21+27{,}76)^2 + ... + (27-27{,}76)^2}{25-1}$$

$$S^2 = \frac{1244{,}56}{24} = 51{,}8567$$

La varianza equivale a 51,8567. Por elevar las unidades al cuadrado, carece de un significado contextual dentro del análisis descriptivo del caso.

5.2.2 Ejemplo: Varianza para datos agrupados

Calcular la varianza a partir de la siguiente tabla de frecuencia (suponga que los datos son poblacionales).

Ni	Lm	Ls	f	Mc
1	[15	17)	2	16
2	[17	19)	5	18
3	[19	21)	13	20
4	[21	23)	4	22
5	[23	25]	1	24
Total			**25**	

SOLUCIÓN

PASO 1: Calcular la media aritmética.

$$\mu = \frac{16x2 + 18x5 + 20x13 + 22x4 + 24x1}{25} = \frac{494}{25}$$

$$\mu = 19{,}76$$

PASO 2: Calcular la varianza

En este punto, la varianza es identificada por S^2.

$$\sigma^2 = \frac{(16-19{,}76)^2.2 + (18-19{,}76)^2.5 + (20+19{,}76)^2.13 + (22-19{,}76)^2.4 + (24-19{,}76)^2.1}{25}$$

$$\sigma^2 = \frac{82{,}56}{25} = 3{,}3024$$

5.2.3 Cálculo de la varianza en Excel

Excel posee dos funciones propias para el cálculo de la media, diferenciando los datos muestrales de los datos poblacionales.

VAR: Calcula la varianza de una muestra.

Formato: VAR(número1;número2;…)
Categoría: Estadísticas

VARP: Calcula la varianza de todos los datos de una población.

Formato: VARP(número1;número2;…)
Categoría: Estadísticas

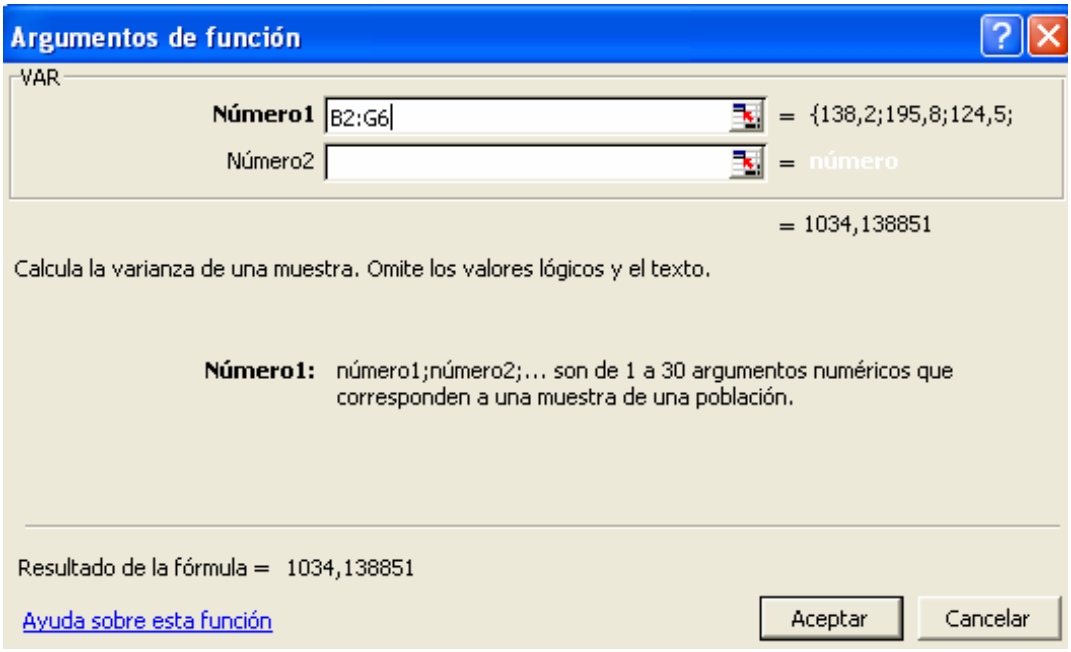
Mostremos su funcionamiento calculando la varianza en ambos casos a partir de los siguientes datos:

138,2	195,8	124,5	101,7	137,1	130,3
110,0	101,4	104,5	128,5	135,5	197,5
159,6	140,7	103,2	134,3	191	180,6
189,9	186,3	116,4	155,3	146,6	199,1
188,4	113,8	121,9	135,7	142,6	125,6

Los datos copiados en Excel desde la celda B2 deberían verse como sigue:

	A	B	C	D	E	F	G	H
1								
2		138,2	195,8	124,5	101,7	137,1	130,3	
3		110,0	101,4	104,5	128,5	135,5	197,5	
4		159,6	140,7	103,2	134,3	191	180,6	
5		189,9	186,3	116,4	155,3	146,6	199,1	
6		188,4	113,8	121,9	135,7	142,6	125,6	
7								
8								

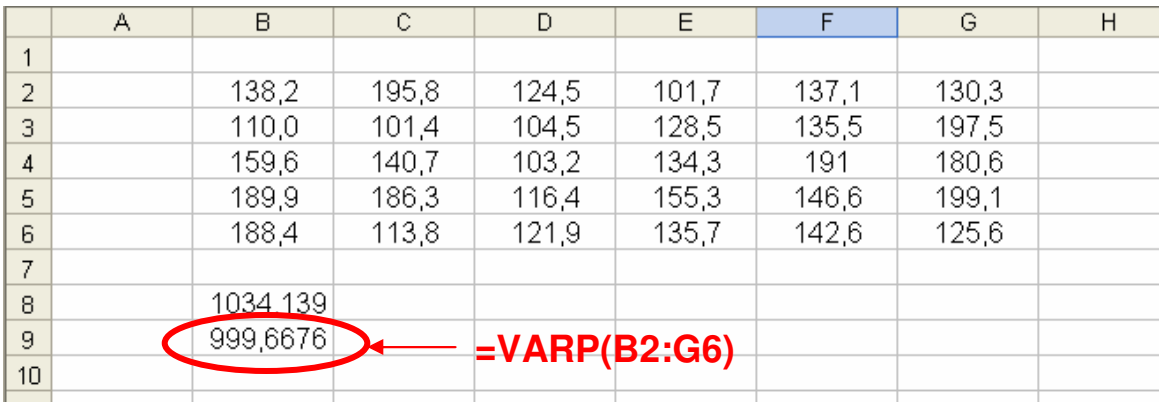
Si los datos provienen de una muestra, emplearemos la función **VAR**, en cuyo denominador se tendría el valor 29 en vez de 30, equivalente al tamaño de la muestra. Activemos esta función en la celda B8.



El resultado de la varianza muestral es de 1034,138051.

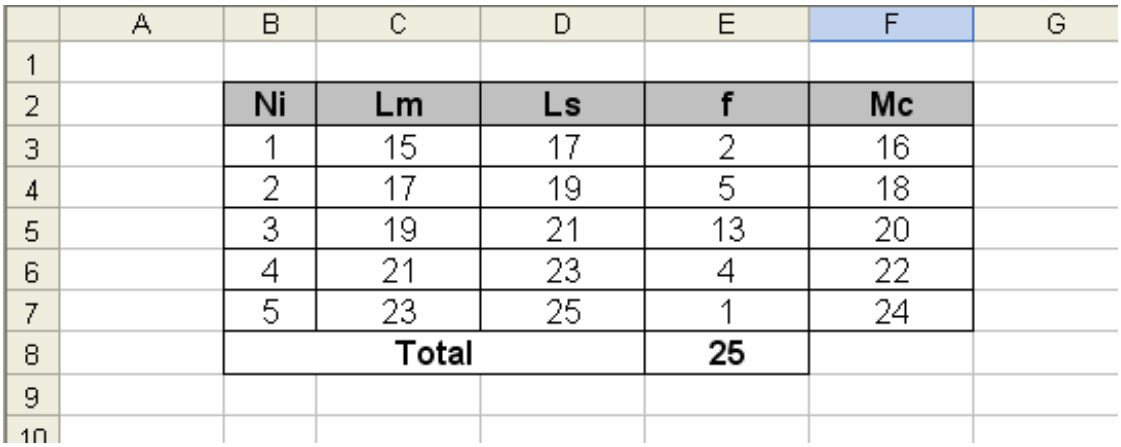
	A	B	C	D	E	F	G	H
1								
2		138,2	195,8	124,5	101,7	137,1	130,3	
3		110,0	101,4	104,5	128,5	135,5	197,5	
4		159,6	140,7	103,2	134,3	191	180,6	
5		189,9	186,3	116,4	155,3	146,6	199,1	
6		188,4	113,8	121,9	135,7	142,6	125,6	
7								
8		1034,139	←	=VAR(B2:G6)				
9								

En la celda B9 calculemos la varianza para datos poblacionales.



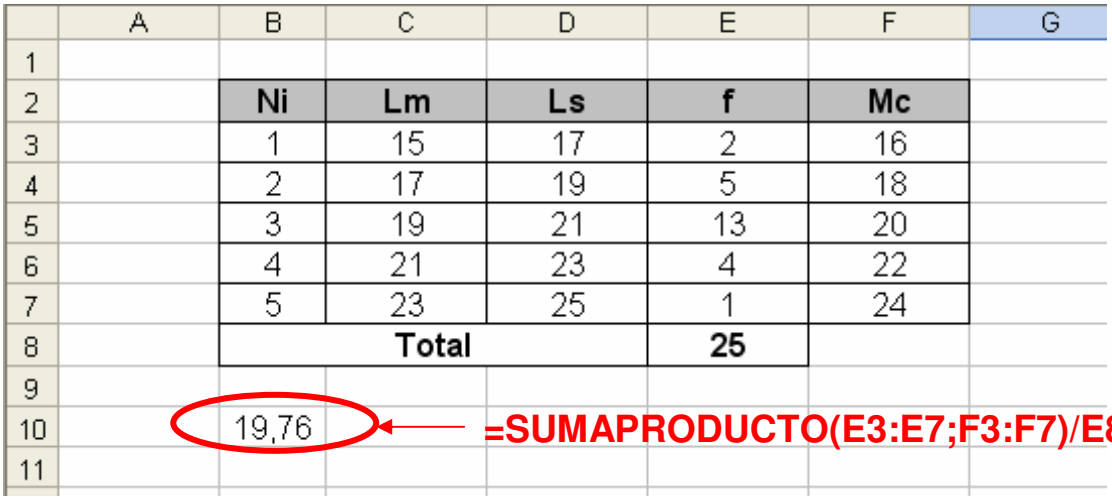
La función de la varianza **VARP**, divide la sumatoria de las distancias al cuadrado por los 30 datos, dando como resultado un valor menor que con la función **VAR** (la varianza para la muestra es un valor más conservador).

Para el cálculo de la varianza en datos agrupados en Excel, tomaremos la tabla de frecuencia dada en el ejemplo 5.2.2.



Ni	Lm	Ls	f	Mc
1	15	17	2	16
2	17	19	5	18
3	19	21	13	20
4	21	23	4	22
5	23	25	1	24
Total			25	

Calculemos la media en la celda B10.



Ni	Lm	Ls	f	Mc
1	15	17	2	16
2	17	19	5	18
3	19	21	13	20
4	21	23	4	22
5	23	25	1	24
Total			25	

19,76 =SUMAPRODUCTO(E3:E7;F3:F7)/E8

En una columna adicional colocaremos las diferencias entre la marca de clase y la media elevadas al cuadrado multiplicadas por su frecuencia.

	A	B	C	D	E	F	G	H
1								
2		**Ni**	**Lm**	**Ls**	**f**	**Mc**	**Distancias**	
3		1	15	17	2	16	28,2752	
4		2	17	19	5	18	15,4880	
5		3	19	21	13	20	0,7488	
6		4	21	23	4	22	20,0704	
7		5	23	25	1	24	17,9776	
8		**Total**			**25**		**82,5600**	
9								
10		19,76						
11								

Analicemos la fórmula empleada desde la celda C3.

```
=(F3-$B$10)^2*E3
```

La celda B10 esta fija indicando la media aritmética. Aparece el operador ^, la cual eleva al cuadrado lo que esta dentro del paréntesis. Esta distancia se multiplica por el número de veces que se repite (por su frecuencia). Al final calculamos su sumatoria.

En la celda B11 calculamos la varianza.

	A	B	C	D	E	F	G	H
1								
2		**Ni**	**Lm**	**Ls**	**f**	**Mc**	**Distancias**	
3		1	15	17	2	16	28,2752	
4		2	17	19	5	18	15,4880	
5		3	19	21	13	20	0,7488	
6		4	21	23	4	22	20,0704	
7		5	23	25	1	24	17,9776	
8		**Total**			**25**		**82,5600**	
9								
10		19,76						
11		3,3024						

=G8/E8

5.3 DESVIACIÓN ESTÁNDAR

Habíamos visto que la varianza transforma todas las distancias a valores positivos elevándolas al cuadrado, con el inconveniente de elevar consigo las unidades de los datos originales.

La desviación estándar soluciona el problema obteniendo la raíz cuadrada de la varianza, consiguiendo así, un valor similar a la desviación media.

Desviación estándar o típica (S o σ): Es igual a la raíz cuadrada de la varianza.

$$S = \sqrt{S^2} \quad \text{o} \quad \sigma = \sqrt{\sigma^2}$$

La S representa la desviación estándar de una muestra, mientras que σ la desviación para todos los datos de una población. Ampliando las fórmulas tenemos:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}}$$

Población

$$S = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n}}$$

Muestra

Aplicamos el mismo procedimiento a las fórmulas para las tablas de frecuencias tipo A.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{Nc} (X_i - \mu)^2 . f_i}{N}}$$

Población

$$S = \sqrt{\frac{\sum_{i=1}^{Nc} (X_i - \overline{X})^2 . f_i}{n}}$$

Muestra

Y para las tablas de frecuencias tipo B.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{Nc} (Mc_i - \mu)^2 . f_i}{N}}$$

Población

$$S = \sqrt{\frac{\sum_{i=1}^{Nc} (Mc_i - \overline{X})^2 . f_i}{n}}$$

Muestra

5.3.1 Ejemplo: Desviación estándar para datos no agrupados

Calcular la desviación estándar al siguiente conjunto de datos muestrales.

220	215	218	210	210
219	208	207	213	225
213	204	225	211	221
218	200	205	220	215
217	209	207	211	218

PASO 1: Calcular la media aritmética.

$$\overline{X} = \frac{220 + 215 + 218 + 210 + 210 + ... + 218}{25} = \frac{5339}{25}$$

$$\overline{X} = 213{,}56$$

PASO 2: Calcular la varianza

En este punto, la varianza es identificada por S^2.

$$S^2 = \frac{(220 - 213{,}56)^2 + (215 - 213{,}56)^2 + (218 + 213{,}56)^2 + (210 - 213{,}56)^2 + ... + (218 - 213{,}56)^2}{25 - 1}$$

$$S^2 = \frac{1030{,}16}{25 - 1} = 42{,}9233$$

PASO 3: Calcular la desviación estándar a partir de la raíz cuadrada de la varianza.

$$S = \sqrt{42{,}9233}$$

$$S = 6{,}5516$$

Los datos se alejan en promedio de la media aritmética en 6,5516 puntos.

5.3.2 Ejemplo: Desviación estándar para datos agrupados

Calcular la desviación estándar a partir de la siguiente tabla de frecuencia. Considere los datos como poblacionales.

No.	Lm	Ls	f	Mc
1	13,20	15,21	15	14,21
2	15,21	17,21	10	16,21
3	17,21	19,21	1	18,21
4	19,21	21,21	4	20,21
5	21,21	23,21	5	22,21
6	23,21	25,21	12	24,21
7	25,21	27,20	1	26,21
Total			48	

PASO 1: Calcular la media aritmética.

$$\mu = \frac{14{,}21x15 + 146{,}21x10 + 18{,}21x1 + 20{,}21x4 + 22{,}21x5 + 24{,}21x12 + 26{,}21x1}{48} = \frac{902}{48}$$

$$\mu = 18{,}7917$$

PASO 2: Calcular la varianza

En este punto, la varianza es identificada por σ^2.

$$\sigma^2 = \frac{(14{,}21 - 18{,}7917)^2.15 + (16{,}21 - 18{,}7917)^2.10 + (18{,}21 - 18{,}7917)^2.1 + ... + (26{,}21 - 18{,}7917)^2.1}{48}$$

$$\sigma^2 = \frac{2789{,}96}{48} = 58{,}1242$$

PASO 3: Calcular la desviación estándar a partir de la raíz cuadrada de la varianza.

$$\sigma = \sqrt{58{,}1242}$$

$$\sigma = 7{,}6239$$

Los datos se alejan en promedio de la media aritmética en 7,6239 puntos.

5.3.3 Cálculo de la Desviación estándar en Excel

Al igual que en la varianza, Excel posee dos funciones para el cálculo de la media, diferenciando los datos muestrales de los datos poblacionales.

DESVEST: Calcula la desviación estándar de una muestra.

Formato: DESVEST(número1;número2;…)
Categoría: Estadísticas

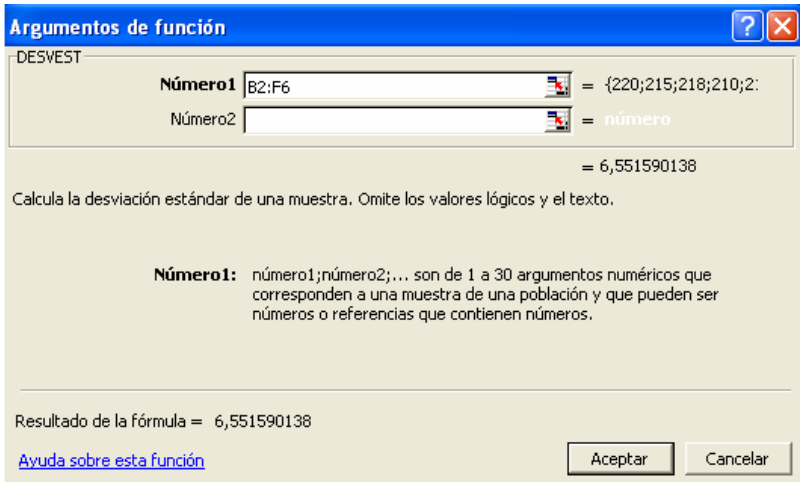
DESVESTP: Calcula la desviación estándar de todos los datos de una población.

Formato: DESVESTP(número1;número2;…)
Categoría: Estadísticas

Tomemos los datos del **ejemplo 5.2.1** para aplicar la fórmula de desviación estándar para datos muestrales. Copie los datos a una hoja en blanco en Excel:

	A	B	C	D	E	F	G
1							
2		220	215	218	210	210	
3		219	208	207	213	225	
4		213	204	225	211	221	
5		218	200	205	220	215	
6		217	209	207	211	218	
7							
8							

En la celda B8 active la función DESVEST, marcando en la primera casilla, losdatos del ejercicio y luego pulsando en el botón aceptar.



El resultado es de aproximadamente 6,5516.

	A	B	C	D	E	F	G
1							
2		220	215	218	210	210	
3		219	208	207	213	225	
4		213	204	225	211	221	
5		218	200	205	220	215	
6		217	209	207	211	218	
7							
8		6,5516					

Para datos agrupados, calcularemos la varianza tal cual como se mostró en la **sección 5.2.3** para luego calcular su raíz cuadrada con la función RAIZ:

RAIZ: Calcula la raíz cuadrada de un número.

Formato: RAIZ(número1)
Categoría: Matemáticas y trigonométricas

Calculemos la raíz cuadrada de una tabla de frecuencia sencilla.

Ni	Clase	f
1	4	15
2	5	10
3	6	1
4	7	4
5	8	5
6	9	12
Total		47

En la celda B11 hallamos la media aritmética de la tabla.

	A	B	C	D	E
1					
2		**Ni**	**Clase**	**f**	
3		1	4	15	
4		2	5	10	
5		3	6	1	
6		4	7	4	
7		5	8	5	
8		6	9	12	
9		**Total**		47	
10					
11		6,2128			
12					

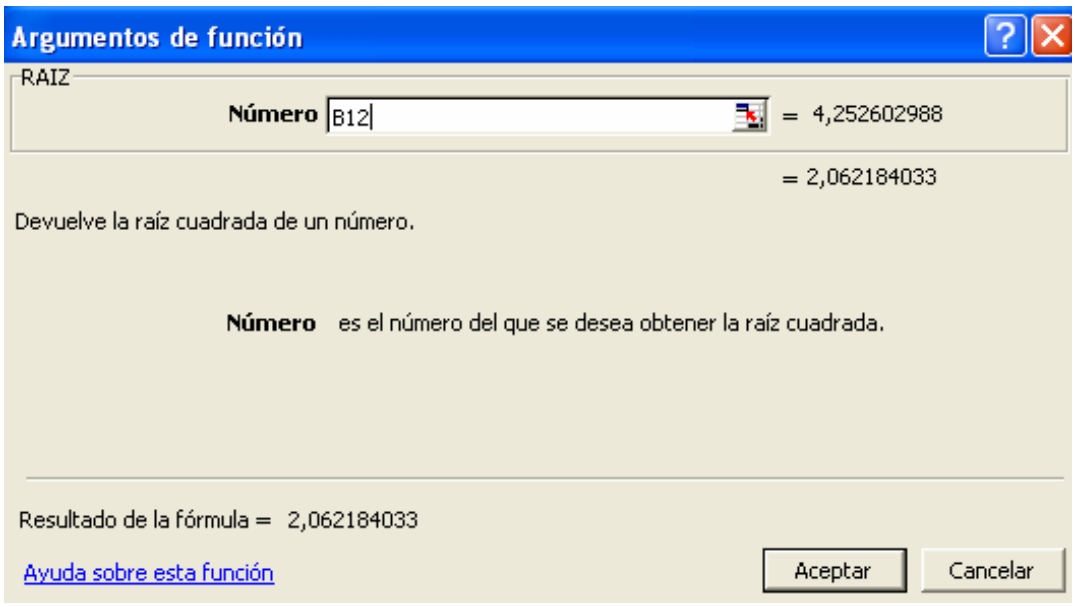
En una columna nueva colocamos las distancias de las clases respecto a la media, multiplicadas por sus frecuencias respectivas.

	A	B	C	D	E	F
1						
2		**Ni**	**Clase**	**f**	**Distancia**	
3		1	4	15	73,44	
4		2	5	10	14,71	
5		3	6	1	0,05	
6		4	7	4	2,48	
7		5	8	5	15,97	
8		6	9	12	93,22	
9		**Total**		47	199,87	
10						
11		6,2128				
12						
13						

Dividimos el total de las distancias al cuadrado por el número de datos (colocamos el resultado en la celda B12).

	A	B	C	D	E	F
1						
2		**Ni**	**Clase**	**f**	**Distancia**	
3		1	4	15	73,44	
4		2	5	10	14,71	
5		3	6	1	0,05	
6		4	7	4	2,48	
7		5	8	5	15,97	
8		6	9	12	93,22	
9		**Total**		47	199,87	
10						
11		6,2128				
12		4,2526				
13						

La desviación será igual a la raíz cuadrada del valor contenido en la celda B12.



La desviación estándar es de 2,0622.

	A	B	C	D	E	F
1						
2		**Ni**	**Clase**	**f**	**Distancia**	
3		1	4	15	73,44	
4		2	5	10	14,71	
5		3	6	1	0,05	
6		4	7	4	2,48	
7		5	8	5	15,97	
8		6	9	12	93,22	
9		**Total**		47	199,87	
10						
11		6,2128				
12		4,2526				
13		2,0622	←	**=RAIZ(B12)**		

5.4 COEFICIENTE DE VARIACIÓN

El coeficiente de variación permite comparar la dispersión entre dos poblaciones distintas e incluso, comparar la variación producto de dos variables diferentes (que pueden provenir de una misma población).

Estas variables podrían tener unidades diferentes, por ejemplo, podremos determinar si los datos tomados al medir el volumen de llenado de un embase de cierto líquido varían más que los datos tomados al medir la temperatura de el liquido contenido en el embase al salir al consumidor. El volumen los mediremos en centímetros cúbicos y la temperatura en grados centígrados.

El coeficiente de variación elimina la dimensionalidad de las variables y tiene en cuenta la proporción existente entre una medida de tendencia y la desviación típica o estándar.

Coeficiente de variación (Cv): Equivale a la razón entre la media aritmética y la desviación típica o estándar.

.

$$Cv = \frac{S}{\overline{X}} \text{ o } Cv = \frac{\sigma}{\mu}$$

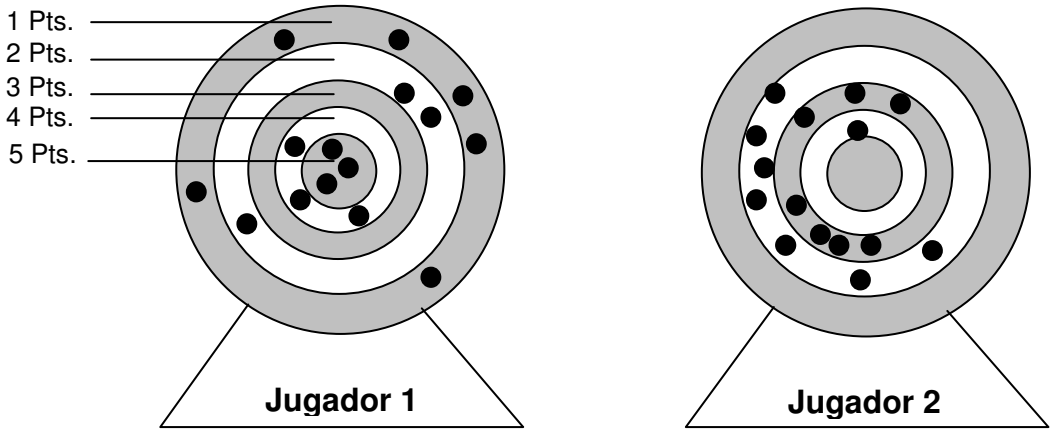
Si envés de la media aritmética se emplea la mediana, obtendremos el coeficiente de variación mediana.

$$Cv = \frac{S}{Me} \text{ o } Cv = \frac{\sigma}{Me}$$

Este índice solo se debe calcular para variables con todo los valores positivos, para dar seguridad de un $\overline{X}$ o μ mayores a cero (un coeficiente de variación positivo).

5.4.1 Ejemplo: Desviación estándar para datos no agrupados

En un juego de tiro al blanco con escopeta de perdigones por dos participantes a un tablero, obtienen el siguiente registro después de 15 disparos cada uno. Determinar el coeficiente de variación para ambos casos.



Disparo	f
1	6
2	3
3	0
4	3
5	3

Disparo	f
1	0
2	7
3	7
4	1
5	0

PASO 1: Calcular las medias aritméticas:

$$\overline{X}_1 = \frac{1x6 + 2x3 + 3x0 + 4x3 + 5x3}{15} = \frac{39}{15} = 2{,}6$$

$$\overline{X}_2 = \frac{1x0 + 2x7 + 3x7 + 4x1 + 5x0}{15} = \frac{39}{15} = 2{,}6$$

PASO 2: Calcular las varianzas

En este punto, la varianza es identificada por S^2.

$$S_1^2 = \frac{(1-2{,}6)^2.6 + (2-2{,}6)^2.3 + (3-2{,}6)^2.0 + (4-2{,}6)^2.3 + (5-2{,}6)^2.3}{15-1}$$

$$S_1^2 = \frac{39{,}6}{14} = 2{,}8286$$

$$S_2^2 = \frac{(1-2{,}6)^2.0 + (2-2{,}6)^2.7 + (3-2{,}6)^2.7 + (4-2{,}6)^2.1 + (5-2{,}6)^2.0}{15-1}$$

$$S_1^2 = \frac{5{,}6}{14} = 0{,}4$$

PASO 3: Calcular la desviación estándar a partir de la raíz cuadrada de la varianza.

$$S_1 = \sqrt{2{,}8286} = 1{,}6818$$

$$S_2 = \sqrt{0{,}4} = 0{,}6325$$

La puntuación de los disparos se aleja en promedio de la media aritmética en aproximadamente 1,6818 para el jugador 1 y 0,6325 para el jugador 2.

PASO 4: Calcular el coeficiente de variación.

$$Cv_1 = \frac{S_1}{\overline{X}_1} = \frac{1{,}6818}{2{,}6} = 0{,}6469$$

$$Cv_2 = \frac{S_2}{\overline{X}_2} = \frac{0{,}6325}{2{,}6} = 0{,}2433$$

El menor coeficiente de variación indica que el jugador 2 presento una dispersión menor de sus puntuaciones respecto a la media, caso contrario al jugador 1 donde la dispersión fue mayor.

5.4.2 Calculo del coeficiente de variación en Excel

Para calcular el coeficiente de variación con ayuda de Excel, debemos calcular primero la media aritmética y la desviación estándar. Por ejemplo, calculemos el coeficiente de variación para los siguientes datos:

	A	B	C	D	E	F	G
1							
2		1	1	1	1	1	
3		1	2	2	2	4	
4		4	4	5	5	5	

Empleando las fórmulas vistas en Excel, se halla la media y desviación (tomando los valores como muestrales):

	A	B	C	D	E	F	G
1							
2		1	1	1	1	1	
3		1	2	2	2	4	
4		4	4	5	5	5	
5							
6		**Media**	2,6	← **=PROMEDIO(B2:F4)**			
7		**Desviación**	1,6818	← **=DESVEST(B2:F4)**			
8							

El coeficiente de variación es el resulta de la división entre la desviación (C7) y la media (C6):

	A	B	C	D	E	F	G
1							
2		1	1	1	1	1	
3		1	2	2	2	4	
4		4	4	5	5	5	
5							
6		**Media**	2,6				
7		**Desviación**	1,6818				
8		**Cv**	0,6469	← **=C7/C6**			

5.5 EJERCICIOS PROPUESTOS

5.5.1 Calcular la desviación media a partir del registro de las siguientes edades de una muestra de 36 personas.

48	19	12	1	19	15
1	5	2	14	18	49
20	30	2	17	46	40
43	45	8	28	23	50
7	44	11	18	21	13
48	3	35	41	49	23

5.5.2 Calcular la desviación media a partir de la siguiente tabla de frecuencia.

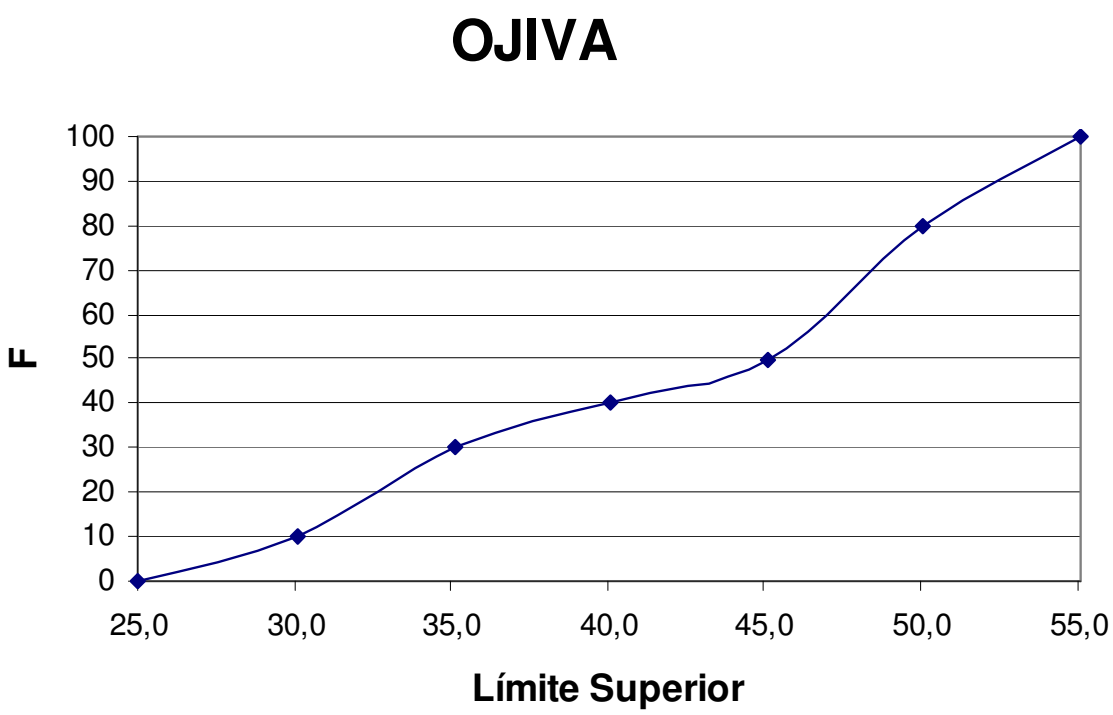
Nc	Lm	Ls	f
1	100,00	150,51	5
2	150,51	201,01	8
3	201,01	251,51	9
4	251,51	302,01	15
5	302,01	352,51	2
6	352,51	403,01	6
7	403,01	453,50	4
Total			**49**

5.5.3 Calcular la desviación media, varianza y desviación estándar a los datos mostrados en los **ejercicios 4.4.1, 4.4.2, 4.4.3 y 4.4.4**

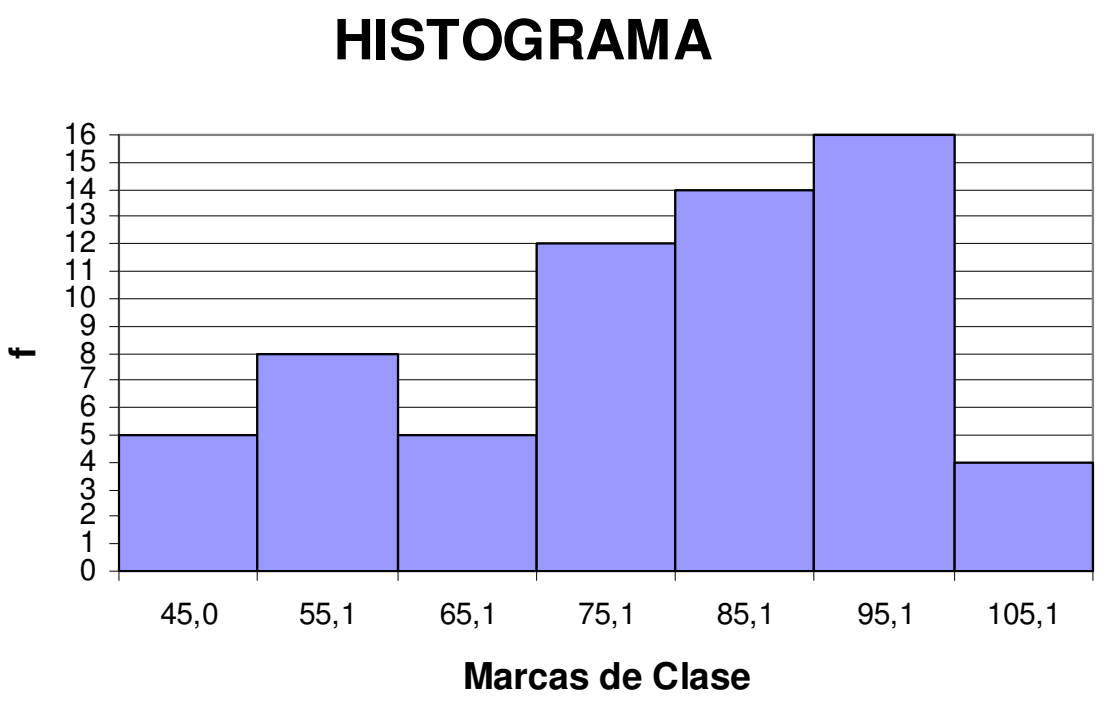
5.5.4 Calcule la desviación media, varianza y desviación estándar a partir de los siguientes datos sin agrupar y agrupándolos en una tabla de frecuencia tipo B (notar la variación de las medidas de dispersión en ambos casos).

49,15	46,17	53,28	49,41	49,00	36,14	41,65	51,75
45,13	43,00	41,95	45,95	52,66	47,50	37,43	48,53
47,24	47,55	51,17	52,69	37,12	49,39	35,20	45,14
35,20	40,59	54,06	47,05	47,04	53,13	53,88	42,33
45,16	35,87	35,02	39,33	48,64	51,83	49,89	36,13

5.5.5 Calcule la desviación media, varianza y desviación estándar a partir del siguiente gráfico de ojiva (dado en el **ejercicio 4.4.6**):



5.5.6 Calcule la desviación media, varianza y desviación estándar a partir del siguiente gráfico de ojiva (dado en el **ejercicio 4.4.7**):



5.6 CASO: EL RÍO MAGDALENA

El cuadro que figura más abajo da los caudales mensuales del rio Magdalena observados durante los meses del abril y mayo, desde 1988 a 2005 (la unidad de medida no se precisa).

AÑO	***ABRIL***	***MAYO***
1988	600	512
1989	227	211
1990	487	469
1991	560	370
1992	521	363
1993	423	272
1994	307	241
1995	390	253
1996	364	408
1997	284	233
1998	415	245
1999	255	199
2000	209	215
2001	230	297
2002	424	309
2003	528	303
2004	258	196
2005	242	166

Se desea ordenar estos datos y efectuar el análisis siguiente:

1. Dar una representación global de los caudales de abril y mayo. Graficar mediante dos histogramas los datos resumidos (***recomendación:*** agrupe los datos empleando tablas de frecuencia con iguales intervalos de clase).
2. Calcular la media de los caudales de abril, y la media de los caudales de mayo.
3. Calcular la desviación típica de los caudales de abril, y la desviación típica de los caudales de mayo.
4. Comparar los caudales de abril con los caudales de mayo, a partir de la información suministrada en la segunda y tercera pregunta.
5. Realizar conclusiones sobre: media, mediana, moda, frecuencias, desviaciones e histogramas de frecuencia.

CAPITULO 6

MEDIDAS DE POSICIÓN

Las medidas de posición equivalen a los valores que puede tomar una variable caracterizados por agrupar a cierto porcentaje de observaciones en la muestra o población.

Las medidas de posición son ideales para obtener información adicional a partir de datos resumidos, es decir, que presentan perdida de información por agrupamiento en intervalos de clase.

CAPITULO 6: MEDIDAS DE POSICIÓN

Medidas de posición: Son indicadores estadísticos que muestran la frecuencia acumulada hasta un valor k cualquiera.

En este capitulo analizaremos tres medidas de posición:

- Percentiles
- Deciles
- Cuartiles

Es necesario revisar nuevamente el concepto de interpolación, ya que la base de estos indicadores es encontrar el valor de la variable a partir de un porcentaje de datos acumulados, de forma similar como se hizo con la mediana.

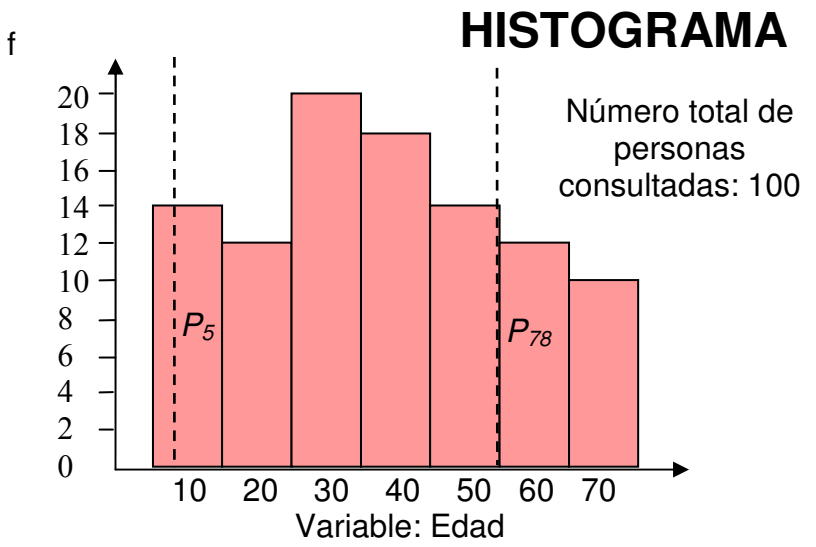
6.1 PERCENTILES

Los percentiles representan los valores de la variable que están por debajo de un porcentaje, el cual puede ser una valor de 1% a 100% (en otras palabras, el total de los datos es divido en 100 partes iguales).

La notación empleada será:

$$P_k$$

Donde *k* es equivalente al porcentaje de datos acumulados, y P_k es el valor de la variable que representa dicho porcentaje. Por ejemplo, P_5 es el valor de la variable que deja por debajo el 5% de los datos. P_{78} será entonces el valor que agrupa el 78% de los datos.



Podemos concluir que P_{50} sería el valor que divide en dos parte iguales la cantidad de datos de la muestra o población siendo equivalente a la mediana.

$$P_{50} = Me$$

Traslademos el gráfico de barra a su respectiva tabla de frecuencia y tratemos de localizar los Percentiles expuestos en el ejemplo:

Nc	Lm	Ls	f	F	h	H
1	[5	15)	14	14	14,00%	14,00%
2	[15	25)	12	26	12,00%	26,00%
3	[25	35)	20	46	20,00%	46,00%
4	[35	45)	18	64	18,00%	64,00%
5	[45	55)	14	78	14,00%	78,00%
6	[55	65)	12	90	12,00%	90,00%
7	[65	75]	10	100	10,00%	100,00%
TOTAL			**100**		**100,00%**	

Podemos concluir fácilmente (con ayuda de las frecuencias acumuladas), que 14 personas (14% del total) están por debajo de los 15 años (podemos aproximarlo a 15 años), lo cual representaría al percentil 14:

$$P_{14} = 15$$

El percentil 5 (P_5) no puede ser calculado directamente, pero podemos concluir que dicho valor se encuentra en el primer intervalo, ya que este acumula el 14% de las personas. No ocurre lo mismo con el percentil 78 (P_{78}) que aparece directamente en la tabla:

Nc	Lm	Ls	f	F	h	H
1	[5	15)	14	14	14,00%	14,00%
2	[15	25)	12	26	12,00%	26,00%
3	[25	35)	20	46	20,00%	46,00%
4	[35	45)	18	64	18,00%	64,00%
5	[45	55)	14	78	14,00%	78,00%
6	[55	65)	12	90	12,00%	90,00%
7	[65	75]	10	100	10,00%	100,00%
TOTAL			**100**		**100,00%**	

En el intervalo 5 se encuentra el percentil 78

$$P_{78} = 55$$

El 78% de las personas consultadas poseen una edad igual o inferior a los 55 años.

6.1.1 Ejemplo: Calculo de percentiles

A partir de la tabla de frecuencia anterior calcular el percentil 5 (P_5)

SOLUCIÓN

PASO 1: Localizar en cuál de los intervalos de clase se encuentra el percentil

Como se había mencionado, el percentil 5 se encuentra en el primer intervalo.

Nc	Lm	Ls	f	F	h	H
1	[5	15)	14	14	14,00%	14,00%
2	[15	25)	12	26	12,00%	26,00%
3	[25	35)	20	46	20,00%	46,00%
4	[35	45)	18	64	18,00%	64,00%
5	[45	55)	14	78	14,00%	78,00%
6	[55	65)	12	90	12,00%	90,00%
7	[65	75]	10	100	10,00%	100,00%
TOTAL			**100**		**100,00%**	

En el intervalo 1 se encuentra el percentil 5

PASO 2: Interpolar los datos para encontrar el percentil. En resumen tenemos que:

	Límite Superior	**H**
	15,00 (Ls_1)	14,00% (H_1)
	5,00 (Ls_0)	0,00% (H_0)
Diferencia	**10,00**	**14,.00%**

En este caso, suponemos un intervalo adicional cuyo límite superior llamaremos Ls_0 equivalente a 5 el cual agrupa 0% de los datos. Entre los dos límites superiores abarcan un total de 14% de los datos. Si queremos llegar al 5% de los datos, debemos incrementar el porcentaje en una cantidad igual.

$$10{,}00 \longrightarrow 14{,}00\%$$
$$\text{Incremento} \longrightarrow 5{,}00\%$$

$$Incremento = \frac{5{,}00\% x 10{,}00}{14{,}00\%} = 3{,}57$$

Para llegar al 5% de los datos, el límite 5 se debe aumentar en 3,57 unidades.

$$P_5 = 5{,}00 + 3{,}57$$

$$P_5 = 8{,}57$$

6.1.2 La fórmula para calcular percentiles

El percentil *k* parte desde límite superior del intervalo anterior al que se encuentra dicho percentil más un incremento

$$P_k = Ls_{i-1} + Incremento$$

El incremento esta dado por:

$$P_k = Ls_{i-1} + A.\frac{(k - H_{i-1})}{(H_i - H_{i-1})}$$

Simplificando aún más la fórmula tenemos:

$$P_k = Ls_{i-1} + A.\frac{(k - H_{i-1})}{h_i}$$

Para expresar la fórmula en frecuencias absolutas tenemos que:

$$P_k = Ls_{i-1} + A.\frac{(n.k - F_{i-1})}{f_i}$$

Aplicando la fórmula al **ejemplo 6.1.1**, concluimos:

$$P_5 = 5 + 10{,}00.\frac{(5{,}00\% - 0{,}00\%)}{14{,}00\%} = 8{,}57$$

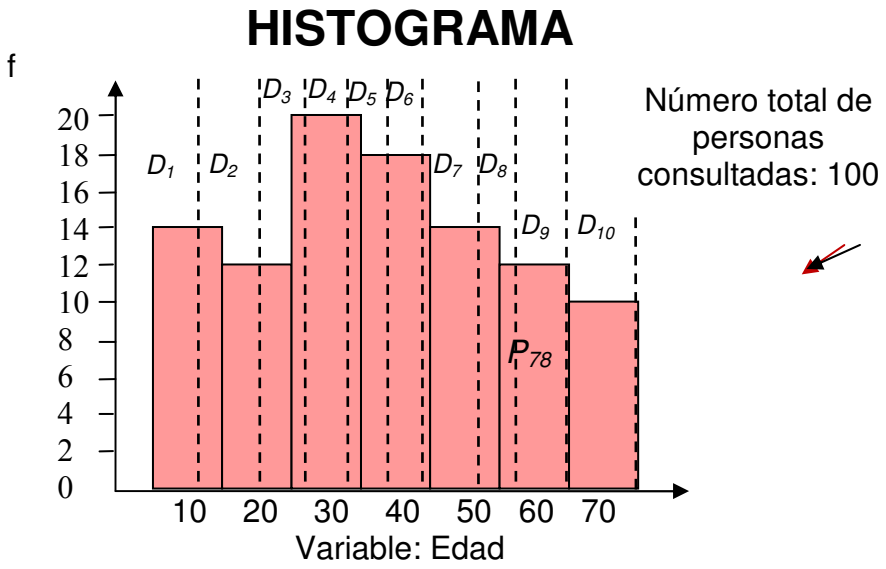
6.2 DECILES

Para los deciles, tomaremos el total de los datos divididos en 10 partes iguales, por tanto, existirán 10 deciles representado como D_k

D_1 = Valor de la variable que agrupa el 10% de los datos.
D_2 = Valor de la variable que agrupa el 20% de los datos.
D_3 = Valor de la variable que agrupa el 30% de los datos.
D_4 = Valor de la variable que agrupa el 40% de los datos.
D_5 = Valor de la variable que agrupa el 50% de los datos.
D_6 = Valor de la variable que agrupa el 60% de los datos.
D_7 = Valor de la variable que agrupa el 70% de los datos.

D_8 = Valor de la variable que agrupa el 80% de los datos.
D_9 = Valor de la variable que agrupa el 90% de los datos.
D_{10} = Valor de la variable que agrupa el 100% de los datos.

Nuestro histograma con los deciles identificados quedaría como sigue:



Las equivalencias entre percentiles y deciles son:

$$P_{10} = D_1$$
$$P_{20} = D_2$$
$$P_{30} = D_3$$
$$P_{40} = D_4$$
$$P_{50} = D_5 = Me$$
$$P_{60} = D_6$$
$$P_{70} = D_7$$
$$P_{80} = D_8$$
$$P_{90} = D_9$$
$$P_{100} = D_{101}$$

6.2.1 Ejemplo: Calculo de deciles

A partir de la tabla de frecuencia dada para ejemplificar los percentiles, encontrar el decil 2.

SOLUCIÓN

PASO 1: Localizar en cuál de los intervalos de clase se encuentra el decil

El Decil 2 se encuentra en el segundo intervalo, ya que este almacena hasta el 26% de los datos.

Nc	Lm	Ls	f	F	h	H
1	[5	15)	14	14	14,00%	14,00%
2	[15	25)	12	26	12,00%	26,00%
3	[25	35)	20	46	20,00%	46,00%
4	[35	45)	18	64	18,00%	64,00%
5	[45	55)	14	78	14,00%	78,00%
6	[55	65)	12	90	12,00%	90,00%
7	[65	75]	10	100	10,00%	100,00%
TOTAL			**100**		**100,00%**	

En el intervalo 2 se encuentra el decil 2

PASO 2: Interpolar los datos para encontrar el decil. En resumen tenemos que:

	Límite Superior	**H**
	25,00 (Ls_2)	26,00% (H_2)
	15,00 (Ls_1)	14,00% (H_1)
Diferencia	**10,00**	**12,00%**

Entre los dos límites superiores abarcan un total de 12% de los datos. Si queremos llegar al 20% de los datos, debemos incrementar el porcentaje acumulado en Ls_1 en un 6%

10,00 ⟶ 12,00%

Incremento ⟶ 6,00%

$$Incremento = \frac{6{,}00\% x 10{,}00}{12{,}00\%} = 5{,}00$$

Para llegar al 20% de los datos acumulados, el límite de 15,00 se debe aumentar en 6,25 unidades.

$$D_2 = 15{,}00 + 5{,}00$$

$$D_2 = 20{,}00$$

6.2.2 La fórmula para calcular deciles

El decil *k* parte desde límite superior del intervalo anterior al que se encuentra dicho decil más un incremento

$$D_k = Ls_{i-1} + Incremento$$

El incremento esta dado por:

$$D_k = Ls_{i-1} + A.\frac{(10\%.k - H_{i-1})}{h_i}$$

Para expresar la fórmula en frecuencias absolutas tenemos que:

$$D_k = Ls_{i-1} + A.\frac{(10\%.n.k - F_{i-1})}{f_i}$$

Aplicando la fórmula al **ejemplo 6.2.1**, concluimos:

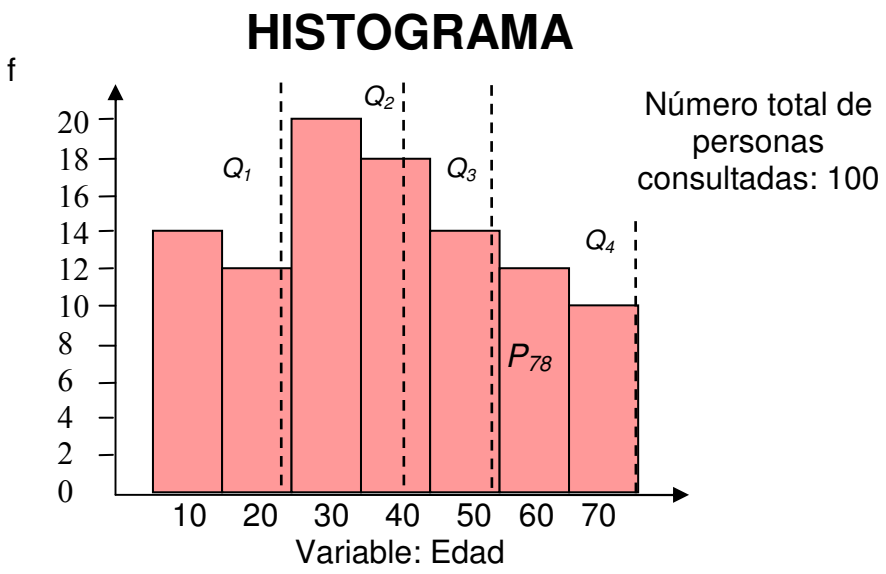
$$D_2 = 15{,}00 + 10{,}00.\frac{(20{,}00\% - 14{,}00\%)}{12{,}00\%} = 20{,}00$$

6.3 CUARTILES

Para los deciles, tomaremos el total de los datos divididos en 4 partes iguales. Denotaremos el cuartel como Q_k.

Q_1 = Valor de la variable que agrupa el 25% de los datos.
Q_2 = Valor de la variable que agrupa el 50% de los datos.
Q_3 = Valor de la variable que agrupa el 75% de los datos.
Q_4 = Valor de la variable que agrupa el 100% de los datos.

El histograma de ejemplo con los cuartiles identificados quedaría como sigue:



Las equivalencias entre percentiles, deciles y cuartiles son:

$$P_{25} = Q_1$$
$$P_{50} = D_5 = Q_2 = Me$$
$$P_{75} = Q_3$$
$$P_{10} = D_{10} = Q_4$$

6.3.1 Ejemplo: Calculo de cuartiles

Calcular el cuartil 3.

SOLUCIÓN

PASO 1: Localizar en cuál de los intervalos de clase se encuentra el cuartil.

El Cuartil 3 se encuentra en el quinto intervalo, ya que este almacena hasta el 78% de datos.

Nc	Lm	Ls	f	F	h	H
1	[5	15)	14	14	14,00%	14,00%
2	[15	25)	12	26	12,00%	26,00%
3	[25	35)	20	46	20,00%	46,00%
4	[35	45)	18	64	18,00%	64,00%
5	[45	55)	14	78	14,00%	78,00%
6	[55	65)	12	90	12,00%	90,00%
7	[65	75]	10	100	10,00%	100,00%
TOTAL			**100**		**100,00%**	

En el intervalo 5 se encuentra el cuartil 3

PASO 2: Interpolar los datos para encontrar el cuartil. En resumen tenemos que:

	Límite Superior	H
	55,00 (Ls_5)	78,00% (H_5)
	45,00 (Ls_4)	64,00% (H_4)
Diferencia	**10,00**	**14,00%**

Entre los dos límites superiores abarcan un total de 14% de los datos. Si queremos llegar al 75% de los datos, debemos incrementar el porcentaje acumulado en Ls_4 en un 11%

$$\begin{array}{rcl} 10{,}00 & \longrightarrow & 14{,}00\% \\ \text{Incremento} & \longrightarrow & 11{,}00\% \end{array}$$

$$Incremento = \frac{11{,}00\% x 10{,}00}{14{,}00\%} = 7{,}86$$

Para llegar al 20% de los datos acumulados, el límite de 45,00 se debe aumentar en 7,86 unidades.

$$Q_3 = 45{,}00 + 7{,}86$$

$$Q_3 = 52{,}86$$

6.3.2 La fórmula para calcular cuartiles

El cuartil *k* parte desde límite superior del intervalo anterior al que se encuentra dicho decil más un incremento

$$Q_k = Ls_{i-1} + Incremento$$

El incremento esta dado por:

$$Q_k = Ls_{i-1} + A.\frac{(25\%.k - H_{i-1})}{h_i}$$

Para expresar la fórmula en frecuencias absolutas tenemos que:

$$Q_k = Ls_{i-1} + A.\frac{(25\%.n.k - F_{i-1})}{f_i}$$

Aplicando la fórmula al **ejemplo 6.3.1**, concluimos:

$$Q_3 = 45{,}00 + 10{,}00.\frac{(75{,}00\% - 64{,}00\%)}{14{,}00\%} = 52{,}86$$

6.4 APLICACIÓN DE PERCENTILES Y CUARTILES EN EXCEL

MS Excel dispone de las funciones ***PERCENTIL*** y ***CUARTIL*** creadas para determinar estos indicadores de posición en datos no agrupados. Copiemos los siguientes datos sin agrupar en una nueva hoja de Excel.

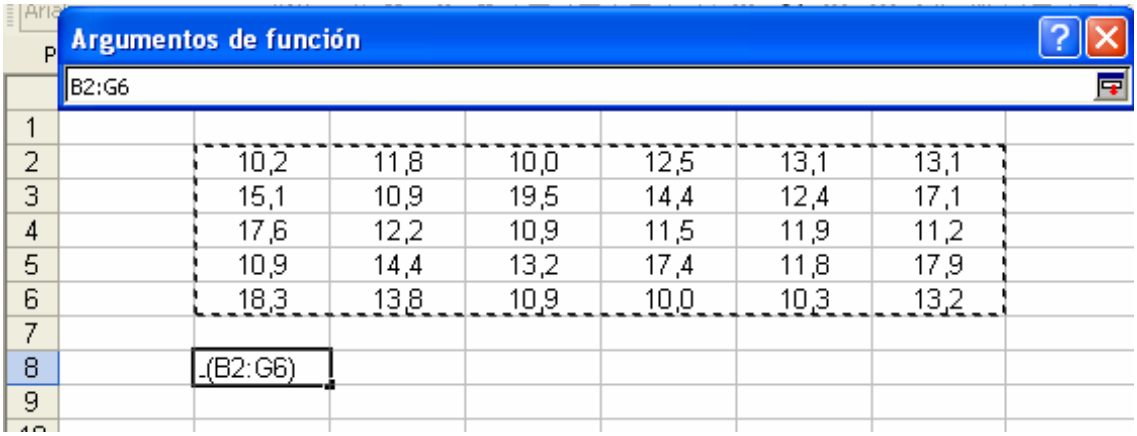
	A	B	C	D	E	F	G	H
1								
2		10,2	11,8	10,0	12,5	13,1	13,1	
3		15,1	10,9	19,5	14,4	12,4	17,1	
4		17,6	12,2	10,9	11,5	11,9	11,2	
5		10,9	14,4	13,2	17,4	11,8	17,9	
6		18,3	13,8	10,9	10,0	10,3	13,2	
7								
8								

Empleemos ahora la función ***PERCENTIL*** desde la celda B8 para calcular el percentil 48.

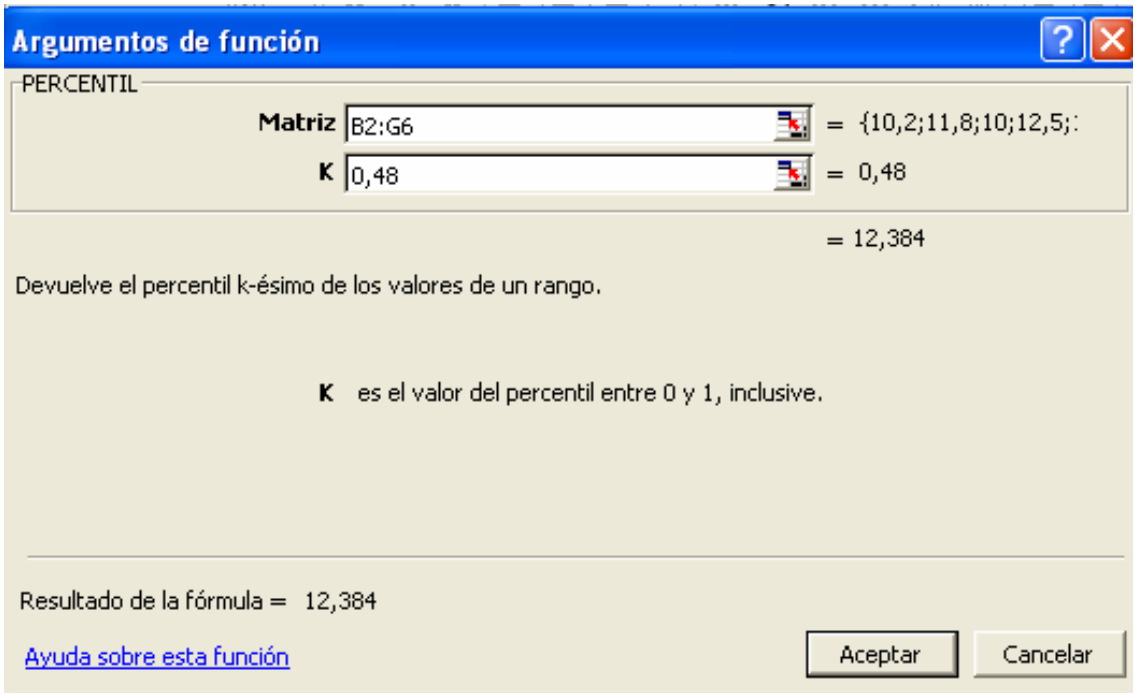
PERCENTIL: Calcula el percentil k de un conjunto de datos. (Siendo k un valor entre 0 y 1)

Formato: PERCENTIL(matriz;k)
Categoría: Estadística

Activemos la fórmula y señalemos en la casilla matriz los datos dados en el ejemplo.



En el valor k colocaremos 0,48 (ya que acepta valores de 0 a 1, siendo 1 el equivalente al 100% de los datos acumulados).



El percentil 48 es equivalente a 12,384.

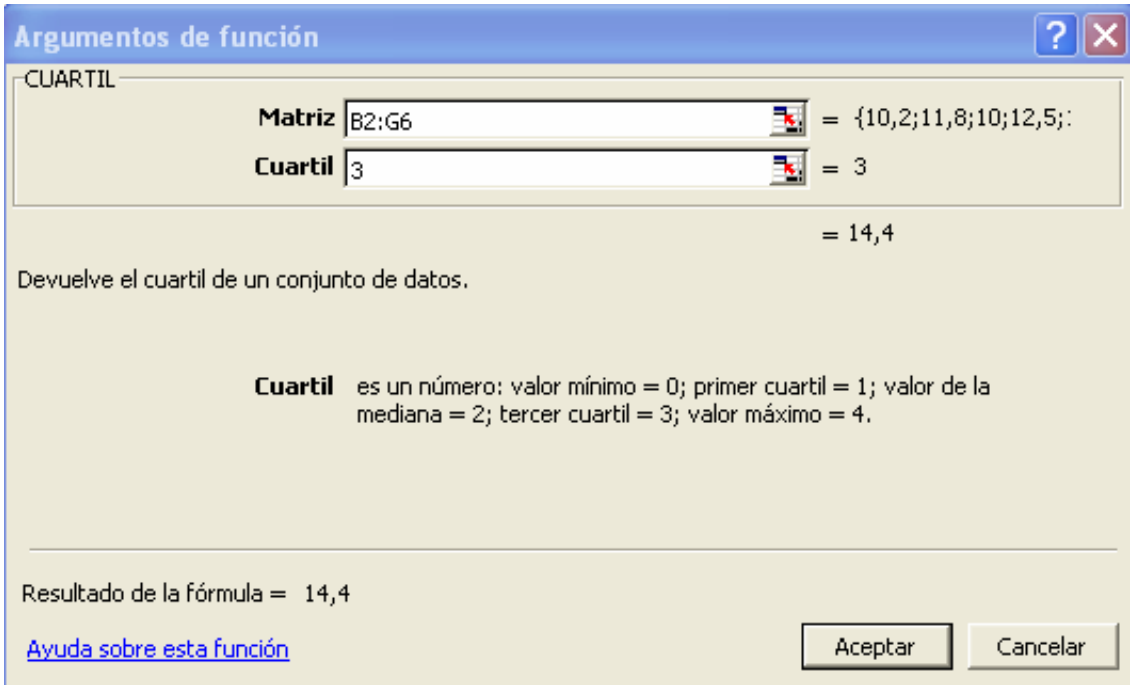
	A	B	C	D	E	F	G	H
1								
2		10,2	11,8	10,0	12,5	13,1	13,1	
3		15,1	10,9	19,5	14,4	12,4	17,1	
4		17,6	12,2	10,9	11,5	11,9	11,2	
5		10,9	14,4	13,2	17,4	11,8	17,9	
6		18,3	13,8	10,9	10,0	10,3	13,2	
7								
8		12,384	←	=PERCENTIL(B2:G6;0,48)				
9								

Para calcular los cuartiles emplearemos la función ***CUARTIL*** con parámetros parecidos a la función ***PERCENTIL***, solo que k representa un valor del 1 al 4 (se incluye el 0 para identificar el valor mínimo de los datos).

CUARTIL: Calcula el cuartil k de un conjunto de datos. (Siendo k un valor entre 0 y 4)

Formato: CUARTIL(matriz;cuartil)
Categoría: Estadística

Hallemos el cuartil 3, el cual el representa el 75% de los datos acumulados. Desde la celda B9 activemos la función señalando en la casilla matriz los datos del ejemplo. En la casilla cuartil escribiremos 3.



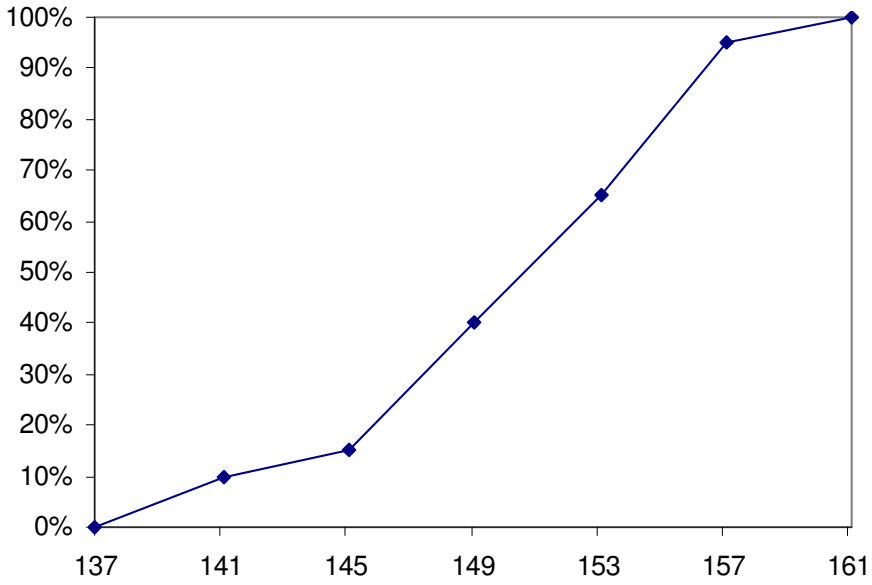
El valor resultante es 14,4.

	A	B	C	D	E	F	G	H
1								
2		10,2	11,8	10,0	12,5	13,1	13,1	
3		15,1	10,9	19,5	14,4	12,4	17,1	
4		17,6	12,2	10,9	11,5	11,9	11,2	
5		10,9	14,4	13,2	17,4	11,8	17,9	
6		18,3	13,8	10,9	10,0	10,3	13,2	
7								
8		12,384						
9		14,4 ←	=CUARTIL(B2:G6;3)					
10								

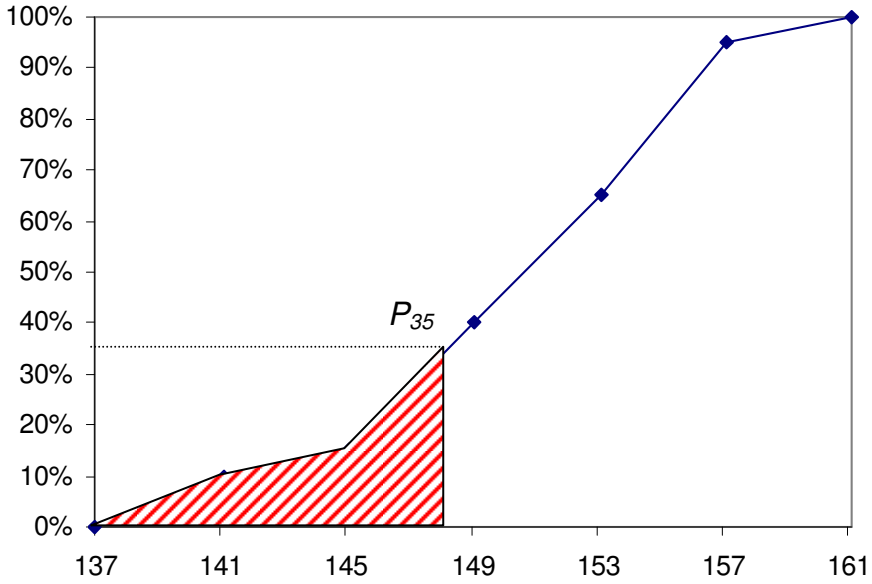
6.5 LAS MEDIDAS DE POSICIÓN Y EL GRÁFICO DE OJIVA

El gráfico de ojiva es ideal para mostrar las medidas de posición, ya que esta registra las frecuencias acumuladas, tanto absolutas como relativas. Analicemos la siguiente tabla de frecuencia con su respectivo gráfico de ojiva:

Nc	Lm	Ls	f	F	h	H
1	137.0	141.1	4	4	10,00%	10,00%
2	141.1	145.1	2	6	5,00%	15,00%
3	145.1	149.1	10	16	25,00%	40,00%
4	149.1	153.1	10	26	25,00%	65,00%
5	153.1	157.1	12	38	30,00%	95,00%
6	157.1	161.1	2	40	5,00%	100,00%
TOTAL			**40**		**100,00%**	



Grafiquemos el percentil 35 (P_{35}) el cual equivale al valor de la variable de 147.5 (localizado en el intervalo 3). Ubicamos el 35% en el eje vertical y lo proyectamos al eje horizontal. El área debajo de la curva representa el 35% de los datos.



6.6 EJERCICIOS PROPUESTOS

6.6.1 A partir de la siguiente tabla de frecuencia calcular:

Nc	Lm	Ls	f	F	h	H
1	[12	15)	13	13	26,00%	26,00%
2	[15	18)	11	24	22,00%	48,00%
3	[18	21)	9	33	18,00%	66,00%
4	[21	24)	7	40	14,00%	80,00%
5	[24	27)	6	46	12,00%	92,00%
6	[27	30]	4	50	8,00%	100,00%
TOTAL			**50**		**100,00%**	

a. Percentil 15
b. Percentil 35
c. Percentil 40
d. Percentil 85
e. Decil 2
f. Decil 6
g. Decil 8
h. Cuartil 1
i. Cuartil 2
j. Cuartil 3

6.6.2 A partir de la siguiente tabla de frecuencia calcular:

Nc	Lm	Ls	f
1	15.50	21.81	30
2	21.81	28.11	35
3	28.11	34.41	20
4	34.41	40.71	5
5	40.71	47.01	10
TOTAL			**100**

a. Mediana
b. Percentil 33
c. Decil 7
d. Cuartil 3

6.6.3 Calcular la el percentil 17 y 47 a la tabla de frecuencia mostrada en el **ejercicios 5.5.2.**

6.6.4 Calcular todos los deciles a partir de los siguientes datos:

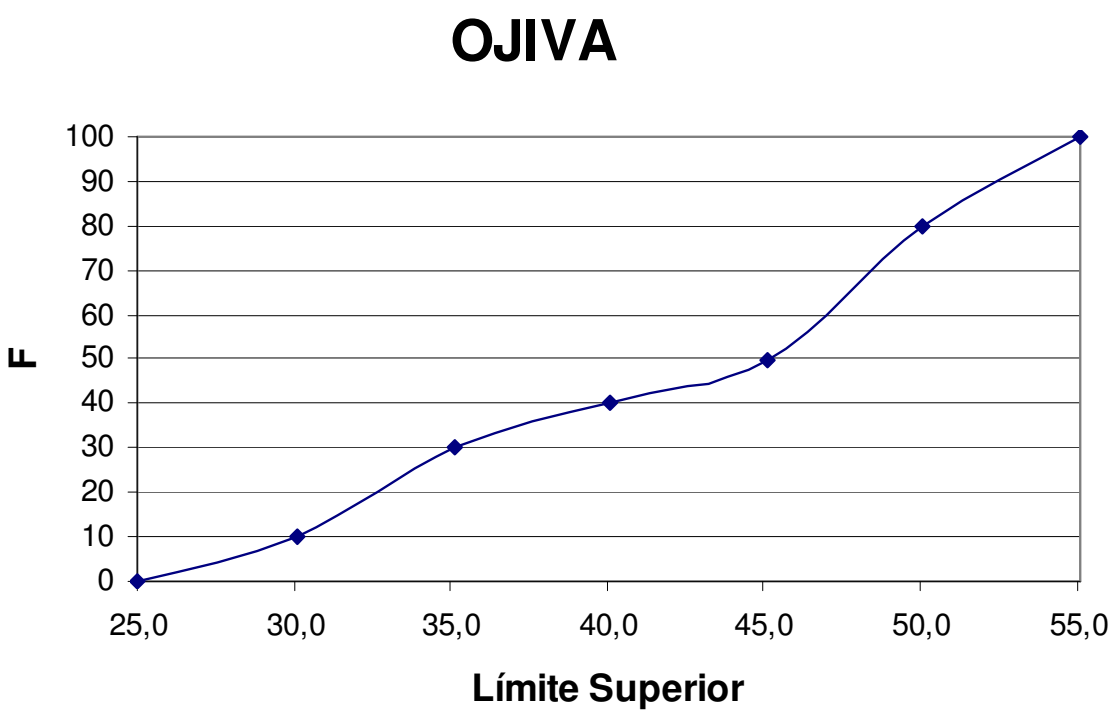
10	15	21	20	23	5	14	15
33	21	29	19	5	14	25	25
30	15	1	26	32	22	30	15
28	27	14	16	21	22	39	34

6.6.5 La siguiente tabla muestra la distribución de frecuencia sobre los ingresos de los trabajadores de un cargo en particular de varias empresas del sector manufacturero (cifras dadas en miles de pesos).

Nc	Lm	Ls	f
1	[452	460)	37
2	[460	468)	23
3	[468	476)	13
4	[476	482)	8
5	[482	490)	7
6	[490	498]	2
TOTAL			**90**

a. ¿Cuántas personas ganan menor de $ 470.000?
b. ¿Cuántas personas ganan menos de $ 482.000?
c. ¿Cuántas personas ganan más de $ 460.000?

6.6.6 Calcule y grafique el percentil 60 a partir del siguiente gráfico de ojiva (dado en el **ejercicio 4.4.6**):



6.7 CUESTIONARIO DE REPASO

Selección Múltiple con Única Respuesta: Marque con una X la respuesta correcta.

1. ¿Cuál de las siguientes equivalencias entre las medidas de posición es correcta?

A. $P_{35} = D_3$
B. $P_{75} = Q_3$
C. $P_{50} = D_6$
D. $P_{75} = Q_4$

2. ¿Cuál de las siguientes equivalencias entre las medidas de posición es correcta?

A. $Me = P_{50} = D_2 = Q_2$
B. $Me = P_{50} = D_5 = Q_4$
C. $Me = P_{50} = D_5 = Q_2$
D. $Me = P_{50} = D_6 = Q_3$

3. Dada la siguiente relación entre límites superiores consecutivos y sus frecuencias relativas acumuladas:

12.5 ⟶ 35,00%
16.5 ⟶ 72,00%

El decil 7 corresponde a:

A. 16,28
B. 15,28
C. 17,28
D. Ninguna de las anteriores

4. Dada la siguiente relación entre límites superiores consecutivos y sus frecuencias relativas acumuladas:

11,0 ⟶ 54,00%
14,0 ⟶ 63,00%

Un valor en la variable de 13,00 corresponde a un porcentaje de:

A. 55,00%
B. 57,00%
C. 60,00%
D. Ninguna de las anteriores

CAPITULO 7

MEDIDAS DE FORMA

Las medidas de forma permiten comprobar si una distribución de frecuencia tiene características especiales como simetría, asimetría, nivel de concentración de datos y nivel de apuntamiento que la clasifiquen en un tipo particular de distribución.

Las medidas de forma son necesarias para determinar el comportamiento de los datos y así, poder adaptar herramientas para el análisis probabilístico.

CAPITULO 7: MEDIDAS DE FORMA

Medidas de forma: Son indicadores estadísticos que permiten identificar si una distribución de frecuencia presenta uniformidad.

En este capitulo analizaremos dos medidas de forma:

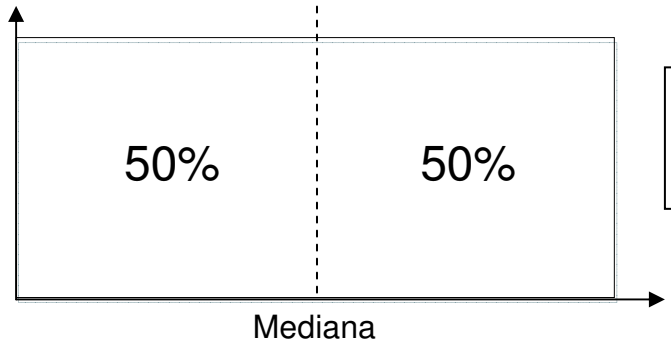
- Coeficiente de asimetría
- Curtosis

Antes de empezar con cada uno de estos indicadores, analizaremos los tipos más comunes de distribución de frecuencia y la relación media, mediana y moda como primera medida para identificar el grado de asimetría en una distribución de frecuencia.

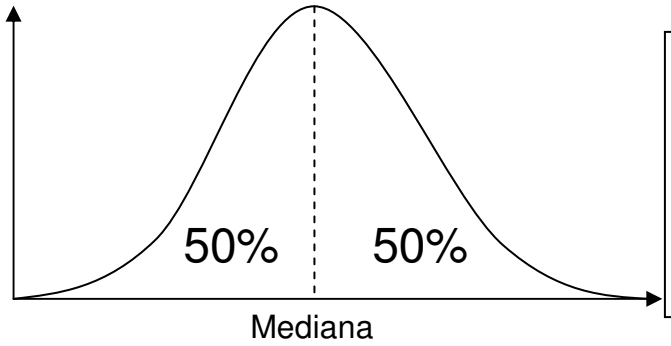
7.1 TIPOS DE DISTRIBUCIONES DE FRECUENCIA MÁS COMUNES

7.1.1 Distribución simétrica

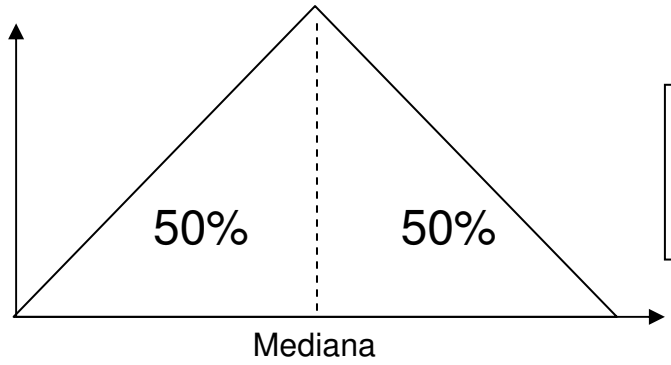
Al dividir una distribución de frecuencia mediante la mediana, ambas áreas resultantes son iguales, es decir, los datos se distribuyen de la misma forma y el área abarcada por ambos lados es equivalente (50% de los datos se encuentran distribuidos en ambas secciones).



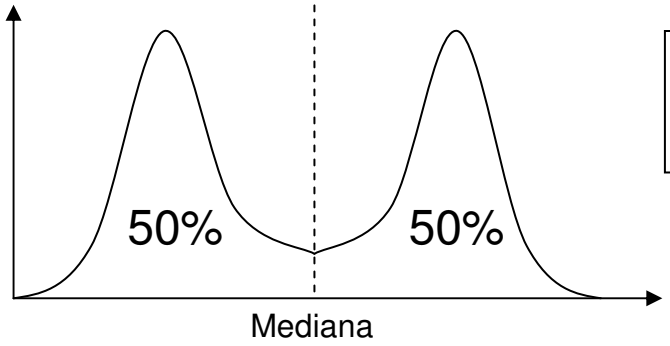
Distribución Uniforme: Las frecuencias tienen todas las mismas alturas



Distribución Simétrica: Los datos se concentran hacia el centro de la distribución. Existe una distribución simétrica con características muy definidas conocida como distribución Normal



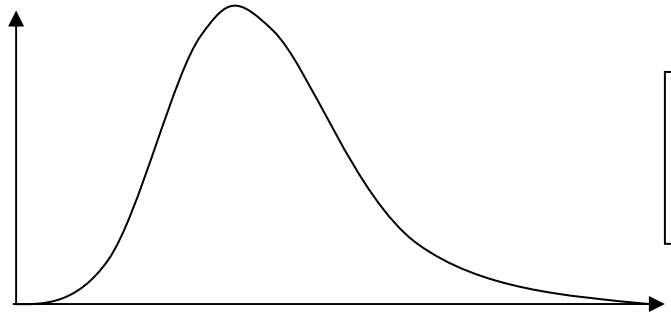
Distribución Triangular: Los datos se distribuyen dando forma a un triangulo.



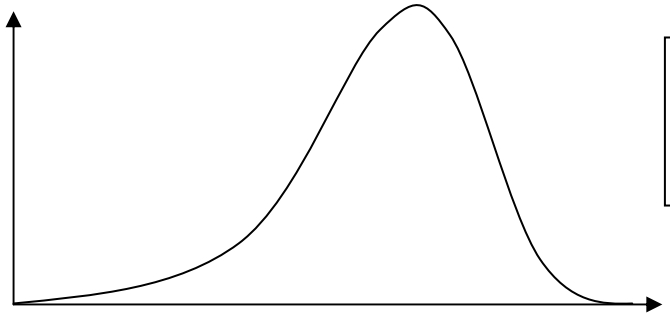
Distribución Binomial Simétrica: Presenta simetría con dos modas.

7.1.2 Distribución asimétrica

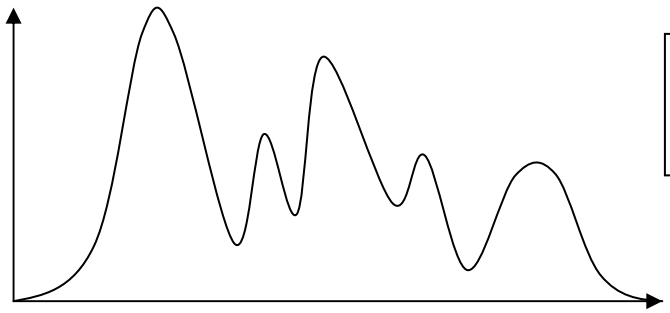
Los datos no se distribuyen de forma uniforme y similar en las áreas que dan como resultado al dividir la distribución de frecuencia por la mediana.



Distribución Sesgada hacia la Izquierda: Los datos se concentran hacia la izquierda de la distribución.



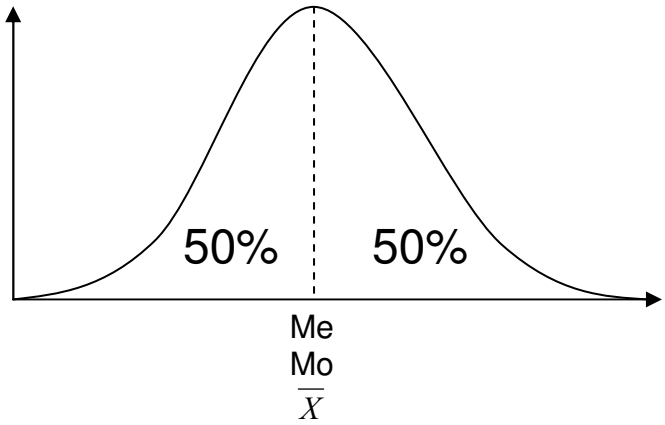
Distribución Sesgada hacia la Derecha: Los datos se concentran hacia la derecha de la distribución.



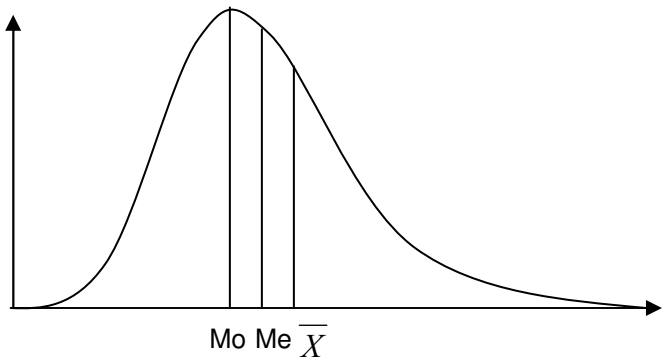
Distribución asimétrica:
No presenta uniformidad en la distribución de los datos.

7.2 RELACIÓN ENTRE LA MEDIA, MEDIANA Y MODA

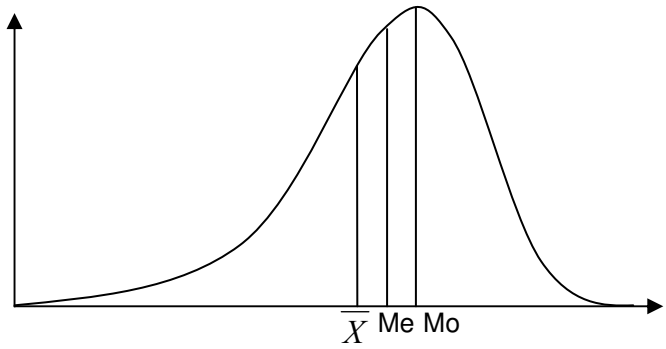
Cuando una distribución de frecuencia es simétrica, la media, mediana y moda coinciden en su valor ($\overline{X}$ = *Me* = *Mo*). En el caso de una distribución binomial simétrica, es necesario calcular el promedio de las modas.



En una distribución sesgada a la izquierda, la moda es menor a la mediana, y esta a su vez menor que la media (*Mo* < *Me* < $\overline{X}$).



En una distribución sesgada a la derecha la relación se invierte, la moda es mayor a la mediana, y esta a su vez mayor que la media (*Mo* > *Me* > $\overline{X}$).



7.2.1 Ejemplo: Relación entre la media, mediana y moda

Calcular la media, mediana y moda de los siguientes datos e interpretar su relación.

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

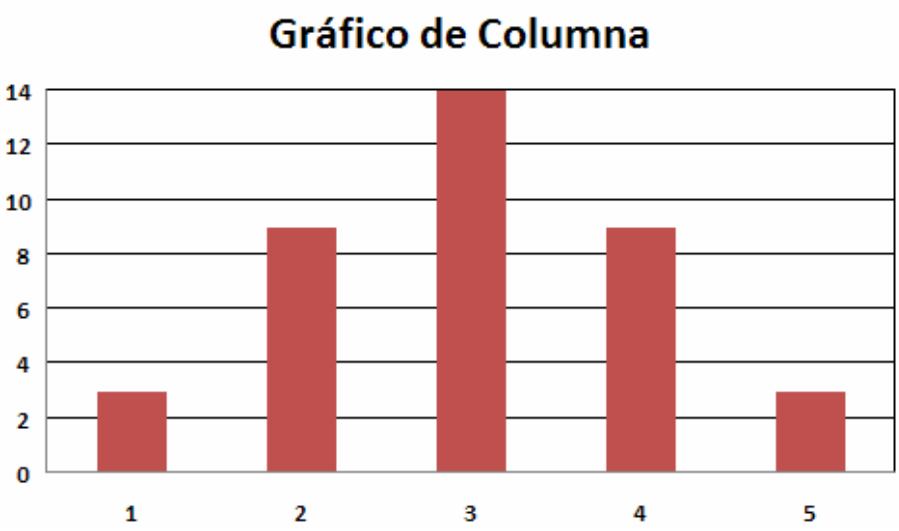
SOLUCIÓN

Se realiza el cálculo de la mediana, moda y media:

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

Media	3,0	← **=PROMEDIO(B2:G7)**
Mediana	3,0	← **=MEDIANA(B2:G7)**
Moda	3,0	← **=MODA(B2:G7)**

En este caso se deduce que fácilmente que los datos representan una distribución simétrica, como se puede observar en el gráfico de barras.



7.3 COEFICIENTE DE ASIMETRÍA

Mide el grado de asimetría de la distribución con respecto a la media. Un valor positivo de este indicador significa que la distribución se encuentra sesgada hacia la izquierda (orientación positiva). Un resultado negativo significa que la distribución se sesga a la derecha.

$$As = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{s} \right)^3 \quad o \quad As = \frac{N}{(N-1)(N-2)} \sum_{i=1}^{N} \left(\frac{X_i - \mu}{\sigma} \right)^3$$

La distribución se considera simétrica si el valor del coeficiente es cero.

7.3.1 Ejemplo: Cálculo del coeficiente de asimetría

Calcular el coeficiente de asimetría a partir de los siguientes datos obtenidos de una muestra.

5	3	1	3	3	3
3	4	3	2	3	3
1	2	3	2	3	2
2	4	2	2	2	2
1	2	1	4	2	2
3	2	3	1	2	3
1	5	6	3	2	1
1	1	2	3	2	1

SOLUCIÓN

PASO 1: Calculamos la desviación estándar de muestra.

$$S = 1{,}1468$$

PASO 2: Calculamos la diferencia de cada valor con respecto a la media, divido por la desviación y luego elevado a la 3.

$$\sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{s}\right)^3 = \left(\frac{2{,}5625}{1{,}1468}\right)^3 + \left(\frac{0{,}5625}{1{,}1468}\right)^3 + \left(\frac{-1{,}4375}{1{,}1468}\right)^3 + \ldots + \left(\frac{-1{,}4375}{1{,}1468}\right)^3$$

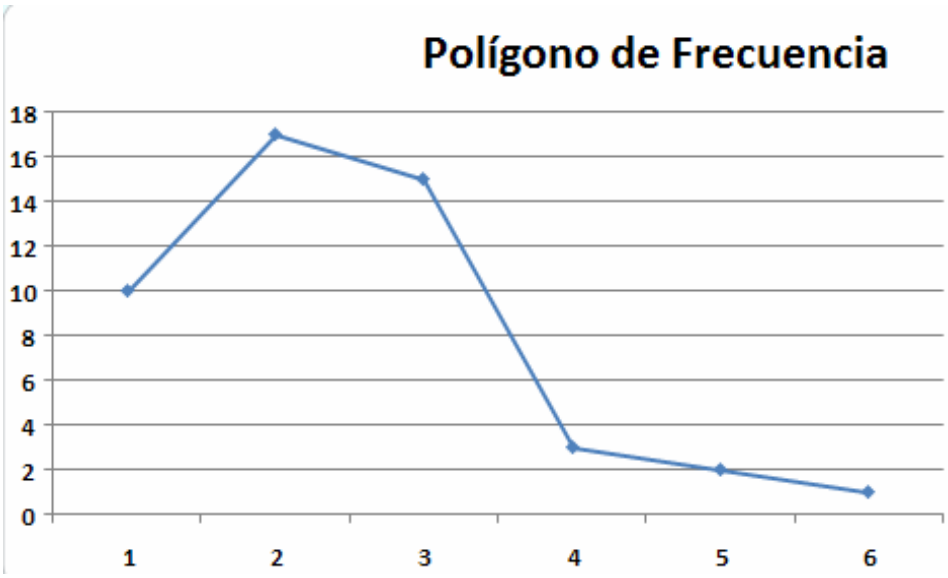
$$\sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{s}\right)^3 = 41{,}0094$$

PASO 3: Se calcula el indicador completo.

$$As = \frac{48}{(48-1)(48-2)}41{,}0094$$

$$As = 0{,}9105$$

Este valor indica que la distribución se orienta hacia la izquierda.



Para calcular este indicador en Excel, simplemente activamos la función **COEFICIENTE.ASIMETRÍA**.

COEFICIENTE.ASIMETRÍA: Devuelve el sesgo de una distribución.

Formato: COEFICIENTE.ASIMETRIA(numero1:numero2…)
Categoría: Estadística

5	3	1	3	3	3
3	4	3	2	3	3
1	2	3	2	3	2
2	4	2	2	2	2
1	2	1	4	2	2
3	2	3	1	2	3
1	5	6	3	2	1
1	1	2	3	2	1

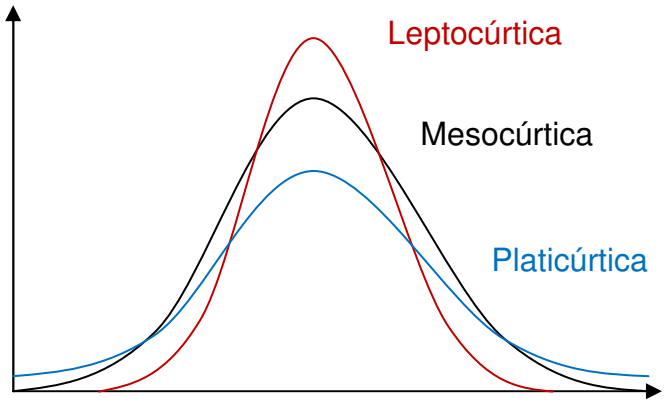
Coeficiente de Asimetría	0,9105

=COEFICIENTE.ASIMETRIA(B2:G9)

7.4 CURTOSIS

Indica que tan apuntada o achatada se encuentra una distribución respecto a un comportamiento normal (distribución normal). Si los datos están muy concentrado hacia la media, la distribución es leptocúrtica (curtosis mayor a 0). Si los datos están muy dispersos, la distribución es platicúrtica (curtosis menor a 0).

El comportamiento normal exige que la curtosis sea igual a 0 (distribución mesocúrtica).



La fórmula empleada para calcular la Curtosis se muestra a continuación (reemplace el valor de n por N en caso de tratar con datos poblacionales):

$$Curtosis = \left[\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{s} \right)^4 \right] - \frac{3(n-1)^2}{(n-2)(n-3)}$$

7.4.1 Ejemplo: Cálculo de la Curtosis

Calcular el coeficiente de asimetría a partir de los siguientes datos obtenidos de una muestra.

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

SOLUCIÓN

PASO 1: Calculamos la desviación estándar de muestra.

$$S = 1{,}1109$$

PASO 2: Calculamos la diferencia de cada valor con respecto a la media, divido por la desviación y luego elevado a la 4.

$$\sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{s}\right)^4 = \left(\frac{2{,}00}{1{,}1109}\right)^4 + \left(\frac{0{,}00}{1{,}1109}\right)^4 + \left(\frac{0{,}00}{1{,}1109}\right)^4 + \ldots + \left(\frac{0{,}00}{1{,}1109}\right)^4$$

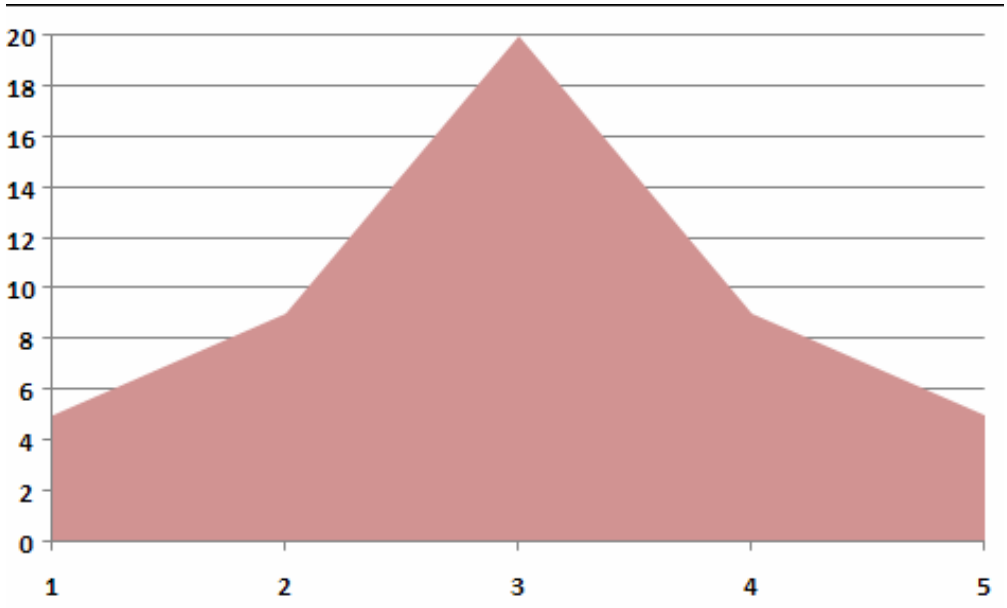
$$\sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{s}\right)^4 = 116{,}8853$$

PASO 3: Se calcula el indicador completo.

$$Curtosis = 0{,}0242x116{,}8853 - 3{,}2014$$

$$Curtosis = -0{,}3757$$

Este valor indica que la distribución es de tipo platicúrtica.



Para calcular este indicador en MS Excel, introduciremos la función llamada **CURTOSIS**.

CURTOSIS: Devuelve la Curtosis de un conjunto de datos.

Formato: CURTOSIS(numero1:numero2…)
Categoría: Estadística

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

Curtosis	-0,3757311

=CURTOSIS(B2:G9)

7.5 EJERCICIOS PROPUESTOS

7.5.1 Determinar las fórmulas del Coeficiente de Asimetría y Curtosis que puedan se utilizados en tablas tipo A y tipo B.

7.5.2 Calcular e interpretar el coeficiente de asimetría y Curtosis para los siguientes datos (realizar gráfico de barra).

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

7.5.3 Calcular e interpretar el coeficiente de asimetría y Curtosis para los siguientes datos (realizar histograma).

101,16	100,42	96,58	99,96	96,22	100,26	99,46	98,28
102,63	97,40	102,69	101,36	102,72	99,13	101,63	100,00
99,19	98,28	104,63	98,39	97,56	99,80	99,83	102,06
102,06	100,87	97,64	99,16	99,25	101,74	100,33	103,55
103,55	99,45	102,68	98,81	99,82	100,74	104,63	98,39
104,14	103,51	100,87	100,13	97,08	101,31	97,64	99,16

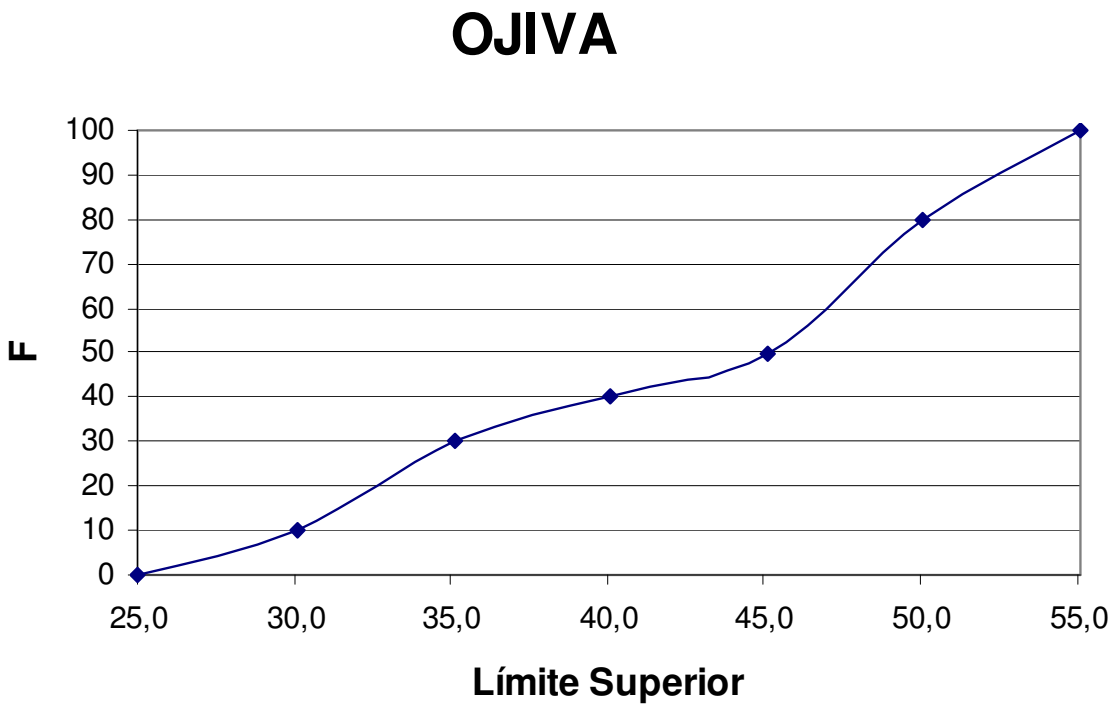
7.5.4 Calcular e interpretar el coeficiente de asimetría y Curtosis para los siguientes datos (realizar gráfico de barra o histograma).

2	13	4	5	11	8	7
7	4	5	6	4	2	7
6	5	8	7	9	4	5
4	5	6	7	6	5	7
3	7	3	4	10	6	6
3	8	8	6	3	7	11
5	6	5	11	2	8	6
9	6	9	9	9	6	3
3	2	8	8	7	2	8

7.5.5 Calcular e interpretar el coeficiente de asimetría y Curtosis a los datos mostrados en los **ejercicios 4.4.1, 4.4.2, 4.4.3 y 4.4.4**

7.5.6 Calcular e interpretar la Curtosis para el **ejemplo 7.3.1**.

7.5.7 Calcular e interpretar el coeficiente de asimetría y Curtosis a partir del siguiente gráfico de ojiva (dado en el **ejercicio 4.4.6**):



ANEXO A. RESUMEN DE SIMBOLOGÍA Y FÓRMULAS ESTADÍSTICAS UNIDAD II

Nombre	Símbolo	Fórmula
Ancho del Intervalo de Clase	A	$A = L_s - L_m$ ó $A = R / Nc$
Frecuencia Absoluta	f	
Frecuencia Absoluta Acumulada	F	$F_i = F_{i-1} + f_i$
Frecuencia Relativa	h	$h_i = \frac{f_i}{n}$
Frecuencia Relativa Acumulada	H	$H_i = H_{i-1} + h_i$
Límite Inferior del Intervalo	Lm	
Límite Superior del Intervalo	Ls	
Marca de Clase	Mc	$M_c = \frac{L_s + L_m}{2}$
Número de Clases o de Intervalos de Clase	Nc	$Nc = \sqrt{n}$ ó $1 + 3{,}22 \log n$
Nuevo Rango	R’	$R' = X_{\max}' - X_{\min}'$ ó $R' = A\ x\ Nc$
Rango	R	$R = X_{\max} - X_{\min}$
Valor Máximo	Xmax	
Valor Mínimo	Xmin	
Valor Máximo Ajustado	Xmax’	
Valor Mínimo Ajustado	Xmin’	

ANEXO B. RESUMEN DE SIMBOLOGÍA Y FÓRMULAS ESTADÍSTICAS UNIDAD IV

Nombre	Símbolo	Fórmula
Media Aritmética para datos Poblacionales no agrupados	μ	$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$
Media Aritmética para datos Muestrales no agrupados	$\overline{X}$	$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$
Media Aritmética para datos Poblacionales agrupados	μ	$\mu = \frac{\sum_{i=1}^{Nc} X_i f_i}{N}$ ó $\mu = \frac{\sum_{i=1}^{Nc} Mc_i f_i}{N}$
Media Aritmética para datos Muestrales agrupados	$\overline{X}$	$\overline{X} = \frac{\sum_{i=1}^{Nc} X_i f_i}{n}$ ó $\overline{X} = \frac{\sum_{i=1}^{Nc} Mc_i f_i}{n}$
Mediana	*Me*	$Me = Ls_{i-1} + A.\frac{(50\% - H_{i-1})}{h_i}$ ó $Me = Ls_{i-1} + A.\frac{(\frac{n}{2} - F_{i-1})}{f_i}$
Moda	*Mo*	$Mo = L_{S-1} + A.\frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i-1})}$
Tamaño de la muestra	*n*	
Tamaño de la Población	*N*	

ANEXO C. RESUMEN DE SIMBOLOGÍA Y FÓRMULAS ESTADÍSTICAS UNIDAD V

Nombre	Símbolo	Fórmula
Desviación Media para datos Poblacionales no agrupados	*Dm*	$Dm = \frac{\sum_{i=1}^{N} \lvert X_i - \mu \rvert}{N}$
Desviación Media para datos Muestrales no agrupados	*Dm*	$Dm = \frac{\sum_{i=1}^{n} \lvert X_i - \mu \rvert}{n}$
Desviación Media para datos Poblacionales agrupados	*Dm*	$Dm = \frac{\sum_{i=1}^{Nc} \lvert X_i - \mu \rvert . f_i}{N}$ $Dm = \frac{\sum_{i=1}^{Nc} \lvert Mc_i - \mu \rvert . f_i}{N}$
Desviación Media para datos Muestrales agrupados	*Dm*	$Dm = \frac{\sum_{i=1}^{Nc} \lvert X_i - \overline{X} \rvert . f_i}{n}$ $Dm = \frac{\sum_{i=1}^{Nc} \lvert Mc_i - \overline{X} \rvert . f_i}{n}$
Varianza para datos Poblacionales no agrupados	σ^2	$\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$
Varianza para datos Muestrales no agrupados	S^2	$S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}$
Varianza para datos Poblacionales agrupados	σ^2	$\sigma^2 = \frac{\sum_{i=1}^{Nc} (X_i - \mu)^2 . f_i}{N}$ $\sigma^2 = \frac{\sum_{i=1}^{Nc} (Mc_i - \mu)^2 . f_i}{N}$

Nombre	Símbolo	Fórmula
Varianza para datos Muestrales agrupados	S^2	$S^2 = \frac{\sum_{i=1}^{Nc} (X_i - \overline{X})^2 . f_i}{n}$ $S^2 = \frac{\sum_{i=1}^{Nc} (Mc_i - \overline{X})^2 . f_i}{n}$
Desviación estándar o típica para datos Poblacionales no agrupados	σ	$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}}$
Desviación estándar o típica para datos Muestrales no agrupados	S	$S = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n}}$
Desviación estándar o típica para datos Poblacionales agrupados	σ	$\sigma = \sqrt{\frac{\sum_{i=1}^{Nc} (X_i - \mu)^2 . f_i}{N}}$ $\sigma = \sqrt{\frac{\sum_{i=1}^{Nc} (Mc_i - \mu)^2 . f_i}{N}}$
Desviación estándar o típica para datos Muestrales agrupados	S	$S = \sqrt{\frac{\sum_{i=1}^{Nc} (X_i - \overline{X})^2 . f_i}{n}}$ $S = \sqrt{\frac{\sum_{i=1}^{Nc} (Mc_i - \overline{X})^2 . f_i}{n}}$
Coeficiente de variación	Cv	$Cv = \frac{S}{\overline{X}}$ o $Cv = \frac{\sigma}{\mu}$

ANEXO D. RESUMEN DE SIMBOLOGÍA Y FÓRMULAS ESTADÍSTICAS UNIDAD VI

Nombre	Símbolo	Fórmula
Percentil k (k = 1 a 100)	P_k	$P_k = Ls_{i-1} + A.\frac{(k - H_{i-1})}{h_i}$ ó $P_k = Ls_{i-1} + A.\frac{(n.k - F_{i-1})}{f_i}$
Decil k (k = 1 a 10)	D_k	$D_k = Ls_{i-1} + A.\frac{(10\%.k - H_{i-1})}{h_i}$ ó $D_k = Ls_{i-1} + A.\frac{(10\%.n.k - F_{i-1})}{f_i}$
Cuartil k (k = 1, 2, 3 o 4)	Q_k	$Q_k = Ls_{i-1} + A.\frac{(25\%.k - H_{i-1})}{h_i}$ ó $Q_k = Ls_{i-1} + A.\frac{(25\%.n.k - F_{i-1})}{f_i}$

ANEXO E. RESUMEN DE SIMBOLOGÍA Y FÓRMULAS ESTADÍSTICAS UNIDAD VII

Nombre	Símbolo	Fórmula
Coeficiente de Asimetría para datos muestrales y poblacionales	*As*	$As = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{s} \right)^3$ O $As = \frac{N}{(N-1)(N-2)} \sum_{i=1}^{N} \left(\frac{X_i - \mu}{\sigma} \right)^3$
Curtosis para datos muestrales y poblacionales	*Curtosis*	$Curtosis = \left[\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{s} \right)^4 \right] - \frac{3(n-1)^2}{(n-2)(n-3)}$ O $Curtosis = \left[\frac{N(N+1)}{(N-1)(N-2)(N-3)} \sum_{i=1}^{N} \left(\frac{X_i - \mu}{\sigma} \right)^4 \right] - \frac{3(N-1)^2}{(N-2)(N-3)}$

ÍNDICE GENERAL DE PALABRAS

C

Censo, 13
Ciencia del Estado, 13
Clasificación de la estadística, 15
Coeficiente de variación, 153
Computing Tabulating Recording Corporation (CTR), 14
Consejo de estado, 13
Cuartíl, 166
Curtosis, 183
Curva de frecuencias, 85
Curvas suavizadas, 85

D

Datos, 13
Decíl, 163
Desviación estándar o típica, 146
Desviación media, 131, 133
Distribución asimétrica, 178
Distribución binomial simétrica, 178
Distribución leptocúrtica, 183
Distribución mesocúrtica, 183
Distribución normal, 183
Distribución platicúrtica, 183
Distribución sesgada a la izquierda, 179
Distribución sesgada hacia la derecha, 178
Distribución sesgada hacia la izquierda, 178
Distribución simétrica, 177
Distribución triangular, 178
Distribución uniforme, 177

E

Epidemiología, 14
Estadista, 13
Estadística, 13, 14
Estadística descriptiva o deductiva, 15
Estadística inferencial o inductiva, 15
Estadístico, 14

F

Frecuencia Absoluta, 18
Frecuencia Absoluta Acumulada, 20
Frecuencia Relativa, 21
Frecuencia Relativa Acumulada, 22

G

Gottfried Achenwall, 13
Gráfico de columna, 69
Gráfico de sectores, 14, 60
Gráficos Estadísticos, 60

H

Herman Hollerith, 14
Herramientas cuantitativas, 13
Histograma, 76

I

International Business Machines (IBM), 14
Intervalo de clase, 32
Investigación estadística, 16

M

Major Greenwwod, 14
Máquina tabuladora, 14
Media aritmética, 104
Mediana, 114
Medidas de dispersión, 131
Medidas de posición, 160
Medidas de tendencia central, 104
Moda, 121
Moisés, 13
Muestra, 15

N

Nuevo Rango, 35
Numero de intervalos, 32

O

Ojiva, 86

P

Padre de la Estadística, 13
Percentíl, 160
Pictograma, 94
Playfair's Statistical Breviary"., 14
Polígono de frecuencias, 81
Político, 13

R

Rango, 19
Relación entre media, mediana y moda, 179, 181

S

Series de tiempo, 14
Sir Francis Galton, 14
Sir John Sinclair, 13
Statista, 13
Statistical Account of Scotland, 13
Statisticum collegium, 13
Statistik, 13
Stato, 13
Status, 13

T

Tabla de frecuencia tipo A, 18
Tablas de Frecuencia, 18
Tablas de frecuencia tipo B, 33
Tarjetas perforadas, 14
The Commercial and Political Atlas, 13
Thomas J. Watson, 14

V

Varianza, 140

W

William Playfair, 13

BIBLIOGRAFÍA

BELLO PÉREZ, Carlos. Pronósticos Empresariales: Como Proyectar su Empresa al Futuro. Ecoe Ediciones. Colombia, año 2000.

CANSADO, Enrique. Estadística General. Centro Interamericano de Enseñanza de Estadística, Segunda Edición. Chile, año 1970.

CHRISTENDEN, Howard B. Estadística Paso a Paso. Editorial Trillas. México, año 2004.

DIXON, John R. Introducción a la Probabilidad. Editorial Limusa - Wiley. México, 1970.

ENCICLOPEDIA BRITÁNICA. (2007, 1 3). *ENCICLOPEDIA BRITANNICA ONLINE.* Retrieved 1 3, 2007 from http://www.britannica.com: http://www.britannica.com/eb/article-9108592/statistics.

EUMED. (2007, 1 2). *Enciclopedia y biblioteca virtual de economía.* Retrieved 1 2, 2007 from http://www.eumed.net: http://www.eumed.net/cursecon/dic/dent/index.htm.

GRASSAU S., Erika. Elementos de Estadística. Editorial Universitaria. Chile, 1960.

GUENTHER, William C. Introducción a la Inferencia Estadística. Mc Graw Hill. Panamá, año 1968.

NAVIDI, William. Estadística para Ingenieros y Científicos. Mc Graw Hill, primera edición. México, 2006.

SUKHATME, Pandurang. Teoría de Encuestas por Muestreo con Aplicaciones. Fondo de Cultura Económica. México, año 1962.

STEVENSON, William. Estadística para Administración y Economía: Conceptos y Aplicaciones. Editorial AlfaOmega, primera edición. México, 2004.

VASQUEZ, Tulio. Problemario de Estadística Económica. Universidad Central de Venezuela. Venezuela, año 1964.

VESSEREAU, André. La Estadística. Editorial Universitaria de Buenos Aires, Séptima Edición. Argentina, año 1970.

WALPOLE, Ronald E. Probabilidad y Estadística. Mc Graw Hill, Cuarta Edición. México, año 1992.

WIKIPEDIA. (2007, 1 2). *Biografía de Sir John Sinclair.* Retrieved 1 2, 2007 from http://en.wikipedia.org/: http://en.wikipedia.org/wiki/Sir_John_Sinclair C_1st_Baronet.

WORDREFERENCE.COM. (2007, 1 3). *WORDREFERENCE.COM English Dictionary.* Retrieved 1 3, 2007 from http://www.wordreference.com: http://www.wordreference.com/definition/statistics.

YAHOO. (2007, 1 3). *YAHOO EDUCATION.* Retrieved 1 3, 2007 from http://education.yahoo.com/: http://education.yahoo.com/reference/dictionary/entry/statistics.