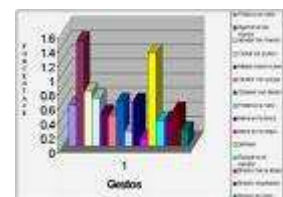
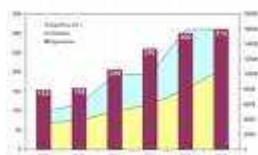
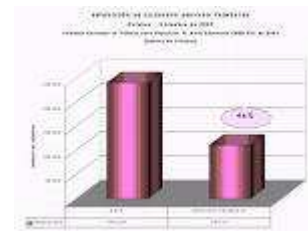
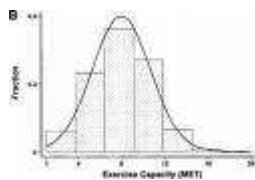
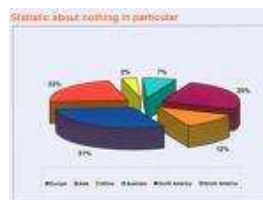
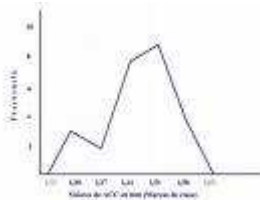




**UNIVERSIDAD NACIONAL DE LOJA**  
**ÁREA DE LA EDUCACIÓN, EL ARTE Y LA COMUNICACIÓN**

# **TALLER DE**

# **ESTADÍSTICA DESCRIPTIVA**



# **APLICADA A LA**

# **INVESTIGACIÓN**

## **MÓDULO 1**

COMPILACIÓN:  
MG. JORGE ARMIJOS ORDÓÑEZ

Loja-Ecuador  
2010

## CONTENIDO

<i>PORTADA</i> .....	1
1. <i>PRESENTACIÓN</i> .....	2
2. <i>OBJETIVOS</i> .....	2
3. <i>METODOLOGÍA</i> .....	2
4. <i>EVALUACIÓN</i> .....	2
5. <i>ACREDITACIÓN</i> .....	3
6. <i>BIBLIOGRAFÍA</i> .....	3
7. <i>CONTENIDOS</i> .....	4

### 1. PRESENTACIÓN

El Taller de “ESTADÍSTICA DESCRIPTIVA APLICADA A LA INVESTIGACIÓN” está dirigido a los estudiantes del módulo uno, del Área de la Educación, el Arte y la Comunicación, como un instrumento para la investigación, que permita el manejo e interpretación de la información empírica en los procesos investigativos que ejecuten en cada uno de los módulos de la formación profesional y luego en su práctica profesional.

Se desarrollará bajo la orientación conceptual y metodológica que orienta el Sistema Académico Modular por Objetos de Transformación SAMOT, e impulsa con rigurosidad la aplicabilidad de los fundamentos científicos, considerando las necesidades de los estudiantes e incorpora la creatividad, valores y desarrollo de la inteligencia como ejes transversales, sin perder la alegría, la creatividad y la interlocución.

### 2. OBJETIVOS

#### 2.1. OBJETIVO GENERAL

Comprender el papel que cumple la Estadística en la investigación, regocijarse con su uso y reconocer el valor de la recopilación, organización, procesamiento, análisis, interpretación y presentación de la información con precisión y perseverancia en la búsqueda de soluciones.

#### 2.2. ESPECÍFICOS

- Reconocer la importancia de la estadística en la investigación.
- Distinguir la importancia de población y muestra.
- Explicar la diferencia entre estadística descriptiva e inferencial.
- Identificar diversos tipos de variables.
- Comprender y utilizar el cálculo de medidas de tendencia central, asociación y dispersión en pequeños y grandes grupos de datos.
- Interpretar información a través de gráficos estadísticos.

### 3. METODOLOGÍA

El proceso será participativo, dinámico, integral y productivo, donde la aprehensión de la realidad que modifica los esquemas de los participantes constituye los aprendizajes significativos.

El trabajo académico, se apoya en procesos y técnicas grupales que promuevan la comunicación en el grupo y el logro de conocimientos individuales significativos que modi-

fiquen los esquemas mentales

Los asistentes al curso cuentan con un documento guía que facilita y habilita la participación en su proceso de formación, así como el uso de la INTERNET y programas de computación.

#### 4. EVALUACIÓN

El proceso de evaluación será permanente y sistemático que permitirá la reorientación y mejoramiento del interaprendizaje, mediante la observación directa por parte del docente y la participación fundamentada los estudiantes.

El taller es de aprobación independiente y se enmarca en el Capítulo II “De la Evaluación, Acreditación y Calificación de los Aprendizajes” del Reglamento de Régimen Académico de la UNL.

#### 5. ACREDITACIÓN

La acreditación se cumplirá a través de la verificación del cumplimiento de:

Tareas extra clase	30%
Trabajo en clase	30%
Pruebas	<u>40%</u>
<b>TOTAL</b>	<b>100%</b>

Debiendo alcanzar un promedio mínimo de 7/10 para aprobar el taller.

#### 6. BIBLIOGRAFÍA

- 6.1. BUSTAMANTE, Luna, 2001, *Estadística Descriptiva*, UTPL, Loja-Ecuador.
- 6.2. DOWNIE y Eath, 1980, *Métodos Estadísticos aplicados*, Editorial Harla, México.
- 6.3. FREUND, Jhon, *Estadística Elemental Moderna*, Editorial Revolucionara
- 6.4. JACK, Levin, *Fundamentos de Estadística en la Investigación Social*.
- 6.5. SPIEGEL, Murray, 1991, *Estadística, McGraw-Hill, Segunda Edición*
- 6.6. VELASCO, Ermel, *Estadística para Educadores y Psicólogos*.
- 6.7. VIZUETE, Alejandro, *Elementos de Estadística aplicados a la Educación*.
- 6.8. <http://www.uaq.mx/matematicas/estadisticas/xu3.html>
- 6.9. <http://www.fisterra.com/mbe/investiga/10descriptiva/10descriptiva.htm>
- 6.10. <http://www.aulafacil.com/CursoEstadistica/Lecc-1-est.htm>

## 7. CONTENIDOS

### INTRODUCCIÓN

Los estudiantes de educación, ha menudo hacen frente al estudio de la estadística con emociones diversas. Ellos saben, de que no pueden seguir adelante a estudios avanzados en sus campos sin un entendimiento básico de la estadística. No es una exageración al afirmar que es necesario entender la estadística en el estudio de las ciencias; el desarrollo de las ideas estadísticas ha sido tal en los últimos años, que esta presente en casi todas las actividades humanas.

La estadística no es simplemente la compilación y presentación de datos en tablas y gráficos, constituye en realidad **la ciencia de tomar decisiones en presencia de la incertidumbre**. Nos enfrentamos con la incertidumbre cuando lanzamos una moneda, cuando experimentamos con un medicamento nuevo, cuando tratamos de decidir cuál de dos procesos de producción es más eficiente, cuando tratamos de predecir los resultados de una lección, cuando un nuevo producto al mercado, cuando cavamos buscando oro, cuando tratamos de decidir si hacemos un viaje a las montañas o a la playa.

Sería presuntuoso decir que la estadística en su estado actual de desarrollo, pueda resolver todas las situaciones que implique inseguridad, pero se están desarrollando métodos nuevos todo el tiempo y la estadística moderna con la ayuda de la informática, proporciona la información necesaria para el análisis en forma lógica y sistemática.

### 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 volúmenes. 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.<sup>1</sup>

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 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 Greenwood (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.

## DEFINICIÓN

Todo lo relacionado, con la recopilación, manipulación, análisis, interpretación y presentación de datos numéricos puede ser clasificado como perteneciente al dominio de la Estadística. Incluye tareas tan diversas como el cálculo del promedio de las calificaciones de un estudiante, la compilación y presentación de datos, sobre nacimientos, defunciones, la evaluación de la precisión, de los proyectiles dirigidos y también el estudio de las leyes que gobiernan el comportamiento de las partículas atómicas dentro de lo que llamamos mecánica estadística.

La palabra Estadística, se utiliza de diversas maneras:

- En el plural, señalan una compilación de datos numéricos, tales como los que se encuentran en las páginas financieras de los periódicos.
- El segundo significado, es la totalidad de los métodos utilizados en la compilación y análisis de datos en la ciencia que permite tomar decisiones.

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 comprensión 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.

## CLASIFICACIÓN

En este sentido la estadística es una rama de las matemáticas aplicadas, que para su mayor comprensión se han dividido: en Estadística Descriptiva, y Estadística Inferencial.

### ESTADÍSTICA DESCRIPTIVA

Debemos entender todos aquellos cálculos y técnicas empleadas para recolectar, manipular, analizar, interpretar y presentar un conjunto de datos.

### ESTADÍSTICA INFERENCIAL

Es aquella que permite determinar a partir de un conjunto de datos, conclusiones, generalizaciones, predicciones, estimados, y permite tomar decisiones en relación con la incertidumbre.

La Estadística utiliza dos tipos de medios matemáticos, para la descripción y análisis de los hechos, se apoya en el cálculo numérico elemental; pero cuando se requiere tomar decisiones se fundamenta en las teorías de las probabilidades. La Estadística es lógica en combinación con procedimientos aritméticos, la lógica le proporciona el método para la toma de datos y la aritmética le proporciona el material sobre el que ha de realizar la inferencial.

La evaluación, el análisis y el control cuidadoso de los riesgos que hay que tomar cuando hacemos generalizaciones o decisiones es una de las principales tareas de la Estadística Inferencial.

### POBLACIÓN

El término población se refiere a la totalidad de observaciones, datos o medidas que se consideren en una situación dada.

**Ejemplos:**

- En un estudio de la preferencia de los votantes en una elección presidencial, la población está constituida por todas las personas empadronadas.
- En un estudio de mercado una investigación de las ventas anuales de los supermercados, la población está constituida por todas las ventas de todos los supermercados.

Una población puede ser finita o infinita.

**Ejemplos:**

- Los alumnos matriculados en el Área de Educación, Arte y Comunicación de la Universidad Nacional de Loja, constituyen una población finita.
- Los alumnos universitarios del Ecuador, constituye una población infinita.

La característica de una población recibe el nombre de **parámetros**.

**Ejemplo:**

- La media aritmética de las estaturas de las alumnas y alumnos del módulo 5 del Área de Educación, Arte y Comunicación de la Universidad Nacional de Loja es un parámetro de la población de los estudiantes del quinto módulo de todos los alumnos de la Universidad Nacional de Loja.

Es claro que, calcular un parámetro de una población con un gran número de elementos es un proceso bastante difícil y en ciertos casos imposibles, por tal motivo es necesario extraer una parte de ella, llamada muestra.

**MUESTRA**

Los datos pueden provenir de una población o muestra. Estos 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.

Es la recolección de una parte de la población, de tal manera que se pueda hacer inferencias con respecto a la población.

**Ejemplos:**

- En un estudio de la preferencia de los votantes en una elección presidencial, la muestra está constituida por las personas empadronadas en la ciudad de Loja.

- En un estudio de mercado una investigación de las ventas anuales de los supermercados, la muestra está constituida por las ventas de los supermercados de Guayaquil.

Una característica de la muestra se conoce como **estadígrafo**.

**Ejemplo:**

- La estatura promedio de los alumnos del séptimo módulo paralelo "A".

En este contexto podemos definir que la inferencia estadística es una técnica, mediante la cual podemos determinar conclusiones, generalizaciones, acerca de los parámetros de la población, basándonos en las informaciones proporcionadas por los estadígrafos de una muestra de esa población.

Considerando las definiciones anteriores, podemos determinar que "la Estadística es un conjunto de teorías y métodos que han sido desarrollados para tratar la recolección, el análisis y la descripción de datos muestrales, con el fin de extraer conclusiones útiles".

Su función principal es apoyar al investigador, en la toma de decisiones sobre el parámetro de la población de la cual procede la muestra.

**MUESTRA REPRESENTATIVA**

Cuando la población es excesivamente grande, cuando el tiempo requerido para observar los elementos es más largo de lo que se desea, cuando el costo de la observación total es superior a los recursos disponibles, o cuando la observación total implica la destrucción de los elementos de la población, es necesario recoger información sólo de una parte de la población, es decir, de una muestra.

Con la muestra se persigue no estudiar la propia muestra, sino estudiar la población, a través de ella.

**Ejemplo:**

- En un laboratorio se recoge una gota de sangre (muestra) para observar la composición de toda la sangre (población).

Para seleccionar una muestra debe tenerse cuidado de que la misma sea realmente representativa de la población y no una muestra sesgada. Esto quiere decir, que cada elemento de la población debe tener igual oportunidad de ser seleccionado, lo cual se consigue aproximadamente de la siguiente manera.

1. Se numera cada miembro de la población con la finalidad de que pueda ser identificado mediante un código numérico.
2. Se escribe cada código numérico en fichas iguales que se colocan en un recipiente convenientemente revueltas.
3. Se selecciona al azar los datos que conforman la muestra.

Para economizar tiempo, se dispone en los libros de Estadística tablas de números aleatorios.



## APLICACIÓN

Los métodos estadísticos han demostrado ser de gran utilidad en una amplia gama de estudios entre los cuales están la educación y la psicología.

Para cumplir con los objetivos utiliza como métodos:

**Recolección de datos.**- que consiste en la obtención de datos relacionados con el problema motivo de estudio, utilizando instrumentos tales como:

- Fichas de observación
- Cuestionarios
- Entrevistas
- Informes
- Memorias, etc.

### Ejemplos:

Entre los estudiantes que pertenecen al Área de la Educación, el Arte y la Comunicación, podemos recoger datos correspondientes: colegio donde provienen, nivel de lectura y presentar esta información en una tabla de datos que facilite su interpretación.

NÓMINA	COLEGIO		NIVEL DE LECTURA	
	FISCAL	PRIVADO	EFICIENTE	DEFICIENTE

- Un educador desea saber si hay relación significativa entre las puntuaciones de un test de aptitud escolar y las calificaciones promedio de un grupo de estudiantes. Si existiera una relación semejante, podría predecir la calificación promedio de un estudiante, basándose en la puntuación de su test de aptitud.
- Es posible comparar dos métodos de enseñanza diferentes para determinar sus eficacias relativas. Se pueden elegir dos grupos de estudiantes y enseñarles por los dos métodos una materia determinada; al final del período de instrucción se administra un test típico a ambos grupos. Basándose en la puntuación promedio obtenida por cada grupo en el test, se puede llegar a una conclusión sobre si los métodos son igualmente efectivos, o si uno de los métodos es significativamente más efectivo que el otro.

Desde hace varios años, el número de estudios y de encuestas sobre materia educativa han aumentado considerablemente. La necesidad de analizar e interpretar esa gran cantidad de información obliga a los estudiantes de educación a tener por lo menos el conocimiento básico de los métodos estadísticos.

## 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.

## DATOS ESTADÍSTICOS

Resulta evidente que toda investigación estadística se ha de referir a un conjunto de personas, animales o cosas. Los elementos que conforman este conjunto pueden ser algo con existencia real como un automóvil, una casa, o algo más abstracto como un voto o un intervalo de tiempo. Los elementos pueden ser también artificiales, creados con sólo el propósito de la investigación, por ejemplo, al estudiar un campo sembrado de trigo es usual dividir el campo en cuadrados o rectángulos, pasando éstos a constituir los elementos de la población.

**CARACTERES.-** Los elementos de la población estadística poseen rasgos, propiedades o cualidades que denominaremos caracteres.

### Ejemplo:

El hombre tiene caracteres tales como: estatura, edad, peso, estado civil, religión, etc.

**VARIABLE.-** Es aquella que puede asumir diferentes valores

### Ejemplo:

- Cuando una característica tal como la edad,

Las variables pueden dividirse a su vez en dos clases:

1. **Las variables cuantitativas** son aquellas que producen valores o datos que se pueden describir mediante números, por ejemplo, la edad de un alumno: 15 años.
2. **Las variables cualitativas** son aquellas que producen datos cuya descripción sólo puede hacerse mediante palabras, por ejemplo, el estado civil de un alumno: soltero.

Los datos de variable cualitativa no son inherentemente numéricos como lo son aquellos que pertenecen a variables cuantitativas, por tanto, se deben convertir a valores numéricos antes de someterlos a cualquier proceso estadístico.

Las variables cualitativas admiten categorías

### Ejemplo

El sexo: masculino y femenino.

Cuando se aplican a los elementos de una población, siempre será posible asignar cada elemento del conjunto a una de dichas categorías y contar luego el número que cae en cada una de ellas. Es a dichos números que se debe aplicar los procesos estadísticos pertinentes.

**Las variables cuantitativas**, a su vez, pueden dividirse en variables discretas y variables continuas:

**Variables discretas.-** Son aquellas que pueden asumir sólo ciertos valores, por lo regular enteros, y no es posible que lleguen a tomar ningún valor comprendido entre dos consecutivos de aquellos.

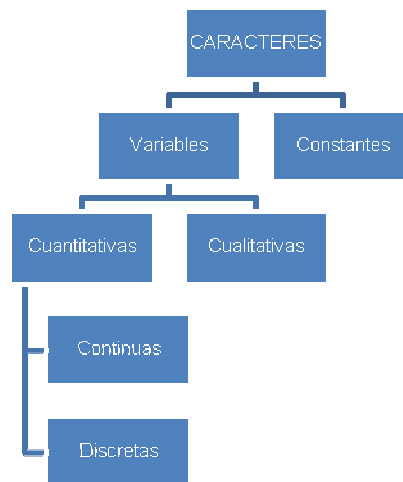
**Ejemplo:**

El número de hermanos de un alumno puede ser: 0, 1, 3, etc., pero no podrá ser 1,5.

**Variables continuas.-** Son aquellas que pueden asumir teóricamente todos los infinitos valores a lo largo de un cierto intervalo.

**Ejemplo:**

La edad de un alumno podría ser cualquier valor entre los 20 y 21 años, por ejemplo, y podría medirse o darse con cualquier grado de exactitud que se desee: 20,3 años, es decir, puede darse utilizando incluso números decimales.



## CONSTANTE

Cuando la medida de un atributo no cambia para los diferentes miembros de una población, no es una variable y se denomina constante.

**Ejemplo:**

Todos los alumnos de Módulo 5 de QQ-BB, tienen como constante “ser alumnos del 1er. Curso”.

**LÍMITES REALES.-** la medida de una observación o dato que pertenece a una variable continua puede darse con cualquier grado de aproximación que se desee; más, si la edad de un alumno ha de darse con aproximación al año más cercano

**Ejemplo:**

Aquel que tiene 20,3 años deberá considerarse como de 20 años de edad; igualmente, el alumno que tenga 19,6 años deberá considerarse de 20 años de edad. En ambos casos las edades son “realmente” de 20,3 años y 19,6 años, respectivamente. Sin embargo, habiéndose acordado dar las edades redondeando al año más cercano, las edades de dichos alumnos se ha dicho que es 20 años. Supongamos además que se

ha determinado que la edad de un alumno es de 19,5 años, ¿qué edad reportaríamos si seguimos aproximando al año más cercano? Las edades de 19 y 20 años, serían igualmente adecuadas; sin embargo, en este caso, la práctica es redondear al número par más próximo, esto es, 20 años.

De manera que, decir que un alumno tiene 20 años no significa afirmar que tiene exactamente 20 años, significa más bien, que realmente su edad está comprendida entre los límites de 19,5 y 20,5 años. 19,5 y 20,5 son los límites reales de 20.

En general, la medida es una variables continua puede darse con cualquier grado de precisión, pero siempre será preferible considerar que dicha medida es el valor medio entre dos límites, llamados límites reales.

### REDONDEO DE DATOS

En la actualidad, con el uso d las computadoras, se pueden obtener miles de cifras decimales o entras; pero en Estadística no se requiere de la precisión absoluta, sino más bien de la aproximación o redondeo de ciertos valores.

Para la aproximación o redondeo se utilizan los siguientes sistemas:

- SISTEMA CONVENCIONAL.- Según el cual:
  - Si el último dígito es menor que 5 se lo suprime

#### Ejemplos:

Cantidad dada	Cantidad redondeada	
9. <u>53</u>	9.5	Redondeado a la décima
7.98 <u>2</u>	7.98	Redondeado a la centésima
19. <u>4</u>	19	Redondeado a la unidad
5,6 <u>30</u>	4,600	Redondeado a centenas

- Si el último dígito es mayor o igual que 5, se lo suprime y el dígito anterior es redondeado a la cifra inmediata superior

Cantidad dada	Cantidad redondeada	
4. <u>47</u>	4.5	Redondeado a la décima
567.96 <u>8</u>	567.97	Redondeado a la centésima
19. <u>6</u>	20	Redondeado a la unidad
98,4 <u>70</u>	984,500	Redondeado a centenas

- SISTEMA INTERNACIONAL

#### Ejemplos:

- Si la fracción decimal es menor que 5 se la deja en la misma cifra, o no se la toma en cuenta para ser retenida

Cantidad dada	Cantidad redondeada	
75,84 <u>35</u>	75,84	Redondeado a dos cifras decimales

- Si la fracción decimal es mayor que 5 se aumenta en una unidad la primera cifra retenida

Cantidad dada	Cantidad redondeada	
324, <u>754</u>	325	Redondeado a cifras enteras

- Si la fracción decimal es exactamente 5 y si le precede al 5 una cifra impar se aumenta una unidad más

Cantidad dada	Cantidad redondeada	
19, <u>5</u>	20	Redondeado a cifras enteras

- Si la fracción decimal es exactamente 5 y si le precede al 5 una cifra par no varia el número

Cantidad dada	Cantidad redondeada	
94, <u>5</u>	94	Redondeado a cifras enteras

## NOTACIÓN SIGMA

**1.3. Sigma.-** La letra griega  $\Sigma$  se utiliza para denotar abreviadamente una suma de términos, todos del mismo tipo.

Ejemplo, si los valores de la variable X son 2, 4, 5 y 9, entonces:

$$\Sigma x = 2 + 4 + 5 + 9$$

Si sólo se van a sumar algunos de los valores, se utilizan subíndices para indicar dichos valores; de este modo,  $\Sigma x_i$ , indica la suma de los valores de la variable X, empezando con el primero ( $i = 1$ ) y terminando con el quinto ( $i = 5$ ):

$$\Sigma X = x_1 + x_2 + x_3 + x_4 + x_5$$

Significa que n observaciones (todas) han de ser sumadas, y a menudo esto se abrevia con los símbolos  $\Sigma X$ .

La expresión  $\sum_{i=1}^n x_i$

### Ejemplos:

Suponiendo que la variables X asume sucesivamente los valores 8, 2, 3, 6, 7, 8, 9, 4, 5, 4, 1, calcular:

$$a) \quad \sum_{i=1}^4 x_i = 8 + 2 + 3 + 6 = 19$$

$$b) \quad \sum_{i=7}^n x_i = 9 + 4 + 5 + 4 + 1 = 23$$

$$c) \quad \sum_{i=1}^n x_i = 8 + 2 + 3 + 6 + 7 + 8 + 9 + 4 + 5 + 4 + 1 = 57$$

### PROPIEDADES:

1. Cuando cada valor de una variable va a ser multiplicado por una constante o dividido entre ella, dicha constante se puede aplicar después de que los valores de la variable se hayan sumado:

$$\Sigma cX = c\Sigma X$$

Supongamos que c es una constante cuyo valor es 2 y X una variable que puede asumir sucesivamente los valores 12, 4, 7 y 10, calcular:

$$\begin{aligned} \Sigma cX &= c\Sigma X \\ &= 2(12 + 4 + 7 + 10) \\ &= 2(33) \\ &= 66 \end{aligned}$$

2. La sumatoria de una constante es igual al producto de la constante y el número de veces que se presenta.

$$\sum_{i=1}^n ci = nc$$

Supongamos que una constante es igual a 5, calcular:

$$\begin{aligned} \sum_{i=1}^6 5i &= 5+5+5+5+5+5 \\ &= 6(5) \\ &= 30 \end{aligned}$$

3. La sumatoria de una suma o diferencia de dos variables es igual a la suma o diferencia, respectivamente, de las sumatorias individuales de las variables.

$$\begin{aligned} \Sigma(x + y) &= \Sigma x + \Sigma y \\ \Sigma(x - y) &= \Sigma x - \Sigma y \end{aligned}$$

Supongamos que X y Y son dos variables que pueden asumir sucesivamente los valores X: 8, 3, 4, 5; Y: 5, 2, 0, 4, calcular  $\Sigma(x + y)$ , y  $\Sigma(x - y)$

Por comodidad se escriben los valores de X y Y en columna e igualmente los valores  $(X + Y)$  y  $(X - Y)$ , de la siguiente manera:

<b>X</b>	<b>Y</b>	<b>(X+Y)</b>	<b>(X-Y)</b>
8	5	13	3
3	2	5	1
4	0	4	4
5	4	9	1
$\Sigma x = 20$	$\Sigma y = 11$	$\Sigma(x + y) = 31$	$\Sigma(x - y) = 9$

$$\begin{aligned} \Sigma(x + y) &= \Sigma x + \Sigma y \\ &= 20 + 11 \\ &= 31 \end{aligned}$$

$$\begin{aligned} \Sigma(x - y) &= \Sigma x - \Sigma y \\ &= 20 - 11 \\ &= 9 \end{aligned}$$

4. La sumatoria de una variable más o menos una constante es igual a la suma de los valores de la variable más o menos, respectivamente, n veces esa constante.

$$\begin{aligned} \Sigma(x + c) &= \Sigma x + n.c \\ \Sigma(x - c) &= \Sigma x - n.c \end{aligned}$$

Supongamos que tres valores de una variable son 2, 4 y 6, calcular:  $\Sigma(x + c)$

$$\begin{aligned} \Sigma(x - c) &= \Sigma x - n.c \\ &= 2 + 4 + 6 - 3c \\ &= 12 - 3c \end{aligned}$$



## ANÁLISIS DE PEQUEÑOS CONJUNTOS DE DATOS

Los métodos estadísticos comprenden el análisis e interpretación de números que provienen de la observación de diversidad de variables, tales como calificaciones de exámenes, porcentajes de respuestas correctas, años de servicio, resistencia de materiales, etc.

Dicho análisis suele realizarse de diversas maneras, dependiendo de si existe una cantidad pequeña de datos (30 ó menos) o si la cantidad de datos que se debe analizar es muy grande.

En el primer caso utilizaremos los métodos que presentamos en este capítulo y, en el segundo caso utilizaremos el agrupamiento de los datos, antes del análisis.

El conjunto de datos, cualquiera que este sea, se puede reducir a unas cuantas medidas numéricas sencillas que resumen el conjunto total. Tales medidas ponen de manifiesto dos características importantes del conjunto: El valor central o típico del conjunto; y, la dispersión de los elementos del conjunto.

### TABLAS DE FRECUENCIA

1. 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.
2. 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.

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

3. Para entender como funcionan las tablas de frecuencia, analicemos el siguiente ejemplo:
4. 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
5. 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.

6. 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

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

### 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.

#### *Ejemplo*

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 número 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 ( $X_{\min}$ ) y el valor mayor ( $X_{\max}$ ), 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 sería igual a 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:

- 1: Contar las veces que se repite cada valor dentro de la muestra.
- 2: Ubicar estas frecuencias en una tabla ordenada.

Grado de Aceptación	Frecuencia (f)
1	0
2	1
3	2
4	3
5	4
<b>TOTAL</b>	<b>10</b>

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:

Donde  $N_c$  representa el número de posibles resultados tabulados en la tabla (que de ahora en adelante se le conocerán como “clases”).

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.

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:
- Grado de aceptación 2:
- Grado de aceptación 3:
- Grado de aceptación 4:
- Grado de aceptación 5:

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:

Para el ejemplo, las frecuencias relativas son:

Grado de aceptación 1:

Grado de aceptación 2: ó 10%

Grado de aceptación 3: ó 20%

Grado de aceptación 4: ó 30%

Grado de aceptación 5: 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).

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.

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	Fa	p	pa
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
<b>TOTAL</b>	<b>10</b>		<b>1,0</b>	

### 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.

## 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):

<i>Azul</i>	<i>Verde</i>	<i>Rojo</i>	<i>Blanco</i>	<i>Rojo</i>	<i>Verde</i>	<i>Rojo</i>
<i>Rojo</i>	<i>Azul</i>	<i>Azul</i>	<i>Blanco</i>	<i>Verde</i>	<i>Azul</i>	<i>Rojo</i>
<i>Verde</i>	<i>Azul</i>	<i>Azul</i>	<i>Rojo</i>	<i>Verde</i>	<i>Azul</i>	<i>Rojo</i>
<i>Verde</i>	<i>Verde</i>	<i>Azul</i>	<i>Rojo</i>	<i>Negro</i>	<i>Negro</i>	<i>Negro</i>
<i>Azul</i>	<i>Verde</i>	<i>Verde</i>	<i>Verde</i>	<i>Blanco</i>	<i>Azul</i>	<i>Rojo</i>
<i>Negro</i>	<i>Negro</i>	<i>Azul</i>	<i>Negro</i>	<i>Blanco</i>	<i>Rojo</i>	<i>Azul</i>
<i>Blanco</i>	<i>Azul</i>	<i>Negro</i>	<i>Azul</i>	<i>Azul</i>	<i>Blanco</i>	<i>Azul</i>
<i>Rojo</i>	<i>Blanco</i>	<i>Rojo</i>	<i>Blanco</i>	<i>Rojo</i>	<i>Blanco</i>	<i>Verde</i>
<i>Blanco</i>	<i>Azul</i>	<i>Verde</i>	<i>Azul</i>	<i>Rojo</i>	<i>Verde</i>	<i>Negro</i>
<i>Blanco</i>	<i>Rojo</i>	<i>Azul</i>	<i>Azul</i>	<i>Rojo</i>	<i>Azul</i>	<i>Blanco</i>

En Excel debería verse así:

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:

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):

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 forma 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:

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 Trigonómicas

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

En la celda C19, ubiquemos la función SUMA de la forma explicada en la función anterior o pulsando sobre el botón . 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:

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

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

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):

## CONSTRUCCIÓN Y CARACTERÍSTICAS DE LAS TABLAS TIPO B

Supongamos que 50 estudiantes han sido examinados de Química y que sus calificaciones individuales son:

60	33	85	52	65	77	84	65	57	74
71	81	35	50	35	64	74	47	68	54
80	41	61	91	55	73	59	53	45	77
41	78	55	48	69	85	67	39	76	60
94	66	98	66	73	42	65	94	89	88

Al presentar las calificaciones podríamos ordenarlas en forma creciente o decreciente, indicando el número de alumnos que obtuvo determinada puntuación. O bien, podríamos presentarlas agrupando las calificaciones para decir, por ejemplo, cuántos alumnos alcanzaron calificaciones entre los 30 y 39 puntos, cuántos entre 40 y 49, etc., es decir, indicando el número de veces que se repiten las calificaciones dentro de cada grupo o categoría de 10 puntos, tal repetición se conoce como “frecuencia de clase”.

Grupo	Calificaciones	Frecuencia
1ro.	30 - 39	4
2do.	40 - 49	6
3ro.	50 - 59	8
4to.	60 - 69	12
5to.	70 - 79	9
6to.	80 - 89	7
7mo.	90 - 99	4
Total		n = 50

El procedimiento se conoce como tabulación y la presentación como “tabla de frecuencias” o “distribución de frecuencias”.

Los grupos o categorías de calificaciones: 30-39, 40-49, etc., se llaman “intervalos de clase”; en tanto que,

Los valores: 30, 40, 50, etc., se llaman “límites inferiores”, y,

Los valores: 39, 49, 59, etc., son los “límites superiores” de los intervalos de clase.

La finalidad que se persigue al presentar las calificaciones agrupadas en una distribución de frecuencias no es únicamente la de simplificar su presentación, sino, sobre todo, la posibilidad de poder obtener mediante técnicas estadísticas varias características de la variable que pueden ser de mucho interés.

Al confeccionar una distribución de frecuencias y a fin de no distorsionar tales características, conviene tener en cuenta que:

- Pocos intervalos de clase facilitan la presentación y reducen los cálculos, pero se pierde mucha información sobre los datos originales;
- Demasiados intervalos de clase proporcionan mucha información sobre los datos originales, pero se pierde el sentido de simplificación y de dificultan poscálculos posteriores;

- En general, es conveniente tener de 10 a 15 intervalos, aunque no hay normas fijas.

El número de calificaciones diferentes que se pueden escribir entre los límites de cada intervalo de clase, incluyendo tales límites, se llama “ancho de clase” o también “amplitud”, o “longitud del intervalo”. Así, el intervalo 30-39 es un intervalo de 10 unidades: 30, 31, 32, 33, 34, 35, 36, 37, 38, 39.

## ¿CÓMO CONSTRUIR UNA DISTRIBUCIÓN DE FRECUENCIAS?

Sigamos considerando el ejemplo de las calificaciones de Química.

1. El primer paso consiste en determinar el recorrido de la variable, llamado también amplitud total, cuyo valor es la diferencia entre las calificaciones: máxima y mínima, aumentada en 1

$$\begin{aligned}\text{Recorrido} &= (\text{calificación máxima} - \text{calificación mínima}) + 1 \\ &= (98 - 33) + 1 \\ &= 66 \text{ puntos}\end{aligned}$$

2. El siguiente paso consiste en determinar el ancho del intervalo de clase ( $i$ ). En la práctica, es frecuente encontrar intervalos de clase de 3, 5 y 7 unidades, aunque también se utiliza el 10 por la facilidad que ofrece al establecer los límites de cada intervalo.
3. La longitud del intervalo debe ir aparejada con el número de intervalos que se desea construir (un criterio, comúnmente aceptado, es considerar de 10 a 15 intervalos de clase). Si dividimos el recorrido de la variable para la longitud del intervalo, obtendremos, aproximadamente, el número de intervalos de la distribución.

$$\text{Número de intervalos} = \frac{\text{Recorrido}}{i}$$

Considerando un ancho de clase de 5 unidades, el número aproximado de intervalos será:

$$\frac{66}{5} = 13,2 = 14 \text{ aprox.}$$

El paso siguiente consiste en señalar los límites del primer intervalo de clase.

Un método corriente es empezar el primer intervalo en un número que sea múltiplo de la amplitud ( $i$ ), pero igual o ligeramente menor que la mínima calificación (33 en nuestro ejemplo).

Después, se escriben los demás intervalos de clase aumentando cada límite en el número de unidades del ancho de clase.

El siguiente paso es el conteo de datos; esto es, se consideran las puntuaciones una por una y se van anotando a la derecha del intervalo de clase correspondiente. Esta columna se sustituye, más tarde, por la columna encabezada por  $f$ ; luego se suman estos valores y el resultado debe ser el número total de datos ( $n$ ).

Observando los pasos anteriores, la distribución de frecuencias correspondientes a las calificaciones de Química, es la siguiente:



<b>INTERVALOS</b>	<b>Nº DE ALUMNOS</b>
30-34	1
35-39	3
40-44	3
45-49	3
50-54	4
55-59	4
60-64	4
65-69	8
70-74	5
75-79	4
80-84	3
85-89	4
90-94	3
95-99	1
	n = 50

Los datos consignados en una distribución de frecuencias se llaman datos agrupados, en tanto que, los datos iniciales se llaman datos no agrupados, datos crudos o puntuaciones brutas.

### **TABLA DE FRECUENCIAS ACUMULADAS**

A veces, es deseable presentar frecuencias de una manera acumulada, especialmente cuando buscamos localizar la posición de un caso en relación con la actuación total de un grupo.

Las frecuencias acumuladas se definen como el número total de casos que tengan cualquier puntaje IGUAL o MENOR que un puntaje dado.

La frecuencia acumulada correspondiente a un intervalo de clase se obtiene al agregar la frecuencia propia de dicho intervalo a las frecuencias de todos los intervalos que están antes de él.

En la siguiente tabla se presenta una distribución de frecuencias absolutas acumuladas de las puntuaciones de 376 estudiantes en un test de habilidad:

<b>INTERVALOS</b>	<b>f</b>	<b>fa</b>
5-9	4	4
10-14	15	19
15-19	23	42
20-24	42	84
25-29	58	142
30-34	64	206
35-39	58	264
40-44	46	310
45-49	32	342
50-54*	20*	362
55-59	12	374
60-64	2	376
	n = 376	

La frecuencia en una distribución de frecuencias indica, el número de casos cuya calificación es IGUAL o MENOR que el límite superior del intervalo al que pertenece dicha frecuencia acumulada. Así: 20\* alumnos del ejemplo anterior alcanzaron una calificación de 54\* puntos o menos.

La representación gráfica de la distribución de frecuencias acumuladas, igual que ciertos cálculos que más adelante mostraremos, permitirá decir qué número de alumnos obtienen una calificación igual o menor que cualquier calificación comprendida dentro de la distribución.

### TABLA DE FRECUENCIAS RELATIVAS

Cuando el investigador estudia distribuciones de igual tamaño total, los datos de frecuencia (f) pueden utilizarse para hacer comparaciones entre los grupos. Sin embargo, generalmente no es posible estudiar distribuciones que tengan exactamente el mismo número de casos; en tales circunstancias, necesitamos un método para estandarizar distribuciones de frecuencia por tamaño. Dos de los métodos más populares y útiles para estandarizar por tamaño y comparar distribuciones son la proporción y el porcentaje.

La proporción compara el número de casos en una categoría dada por el tamaño total de la distribución. Para convertir cualquier frecuencia f en una proporción p, dividimos el valor de f por el número total de casos de la distribución (n):

$$p = \frac{f}{n}$$

El porcentaje compara la frecuencia de ocurrencia en una categoría por cada 100 casos. Para calcular el porcentaje (p) correspondiente a una frecuencia dada f, basta multiplicar la proporción por 100:

$$p = \frac{f}{n} (100)$$

La siguiente es la tabla de frecuencias relativas que corresponde a los datos de la distribución anterior:

INTERVALOS	f	p	P
5-9	4	0.011	1.1
10-14	15	0.040	4.0
15-19	23	0.061	6.1
20-24	42	0.112	11.2
25-29	58	0.154	15.4
30-34	64	0.170	17.0
35-39	58	0.154	15.4
40-44	46	0.122	12.2
45-49	32	0.085	8.5
50-54	20	0.053	5.3
55-59	12	0.032	3.2
60-64	2	0.005	0.5
	n = 376	0.999	99.9%

Los cálculos de la tabla que antecede se han realizado de la siguiente manera.

**Ejemplo:** la proporción  $p$  que corresponde a la frecuencia 32 del intervalo de clase 45-49:

$$\begin{aligned} p &= \frac{f}{n} \\ &= \frac{32}{376} \\ &= 0,085 \end{aligned}$$

El porcentaje  $P$  que corresponde a la frecuencia del mismo intervalo:

$$\begin{aligned} p &= \frac{f}{n} (100) \\ &= 0,085 (100) \\ &= 8,5\% \end{aligned}$$

NOTA.- Cuando el número total de casos es pequeño, los tantos por ciento son inestables. Es decir, una variación en un caso puede producir un cambio relativamente grande en el tanto por ciento. Cuando el número de casos es menor a 100, debería evitarse el empleo de los tantos por ciento.

### TABLA DE FRECUENCIAS RELATIVAS ACUMULADAS

Además de la frecuencia acumulada, también podemos construir una distribución que indique porcentajes acumulados ( $Pa$ ).

Es decir, una columna de la distribución que indique el porcentaje de casos cuya calificación es IGUAL o MENOR QUE un puntaje determinado.

La siguiente tabla muestra las frecuencias relativas acumuladas correspondientes a la distribución anterior:

INTERVALOS	f	fa	Pa
5-9	4	4	1.1
10-14	15	19	5.0
15-19	23	42	11.2
20-24	42	84	22.3
25-29	58	142	37.7
30-34	64	206	54.8
35-39	58	264	70.2
40-44	46	310	82.4
45-49	32	342	90.7
50-54	20	362	96.3
55-59	12	374	99.5
60-64	2	376	100.0
	n = 376		

En este caso, el cálculo del  $Pa$  correspondiente al intervalo de clase 45-49, por ejemplo, se obtiene dividiendo la  $fa$  342 entre  $n = 376$  y multiplicando por 100.

Es decir, expresando mediante fórmula:

$$Pa = \frac{fa}{n} (100)$$

En dicho intervalo se puede observar que el 90,7% de los alumnos alcanzó una calificación IGUAL o MENOR que el límite superior 49.

### Ejemplo 2 (OTRA FORMA)

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

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
<b>Total</b>	<b>30</b>

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

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

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.

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':

**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 cálculo se realiza multiplicando el Ancho ajustado por el Número de Intervalos:

El rango se incremento en cuatro años.

El incremento se le sumará al valor Máximo ( $X_{\max}'$ ) o restará al valor Mínimo ( $X_{\min}'$ ).

En este caso optaremos por aumentar el valor Máximo y reducir el valor Mínimo en dos.

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')

**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.

## EL PROBLEMA DE LA AMBIGÜEDAD EN LAS TABLAS DE FRECUENCIA TIPO B

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

### 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 Límite 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 Límite 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.

### Segunda Solución

Se convierten los Límites 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.

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

Ni	Lm	Ls	F	Fa	P	Pa	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
		<b>Total</b>	<b>30</b>		<b>1,00</b>		

### 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 ( $N_c$ ).

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

**Paso 2:** Determinar el ancho de cada intervalo.

Se determina el rango como primera medida.

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

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

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

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

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

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 4:** Determinar los intervalos de clases iniciales.

**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
		<b>Total</b>	<b>54</b>		<b>1,00</b>		

### 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.

### 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 (dígítelos 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 creamos una tabla resumen a partir de la celda B10, que empiece por la cantidad de datos (n) y el número de intervalos

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:

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 Trigonómicas

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 ( $X_{min}$ ), valor máximo ( $X_{max}$ ), 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 (número 1; número 2;...)  
**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:

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).

**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 Trigonómicas

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:

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 ( $X_{min}'$ ), siendo su límite superior 13 ( $X_{min}'$  más el ancho de clase).

El límite inferior de las siguientes clases es igual al límite superior de su clase anterior:  
 Podremos arrastrar esta fórmula hasta el último intervalo de clase (C28):

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:

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

**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:**

**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:**

**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)**

### EJERCICIOS PROPUESTOS

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

- 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

- Elabore una tabla de frecuencia que permita resumir los datos
  - ¿Que porcentaje de personas valoró la Gestión del Departamento como Buena?
  - ¿Cuántas personas valoraron la gestión como Excelente y Buena?
  - Interprete  $f_3$ ,  $F_3$ ,  $P_3$  y  $Pa_3$ .
- 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

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

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  $Fa_2$ ,  $Fa_4 - Fa_2$ ,  $Fa_4 - f_3$ ,  $f_3 + f_2$ .

c) Interprete  $P_1 + P_2$ ,  $Pa_3 - Pa_2$ ,  $Pa_3 - h_2$

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

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.

9. A partir de la siguiente tabla de frecuencias con datos parciales:

Nc	Lm	Ls	f	Fa	P	Pa
1	(10	14]		10		
2	(14	18]		15		
3	(18	22]		31		
4	(22	26]		42		
5	(26	30]		55		
		<b>TOTAL</b>				

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

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	Fa	P	Pa
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
		<b>TOTAL</b>	<b>150</b>		<b>1,00</b>	

Reconstruya la tabla de frecuencia.

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

### CUESTIONARIO DE REPASO

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

- Las Tablas de Frecuencia Tipo A se caracterizan por:
  - Trabajan solo con datos cualitativos.
  - Agrupan datos cuyo Rango es bajo.
  - Agrupan datos cuyo Rango es alto.
  - Presentan más intervalos que en las Tablas Tipo B.
- Las frecuencias relativas se diferencian de las frecuencias absolutas porque:
  - Las frecuencias relativas se establecen de acuerdo a una base.
  - Las frecuencias relativas se expresan como porcentaje.
  - La suma de las frecuencias relativas es igual a 1.
  - La A y B.
- En una tabla de frecuencia,  $Fa_4 - Fa_2$  es lo mismo que:
  - $Fa_4 - f_3$
  - $Fa_2 + f_3$
  - $Fa_3$
  - $f_3 + f_4$
- En una tabla de frecuencia,  $P_3 + Pa_2$  es lo mismo que:
  - $Fa_3$
  - $Fa_4$
  - $P_3$
  - $P_3 - P_4$

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



## REPRESENTACIONES GRÁFICAS

Los gráficos estadísticos más usados que parten de resúmenes o tablas de frecuencias.

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

**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

### 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.

#### *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
<b>Total</b>	<b>140</b>

### 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	Frecuencia	P
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
<b>Total</b>	<b>140</b>	<b>1</b>

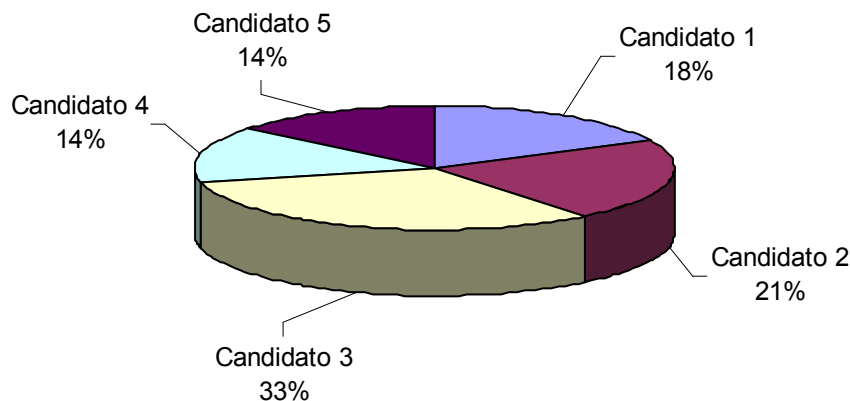
**PASO 2:** Determinar los ángulos que representan las porciones dentro de la figura para cada clase.

Un círculo esta formado por un ángulo de  $360^\circ$ . La porción correspondiente al Candidato 1 equivale a un 17,86% de esos  $360^\circ$ , es decir,  $64,296^\circ$ . 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	Frecuencia	P	Ángulo
Candidato 1	25	0,1786	$64,296^\circ$
Candidato 2	30	0,2142	$77,112^\circ$
Candidato 3	45	0,3214	$115,704^\circ$
Candidato 4	20	0,1429	$51,444^\circ$
Candidato 5	20	0,1429	$51,444^\circ$
<b>Total</b>	<b>140</b>	<b>1</b>	<b><math>360^\circ</math></b>

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



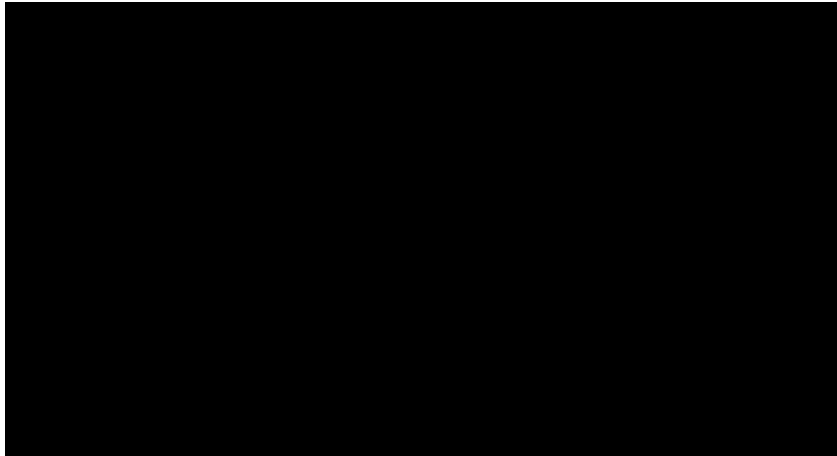
### **Características de los gráficos de sectores**

- 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 ( $360^\circ$ ).

**Ejemplo 2**

Los números de la siguiente tabla representan el número de alumnos del último curso de cinco carreras universitarias.

<b>CARRERAS</b>	<b>f</b>	<b>p</b>	<b>%</b>
Ciencias	440	0.48	48
Derecho	220	0.24	24
Medicina	120	0.13	13
Economía	80	0.09	9
Educación	60	0.06	6
<b>TOTAL</b>	920		100



## CONSTRUCCIÓN DE LOS GRÁFICOS DE SECTORES EN EXCEL

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

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 GRÁFICOS, 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:

## 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.

### *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
<b>Total</b>	<b>88</b>

### SOLUCIÓN

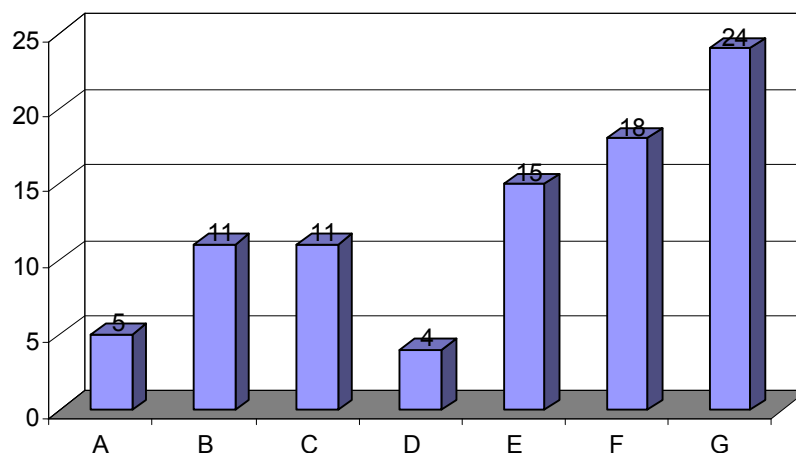
Para crear un gráfico de barras, seguiremos 2 sencillos pasos:

**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.

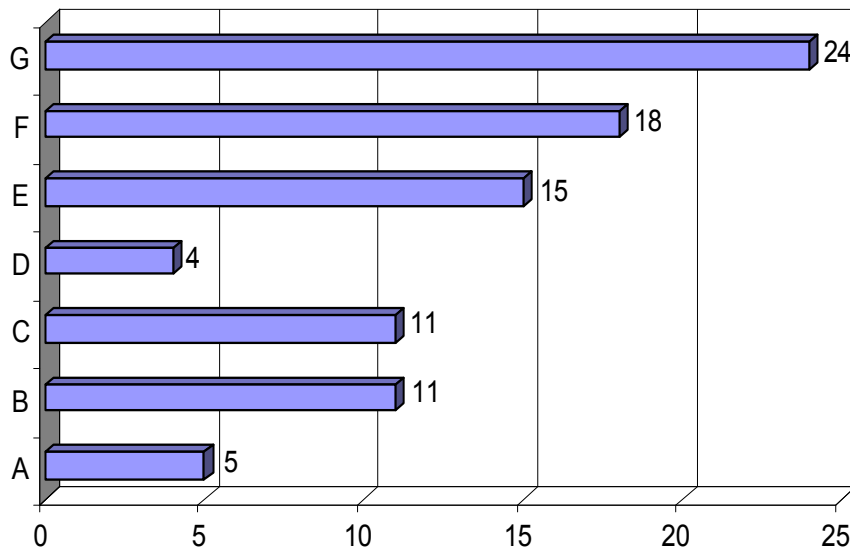
**PASO 2:** A cada clase se representa con una columna (o barra) cuya altura concuerda con su frecuencia expuesta en el eje vertical. Para la clase A con frecuencia 5, tenemos:

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):



### **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.

### **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:*

*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.*

*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:*

## 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.

Es un gráfico de barras verticales cuyas bases sobre el eje horizontal tiene un ancho igual a la longitud del intervalo de clase y cuyas alturas son numéricamente iguales a las frecuencias de clase.

Pasos para dibujar:

1. Se busca el punto medio de cada uno de los intervalos de clase
2. Se trazan líneas verticales en estos valores cuyas alturas son iguales a las frecuencias de cada intervalo, respectivamente:
3. Se toma la mitad de la longitud del intervalo de clase a ambos lados de estas verticales y se dibujan barras, sin dejar espacios entre ellas.

### Ejemplo 1

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

INTERVALOS	F	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
26.1-28,1	8	28.1
<b>TOTAL</b>	<b>92</b>	

### SOLUCIÓN

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

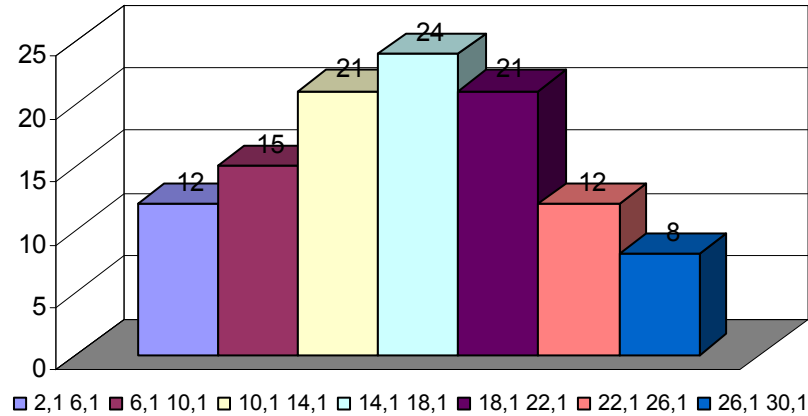
**PASO 1:** Representar las escalas en los ejes horizontal y vertical del primer cuadrante de un plano de cartesiano. En el eje vertical colocaremos las frecuencias y en el eje horizontal las marcas de clases.

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.

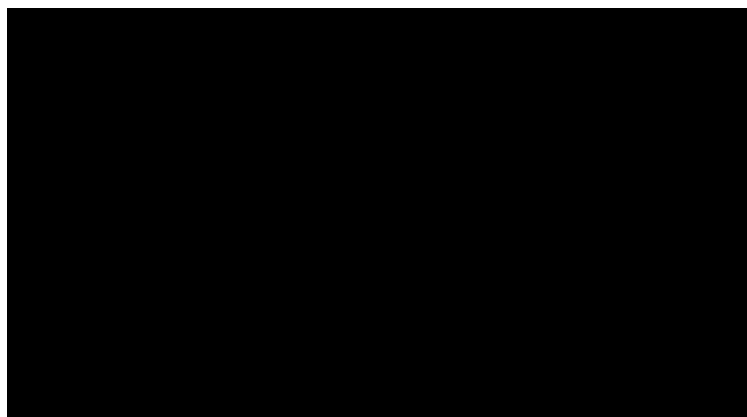
El gráfico final quedaría:



## Ejemplo 2

Trazar el histograma de frecuencias correspondiente al siguiente conjunto de datos:

INTERVALO	f	PUNTO MEDIO
25-29	1	27
30-34	0	32
35-39	3	37
40-44	6	42
45-49	6	47
50-54	6	52
55-59	7	57
60-64	4	62
65-69	4	67
70-74	1	72
75-79	1	77
80-84	1	82
	n = 40	





## OBSERVACIONES

- Si se unieran los puntos medios de las bases superiores de cada rectángulo del histograma, se obtendría el polígono de frecuencias trazado por los mismos datos.
- En un mismo sistema coordenado sólo puede situarse un histograma, esto impide que pueda hacerse comparaciones y explicar diferencias como se hizo con el polígono de frecuencias.
- La altura de cada barra indica la frecuencia de cada intervalo.

## 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.

### Ejemplo 3

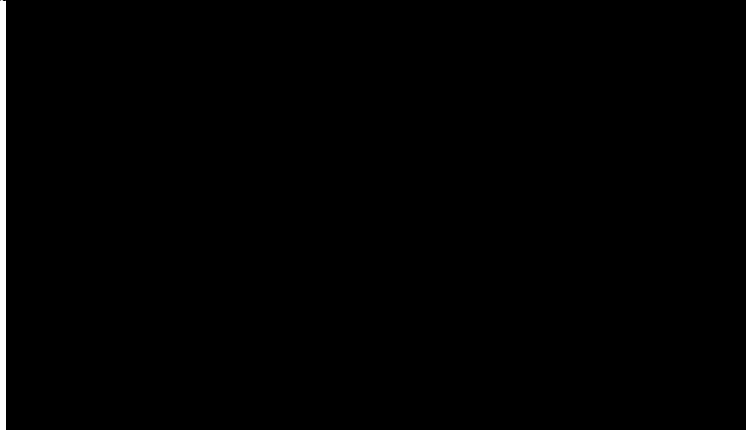
La siguiente tabla contiene información sobre el número de accidentes que ocurren diariamente, durante 50 días, en un enorme estacionamiento:

6 - 9 - 2 - 7 - 0 - 8 - 2 - 5 - 4 - 2 - 6 - 7 - 5 - 4 - 4 - 4 - 4 - 2 - 5 - 6 - 3 - 5 - 6 - 3 - 3 - 8 - 8 - 4 - 4 - 4 - 7 - 7 - 6 - 0 - 6 - 4 - 7 - 5 - 3 - 7 - 1 - 3 - 8 - 5 - 1 - 2 - 3 - 6 - 5 - 0

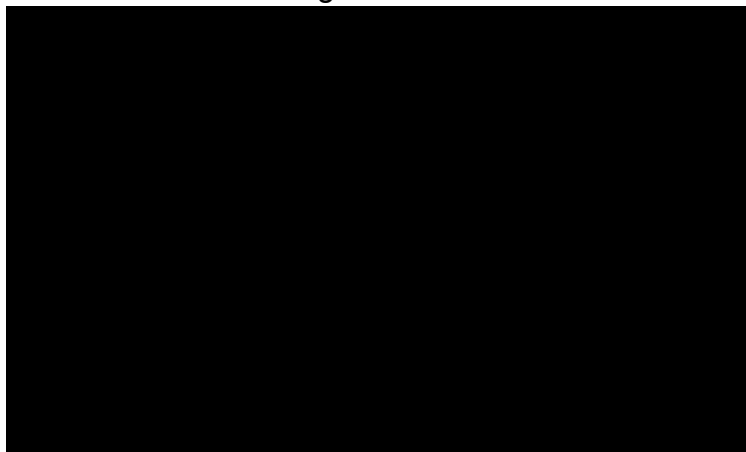
**Utilizando intervalos obtendríamos:**

Representando gráficamente estas dos distribuciones tenemos:

En el primer caso se trazan barras separadas (pueden ser verticales u horizontales)



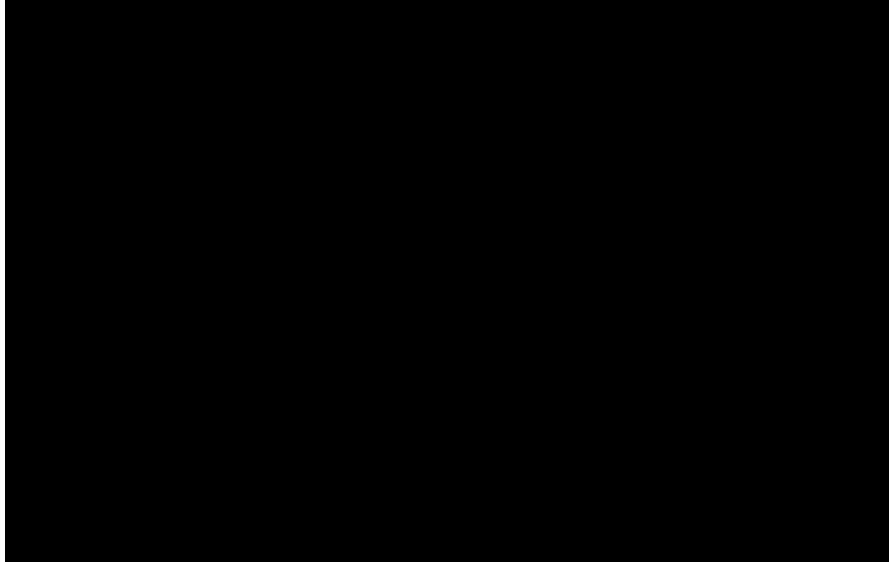
En el segundo caso se traza un histograma



Las distribuciones de datos nominales y jerarquizados se representan generalmente mediante diagramas de barras separadas, lo que significa que las categorías no se tocan.

#### **Ejemplo 4**

Venta de gaseosas en un día



#### **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:*

*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:*

*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:*

## EL POLÍGONO 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.

### **Ejemplo**

El valor usado como la mejor estimación para las puntuaciones que caen en un intervalo de clase se llama marca de clase o punto medio porque está localizado en la mitad del intervalo.

### **MARCA DE CLASE $M_c$**

El punto medio de un intervalo se averigua sumando los límites y dividiendo para dos. Así:

$$\frac{30 + 34}{2} = 32$$

El intervalo que tiene como límites 30-34, tiene como punto medio

Obsérvese que el punto medio será entero si es que el ancho de clase del intervalo es impar y, será decimal si es que "i" es par; de aquí la conveniencia de utilizar preferentemente valores impares como longitudes o anchos de clase (son frecuencias los valores 3, 5, 7 y también 10).

Al construir el polígono de frecuencias se utiliza un sistema coordinado rectangular en el cual se llevan, sobre el eje de las X, los valores de los puntos medios de todos y cada uno de los intervalos de clase y, sobre el eje de las Y, los valores de las frecuencias correspondientes.

Los puntos que resultan se unen mediante segmentos de recta, teniéndose de esta manera una línea poligonal, la misma que, para cerrarla se sienta sobre el eje X, utilizando medio intervalo anterior al primer intervalo de clase y medio intervalo posterior al último.

### **REGLA DE LOS TRES CUARTOS DE ALTURA.**

Para las representaciones gráficas de las distribuciones de frecuencia el eje vertical debe hacerse de tal modo que la altura del punto máximo (resultado que está asociado con la frecuencia más alta) sea aproximadamente igual a tres cuartos de la longitud de eje horizontal.

$$alto = \frac{3}{4} (ancho)$$

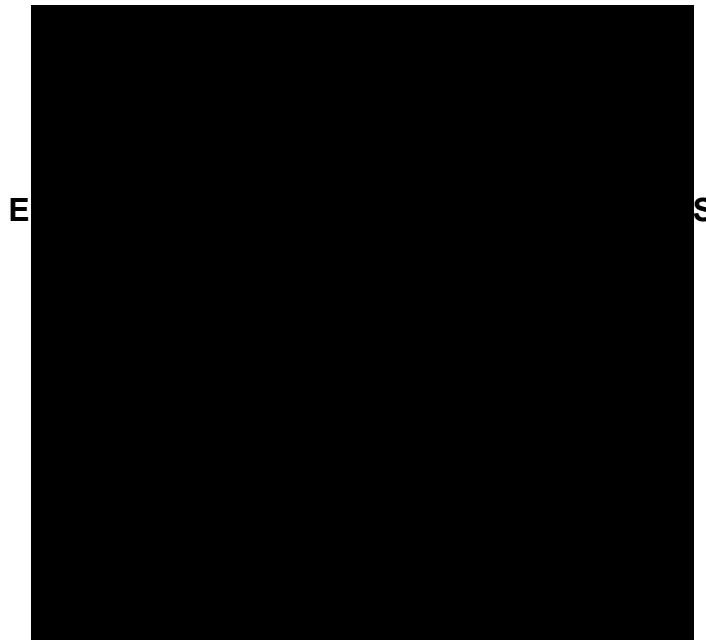
A continuación se muestran los resultados obtenidos por un grupo de estudiantes de cuarto grado en una prueba de habilidad de lectura.

Representar tales resultados mediante un polígono de frecuencias:

INTERVALO	f	PUNTO MEDIO
14-16	2	15
17-19	4	18
20-22	6	21
23-25	9	24
26-28	11	27
29-31	22	30
32-34	35	33
35-37	24	36
38-40	12	39
41-43	8	42
44-46	5	45
47-49	3	48
50-52	2	51
	n = 143	

$$\begin{aligned} & \frac{\text{Max. freq.}}{11} \\ &= \frac{35}{11} \\ &= 3 \text{ aprox.} \end{aligned}$$

Esto es, cada una de las 11 divisiones que deben hacerse sobre el eje Y, valdrá 3f aproximadamente.



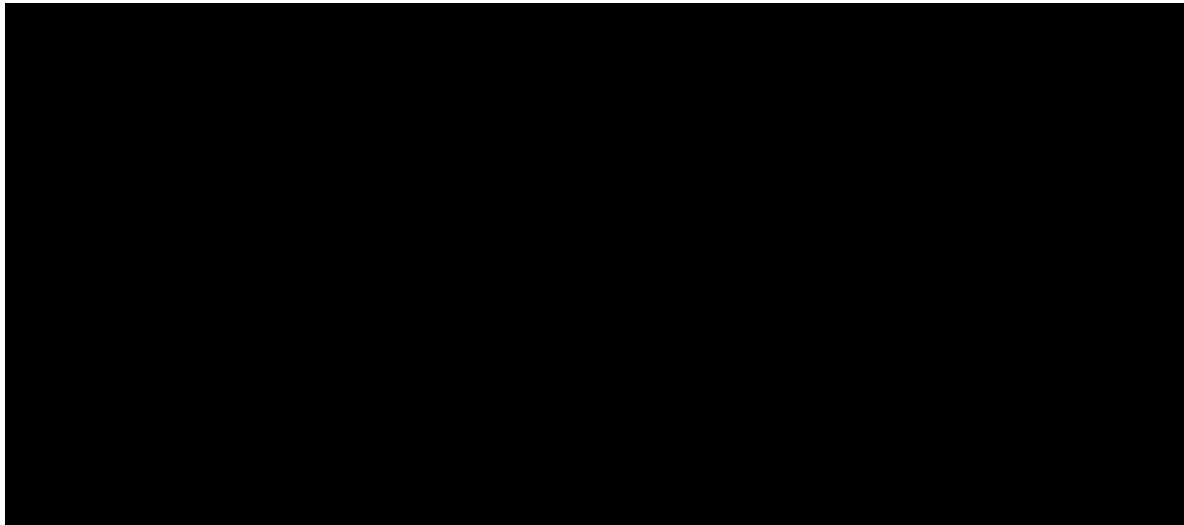
Con mucha frecuencia, los datos de las distribuciones que se desean comparar gráficamente mediante polígonos de frecuencias difieren considerablemente entre sí, a tal punto que se presentan serias dificultades en su representación sobre un mismo eje coordenado; en tales casos, se suelen convertir las frecuencias absolutas en frecuencias relativas, generalmente expresadas como porcentajes.

Trazar el polígono de frecuencias relativas correspondiente a los datos de la tabla.

### PUNTUACIONES DE 376 ESTUDIANTES EN UN TEST DE HABILIDAD

INTERVALO	f	Mc	%
5-9	4	7	1.1
10-14	15	12	4.0
15-19	23	17	6.1
20-24	42	22	11.2
25-29	58	27	15.4
30-34	64	32	17.0
35-39	58	37	15.4
40-44	46	42	12.2
45-49	32	47	8.5
50-54	20	52	5.3
55-59	12	57	3.2
60-64	2	62	0.5
	n = 376		99.9 %

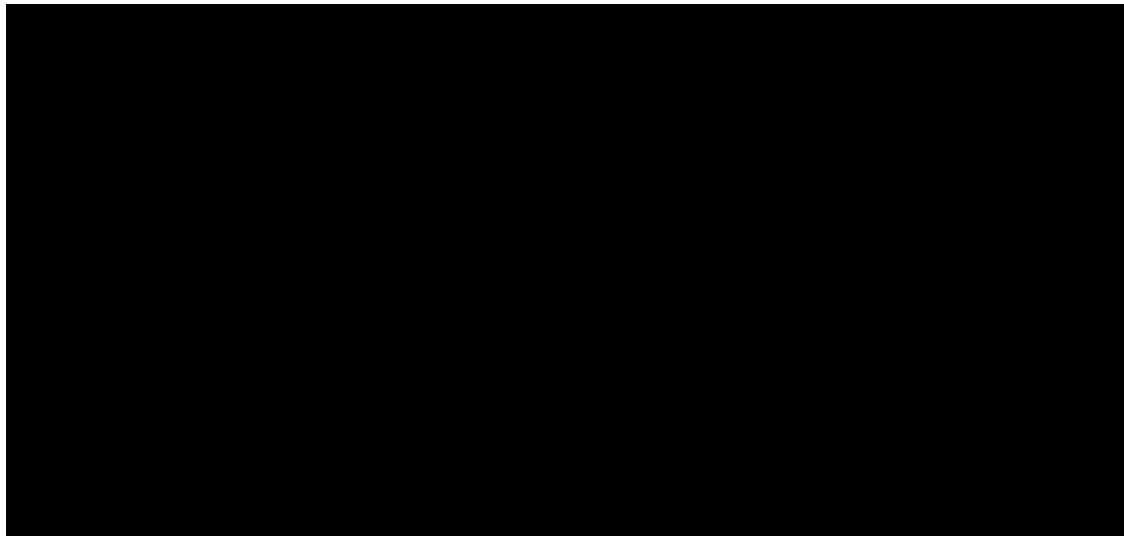
Para trazar el polígono de frecuencias relativas se lleva las puntuaciones en forma usual, esto es, los puntos medios de los intervalos sobre el eje **X** y, las frecuencias relativas (porcentajes) sobre el eje **Y**, como se indica en el siguiente gráfico



Trazar, en un mismo eje coordenado, los polígonos de frecuencias correspondientes a las puntuaciones de un test de inteligencia aplicado a los niños de dos escuelas diferentes:

PUNTUACIONES	ESCUELA A	ESCUELA B	P <sub>1</sub>	P <sub>2</sub>
	F <sub>1</sub>	F <sub>2</sub>		
60-69	1	5	0.6	1
70-79	4	5	2.2	1
80-89	18	10	10.0	2
90-99	32	10	17.8	2
100-109	69	50	38.3	10
110-119	31	180	17.2	36
120-129	12	170	6.7	34
130-139	8	40	4.4	8
140-149	4	20	2.2	4
150-159	1	10	0.6	2
	n = 180	n = 500	100 %	100 %

### DISTRIBUCIÓN DE LAS PUNTUACIONES DE UN TEST DE INTELIGENCIA DE LOS NIÑOS DE DOS ESCUELAS



#### *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 (triángulo rectángulos iguales), manteniendo el área global en el gráfico.

#### **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.

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

## CURVA DE FRECUENCIAS ACUMULADAS

Refiriéndonos a calificaciones de los alumnos, resulta frecuente tener que contestar preguntas tales como:

- ¿Cuántos alumnos tienen calificaciones iguales o menores que...?
- ¿Cuántos alumnos tienen calificaciones superiores a...?

En tales casos conviene ordenar las calificaciones en una tabla de frecuencias acumuladas, según se muestra en el siguiente ejemplo:

Supongamos que 26 alumnos realizan un examen (sobre 10 puntos), obteniendo las siguientes calificaciones:

CALIFICACIONES X	f	Fa
0	0	0
1	2	2
2	4	6
3	4	10
4	3	13
5	2	15
6	5	20
7	8	24
8	0	24
9	2	26
10	0	26
n = 26		

La variable X presenta los valores posibles de las calificaciones de (0 a 10 puntos). La columna f indica el número de estudiantes que tienen una calificación determinada (la calificación correspondiente de la primera columna).

La columna de Fa muestra el número de estudiantes que tienen una calificación “igual o menor que” la calificación correspondiente de la primera columna.

### Ejemplo

15 alumnos tienen una calificación igual o menor que 5 puntos; igualmente podemos afirmar que 11 alumnos (el complemento de 15 a 26) tienen una calificación superior a 5 puntos.

Cuando los datos se han agrupado por intervalos de clase, la tabla de frecuencias acumuladas puede ser representada gráficamente tomando como abscisa los límites reales superiores de los intervalos de clase y como ordenadas longitudes proporcionales a las frecuencias acumuladas.

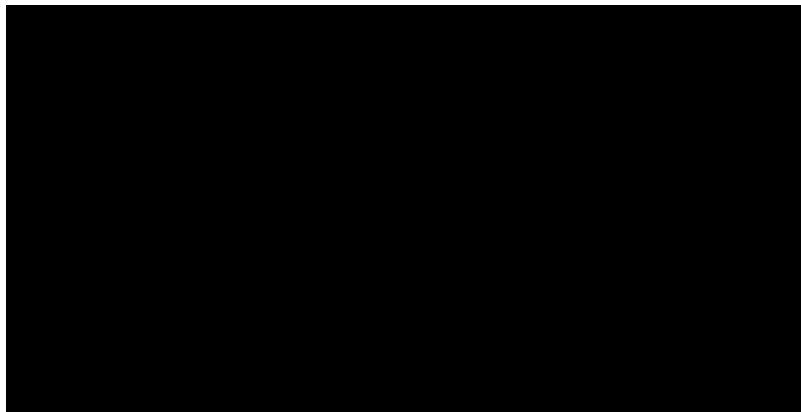
La gráfica de frecuencias acumuladas se llama ojiva

## OJIVA

Representar gráficamente los siguientes valores:

<b>INTERVALOS</b>	<b>f</b>	<b>Fa</b>
90-99	3	3
100-109	14	17
110-119	16	33
120-129	11	44
130-139	4	48
140-149	2	50
	n = 50	

### Distribución de las calificaciones



Esta ojiva indica el número de personas que registran una puntuación “igual o menor que” un cierto valor.

### Ejemplo

48 alumnos obtuvieron un puntaje igual o inferior a 139.5 (observe el gráfico anterior); igualmente, podemos afirmar que 2 alumnos (el complemento de 48 a 50) alcanzaron un puntaje superior a 139.5.

### CURVA DE FRECUENCIAS RELATIVAS ACUMULADAS

La tabla de frecuencias relativas acumuladas correspondiente al ejercicio anterior es la siguiente:

<b>INTERVALOS</b>	<b>f</b>	<b>Fa</b>
90-99	3	3
100-109	14	17
110-119	16	33
120-129	11	44
130-139	4	48
140-149	2	50
	n = 50	

Gráficamente se representa de la siguiente manera los límites reales superiores de cada intervalo de clase se sitúan sobre el eje horizontal (eje x).



Los tantos por ciento se ubican siempre sobre el eje vertical (eje Y). Tales valores estarán comprendidos entre 0 y 100:

- Cada punto del gráfico acumulativo asocia a los intervalos reales superiores de cada intervalo el tanto por ciento acumulado correspondiente
- La línea curva trazada por los puntos anteriores se prolonga un intervalo hasta alcanzar el eje X.

Las curvas correspondientes a las distribuciones acumulativas porcentuales tienen la forma de una S, razón por la cual se denominan curvas de la S

### **LECTURA DE LOS CENTILES A PARTIR DE LA CURVA ACUMULATIVA DE FRECUENCIAS RELATIVAS**

**LOS CENTILES**, también conocidos como percentiles, son puntuaciones de la distribución que dejan por debajo de sí un determinado tanto por ciento de los datos.

#### **Ejemplo:**

El centil 88, es la puntuación de la distribución tal que el 88 % de los datos están por debajo de él.

Si la curva acumulativa de frecuencias relativas se ha trazado con cuidado sobre papel milimetrado, se puede leer en ella el valor de cada centil, con gran precisión.

Con este fin, se traza una perpendicular al eje Y en el punto que corresponde (%) deseado, hasta cortar la curva. A partir de este punto, se traza la perpendicular al eje X.

El pie de esta perpendicular señala el valor correspondiente al centil deseado.

#### **Ejemplo:**

Utilizando el gráfico anterior, determinar los puntajes que corresponden a los centiles:

6, 35 y 66

C6 = 99.5 puntos

C35 = 109.5 puntos

C66 = 119.5 puntos

## MEDIDAS DE TENDENCIA CENTRAL

Las medidas que se utilizan para indicar un valor que tiende a ser el más representativo del conjunto son:

- La media aritmética
- La mediana; y,
- La moda.

### 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  $\mu$ ;
- ☞ En el caso de que estemos trabajando con una muestra, el símbolo será

**Media aritmética ( $\mu$  o  $\bar{x}$ ):** 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.

### MEDIA ARITMÉTICA PARA DATOS NO AGRUPADOS

#### Ejemplo

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:

3,23,12,44,03,5

3,03,53,84,24,0

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

#### SOLUCIÓN

Aplicando la fórmula para datos no agrupados tenemos:

	2,4
	3
	3,1
	3,2
	3,5
	3,5
	3,8
	4
	4
	4,2
SUMA	34,7

$$\bar{X} = \frac{\sum X}{n}$$

$$\bar{X} = \frac{34,7}{10}$$

$$\bar{X} = 3,47$$

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.

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.

**La media aritmética de un conjunto de datos no agrupados** se calcula sumando los valores del conjunto y dividiendo esta suma para el número de dichos datos.

Utilizando la notación sigma, la media aritmética se expresa de la siguiente manera:

$\bar{X}$  = media aritmética

$\Sigma$  = "suma de"

X = los datos del conjunto

n = el número total de datos del conjunto

$$\bar{X} = \frac{\sum X}{n}$$

Sin embargo, esta fórmula sirve para obtener la media aritmética de un pequeño conjunto de datos.

## Ejemplo 2

La media aritmética de las calificaciones de estadística 9, 8 y 10 es:

$$\bar{X} = \frac{\sum x}{n}$$

$$\bar{X} = \frac{9 + 8 + 10}{3}$$

$$\bar{X} = \frac{27}{3}$$

$$\bar{X} = 9$$

**Para un mayor número de datos**, cuando uno o varios de tales datos se repiten más de una vez, resulta más práctico calcular la media aritmética utilizando la siguiente fórmula:

$$\bar{X} = \frac{\sum fX}{n}$$

## Ejemplo:

Calcular la media aritmética del siguiente conjunto de datos:

8	-	2	-	3	-	4	-	7	-	2	-	5
8	-	6	-	7	-	6	-	5	-	5	-	4
6	-	5	-	4	-	3	-	3	-	2	-	3
6	-	5	-	4	-	5	-	7	-	6	-	4

La siguiente tabla ilustra los cálculos necesarios:

<b>X</b>	<b>f</b>	<b>f.X</b>
2	3	6
3	4	12
4	5	20
5	6	30
6	5	30
7	3	21
8	2	16
N = 28		$\sum fX = 135$

$$\bar{X} = \frac{\sum fX}{n}$$

$$\bar{X} = \frac{135}{28}$$

$$\bar{X} = 4,82$$

La fórmula para calcular la media supone que cada uno de los datos es de igual importancia. En términos generales, esto suele ser así, no obstante, hay excepciones:

### Ejemplo:

Un profesor indica que recibirá dos exámenes mensuales, cada uno equivaldrá el 30% de la calificación total; y, un examen final que corresponderá al 40% restante. Si un alumno obtiene 80 en el primer examen, 90 en el segundo y 96 en el final, ¿cuál es su promedio?

<b>Examen</b>	<b>Calificación X</b>	<b>Ponderación W</b>	<b>WX</b>
1er.	80	30	2.400
2do.	90	30	2.700
3ro.	96	40	3.840
		$\sum W = 100$	$\sum WX = 8940$

$$\bar{X} = \frac{\sum WX}{\sum W}$$

$$\bar{X} = \frac{8.940}{100}$$

$$\bar{X} = 89,40 \text{ puntos}$$

### MEDIA DE VARIAS MEDIAS

A veces no se dispone de los datos originales de un conjunto de datos y, más bien, se conocen las medias aritméticas de dos o más grupos en los que se ha dividido dicho conjunto y, se desea hallar la media de todo el conjunto, como si se tratara de un solo grupo. Esto se puede calcular suponiendo que las medias de cada grupo son los datos de un nuevo conjunto y el número de elementos de cada grupo constituyen los pesos o ponderaciones con los cuales se obtiene una media ponderada, llamada en este caso media de varias medias.

**Ejemplo:**

Se desea obtener la media aritmética de las edades de los 50 alumnos de un curso determinado.

Para realizar el trabajo en equipo, se divide el curso en tres grupos diferentes y se encomienda a tres compañeros del mismo curso obtener los datos y calcular la media aritmética de cada grupo; luego se reciben los resultados y se realiza el cálculo final, de la siguiente manera:

<b>Gru- po</b>	<b>Datos media ( )</b>	<b>Número de alumnos (n)</b>	<b>n.X</b>
A	18,7	12	224,4
B	20,1	22	442,2
C	19,5	16	312,0
		$\Sigma n = 50$	$\Sigma nX = 978,6$

$$\bar{x} = \frac{\Sigma nX}{\Sigma n}$$

$$\bar{x} = \frac{978,6}{50}$$

$$\bar{x} = 19,57 \text{ años}$$

Esta medida, basada en las medias de cada grupo, recibe el nombre de media de medias y es equivalente a la media aritmética calculada mediante la fórmula, siempre que se conocieran todas y cada una de las edades de los alumnos integrantes del curso en referencia.

**PROPIEDADES DE LA MEDIA.-** La media aritmética presenta ciertas propiedades que la convierten en la medida de tendencia central que se utiliza con más frecuencia:

1. La media es sensible a cada valor del conjunto de datos. Así, si cambia algún valor del conjunto, la media de tal conjunto también cambiará.
2. Si se suma una constante a cada valor del conjunto de datos, la media aritmética aumentará su valor en dicha constante. De manera que, si sumamos 10 a cada valor del conjunto, la media aumentará en 10. En forma similar, al restar a cada valor del conjunto de datos una constante, o bien, multiplicar o dividir cada valor por una constante, eso hará que la media aritmética disminuya en la misma cantidad o resulte multiplicada o dividida por dicha constante.
3. La suma de las desviaciones de los datos de un conjunto a partir de su media aritmética, siempre es cero.

Se entiende por desviación (d) de cada uno de los datos de un conjunto desde la media aritmética, la diferencia entre cada uno de los datos y la media:

$$d = X - \bar{X}$$

Por considerar que la propiedad anterior es muy importante, comprobar que se cumple para el siguiente conjunto de datos: 10, 12, 14, 15 y 17.

<b>X</b>	
10	- 3.6
12	- 1.6 <u>(- 5.2)</u>
14	0.4
15	1.4
17	3.4 (5.2)
68	$\Sigma d = 0$

Así pues, la suma de las desviaciones es cero.

$$\bar{X} = \frac{\Sigma x}{n}$$

$$\bar{X} = \frac{68}{5}$$

$$\bar{X} = 13.6$$

## LA MEDIANA

Es el punto que deja igual número de datos a cada lado de él, siempre que la distribución esté ordenada.

Es decir que, cuando los datos estén desordenados, antes de calcular la mediana se deberán ordenar de menor a mayor.

En el cálculo de la mediana se pueden presentar dos casos:

1. Cuando el número de datos es impar, la mediana es el valor que se encuentra en el centro de la serie ordenada de datos. La posición de la mediana dentro de la serie ordenada puede hallarse mediante la fórmula:

$$\text{Posición} = \frac{n + 1}{2}$$

### **Ejemplo:**

Los alumnos de 6to grado de una Escuela son 11 y han obtenido los siguientes puntajes en una prueba de Botánica: 14, 12, 10, 8, 18, 17, 8, 19, 19, 8 y 9 hallar la mediana de dichas calificaciones.

Ordenando los puntajes: 8, 8, 8, 9, 10, 12, 14, 17, 18, 19, 19 Tomando en cuenta que los puntajes son 11, determinamos la posición de la mediana:

$$\text{Posición} = \frac{n + 1}{2}$$

$$\text{Posición} = \frac{11 + 1}{2}$$

$$\text{Posición} = 6\text{to. lugar}$$

El 6to lugar en la ordenación lo ocupa el puntaje 12. Así pues, la mediana de las calificaciones es 12.

2. Cuando el número de datos es par, la mediana es igual a la semisuma de los dos valores centrales de la serie ordenada.

### **Ejemplo:**

Los alumnos de 5to grado de una escuela son 10 y han obtenido los siguientes puntajes en una prueba de Botánica: 12, 7, 10, 16, 18, 10, 16, 13, 18 y 19. Hallar la mediana de tales puntajes.

Ordenando los puntajes: 7, 10, 10, 12, 13, 16, 16, 18, 18, 19 Tomando en cuenta que los puntajes son 10, determinamos la posición de la mediana:

$$\text{Posición} = \frac{n + 1}{2}$$

$$\text{Posición} = \frac{10 + 1}{2}$$

$$\text{Posición} = 5,5 \text{ lugar}$$

Esto quiere decir que, la mediana es el valor medio entre el 5to y el 6to lugar de la distribución ordenada. Así pues:

$$\text{Mediana} = \frac{13 + 16}{2}$$

$$\text{Mediana} = 14,5$$

## LA MODA

Es el valor que se repite con mayor frecuencia en una serie de datos.

### **Ejemplo:**

La moda de las calificaciones de 6to. Grado, del ejercicio anterior es 8 puntos, pues, dicho valor se repite con más frecuencia que los demás.

### **Ejemplo:**

La moda de la serie de puntajes del 5to. Grado, del ejercicio anterior es 10 y 18 puntos, pues, dichos valores son los que se repiten con más frecuencia que los otros.

Nota.- Así como existen series bimodales, existen otras que no tienen moda.

## MEDIDAS DE DISPERSIÓN

Para describir en forma adecuada un conjunto de datos, a más de obtener información respecto a la parte media del conjunto ordenado de números, es conveniente también tener una medida que nos indique si los valores están relativamente cercanos el uno del otro o si se encuentran dispersos.

Estudiaremos cuatro medidas de dispersión:

- la amplitud total,
- La desviación media,
- La varianza; y,
- La desviación estándar.

Todas estas medidas, excepto la primera, toman a la media aritmética como punto de referencia. En cada caso, un valor cero indica que no hay dispersión, en tanto que la dispersión es mayor a medida que se hace mayor el valor de cualquiera de las medidas de dispersión que hemos enunciado.

### **LA AMPLITUD TOTAL**

La amplitud de variación de un conjunto de números es una medida sencilla de calcular. Se puede expresar en dos formas:

- a) Dando la diferencia entre los valores mayor y menor del conjunto de datos.
- b) Identificando los números mayor y menor del conjunto.



**Ejemplo:**

Hállese la amplitud de los siguientes conjuntos de números; 1, 5, 7 y 13.  
14, 3, 17, 73 y 8

Números	diferencia	del más bajo al más alto
1, 5, 7, 13	12	1 a 13
14, 3, 17, 73, 8	70	3 a 73

Este último método tiende a ser más informativo. Así, saber sólo que la amplitud de un conjunto de números es 44, no dice nada más respecto de los demás números, sin embargo, si se informa que la amplitud de dichos números es de 300 a 344, se proporciona mejor información acerca de la magnitud de los números del conjunto.

La principal limitación de la amplitud es que considera solamente los valores extremos de un conjunto y no proporciona mayor información respecto de los demás valores intermedio del conjunto.

**LA DESVIACIÓN MEDIA**

Esta medida de variabilidad mide la desviación promedio de los valores de un conjunto de números con respecto a la media aritmética del conjunto; es decir, se basa en las diferencias de cada uno de los datos del conjunto con relación a la media .

Se obtiene sumando los valores absolutos de las desviaciones y dividiendo para el total de casos:

$$Dm = \frac{\sum |X - \bar{X}|}{n}$$

**Ejemplo:**

Hállese la desviación media para el siguiente conjunto de números. 2, 4, 6, 8, 10.

Para mayor comodidad de los cálculos, los datos se disponen de la siguiente manera:

<b>X</b>		
2	-4	4
4	-2	2
6	0	0
8	+2	2
10	+4	4
<b><math>\Sigma x = 30</math></b>		

$$Dm = \frac{\sum |X - \bar{X}|}{n}$$

$$Dm = \frac{12}{5}$$

$$Dm = 2,4$$

$$\bar{X} = \frac{\sum X}{n}$$

$$\bar{X} = \frac{30}{5}$$

$$\bar{X} = 6$$

Los pasos necesarios para calcular la Dm. son los siguientes:

1. Calcular la media de los datos.
2. Restar la media de cada valor del conjunto de datos, es decir, calcular las desviaciones.
3. Sumar los valores absolutos de las desviaciones.
4. Dividir esta suma entre el número de datos.

Para datos agrupados en una distribución de frecuencia debe usarse la siguiente fórmula:

$$Dm = \frac{\sum f|X - \bar{X}|}{n}$$

**Ejemplo:**

Hállese la desviación media del siguiente conjunto de calificaciones:

1, 1, 2, 2, 2, 2, 3, 3, 3, 3 4, 4, 4, 5, 5, 5, 5, 6, 6, 8

Para comodidad de los cálculos, los datos se disponen mediante la siguiente distribución:

<b>X</b>	<b>f</b>	<b>f.X</b>		
1	2	2	-2.7	5.4
2	4	8	-1.7	6.8
3	4	12	-0.7	2.8
4	3	12	+0.3	0.9
5	4	20	+1.3	5.2
6	2	12	+2.3	4.6
8	1	8	+4.3	4.3
	<b>n = 20</b>	<b>∑ f.x = 74</b>		

$$Dm = \frac{\sum f|X - \bar{X}|}{n}$$

$$Dm = \frac{30}{20}$$

$$Dm = 1,5$$

$$\bar{X} = \frac{\sum fX}{n}$$

$$\bar{X} = \frac{74}{20}$$

$$\bar{X} = 3,7$$

**Ejercicio:**

Hállese la Dm. del mismo conjunto de calificaciones, sin necesidad de construir previamente la distribución de frecuencias.

## LA VARIANZA

La varianza ( $S^2$ ) se define como la suma de los cuadrados de las desviaciones de cada uno de los datos con respecto a la media, dividido todo por el número total de datos.

La varianza o variancia de un pequeño conjunto de datos se calcula casi de la misma forma que la Dm, con la diferencia de que las desviaciones se elevan al cuadrado antes de ser sumadas.

Algebraicamente la varianza se expresa y se calcula mediante la siguiente fórmula:

$$S^2 = \frac{\sum(X - \bar{X})^2}{n}$$

### Ejemplo:

Hállese la varianza del conjunto: 2, 4, 6, 8, 10.

<b>X</b>		
2	-4	16
4	-2	4
6	0	0
8	+2	4
10	+4	16

$$\bar{X} = \frac{\sum x}{n}$$

$$\bar{X} = \frac{30}{5}$$

$$\bar{X} = 6$$

$$S^2 = \frac{\sum(X - \bar{X})^2}{n}$$

$$S^2 = \frac{40}{5}$$

$$S^2 = 8$$

Los pasos necesarios para calcular la varianza son los siguientes:

1. Calcular la media de los datos.
2. Restar la media de cada valor del conjunto.
3. Elevar al cuadrado cada una de estas desviaciones.
4. Sumar los cuadrados de las desviaciones; y,
5. Dividir para n.

Para datos presentados en una distribución de frecuencias, la fórmula de la varianza es la siguiente:

$$S^2 = \frac{\sum f(X - \bar{X})^2}{n}$$

Obsérvese que la fórmula es más general que la anterior, porque la frecuencia de cada dato, en el caso de que ninguno se repite, es uno,  $f = 1$ . Por esta razón debemos considerar la  $f$  implícita cuando no se dé.

Ejemplo: Hállese la varianza del siguiente conjunto de datos: 1, 2, 3, 3, 4, 4, 5, 5, 5, 5, 7, 7, 7, 8, 9

X	f	fX			
1	1	1	-4	16	16
2	1	2	-3	9	9
3	2	6	-2	4	8
4	2	8	-1	1	2
5	4	20	0	0	0
7	3	21	2	4	12
8	1	8	3	9	9
9	1	9	4	16	16
n = 15		∑fX = 75			

$$s^2 = \frac{\sum f(x - \bar{x})^2}{n} \qquad \bar{x} = \frac{\sum fX}{n}$$

$$s^2 = \frac{72}{15} \qquad \bar{x} = \frac{75}{15}$$

$$s^2 = 4,8 \qquad \bar{x} = 5$$

### Ejercicio:

Hállese la varianza del conjunto de datos anterior, sin agruparlos en una distribución de frecuencias.

Otra fórmula que suele emplearse en el cálculo de la varianza es:

$$s^2 = \frac{\sum X^2}{n} - \left( \frac{\sum X}{n} \right)^2$$

Esta fórmula, algunas veces, es más fácil de utilizar que la anterior, ya que no se requiere calcular cada una de las desviaciones.

### Ejemplo:

Hállese la varianza del siguiente conjunto de números: 2, 4, 6, 8, 10.

X	X <sup>2</sup>
2	4
4	16
6	36
8	64
10	100
∑X = 30	∑X <sup>2</sup> = 220

n = 5

$$\bar{x} = \frac{\sum X}{n}$$

$$\bar{x} = \frac{30}{5}$$

$$\bar{x} = 6$$

Cuando los datos se han organizado en una tabla de frecuencias, la fórmula para calcular la varianza es:

$$s^2 = \frac{\sum fX^2}{n} - (\bar{X})^2$$

**Ejemplo:**

Hállese la varianza del siguiente conjunto de datos: 1, 2, 3, 3, 4, 4, 5, 5, 5, 5, 7, 7, 7, 8.

X	f	f.x	f.x <sup>2</sup>
1	1	1	1
2	1	2	4
3	2	6	18
4	2	8	32
5	4	20	100
7	3	21	147
8	1	8	64
	<b>n = 14</b>	<b>66</b>	<b>366</b>

$$\bar{X} = \frac{\sum x}{n}$$

$$\bar{X} = \frac{66}{14}$$

$$\bar{X} = 4,71$$

$$s^2 = \frac{\sum f.x^2}{n} - (\bar{X})^2$$

$$s^2 = \frac{366}{14} - (4,71)^2$$

$$s^2 = 26,14 - 22,18$$

$$s^2 = 3,96 \text{ u}^2$$

La varianza es una medida de dispersión en la que hallamos el promedio de las desviaciones al cuadrado. Esto significa que, por ejemplo, para el caso de las notas de los estudiantes, la unidad de varianza es puntos<sup>2</sup>. Para superar esta inconveniencia y disponer de una medida de dispersión para los puntajes, que se exprese en puntos, se halla la raíz cuadrada de la varianza, llamada desviación estándar o desviación típica.

## DESVIACIÓN ESTANDAR

Así, pues, la desviación estándar es simplemente la raíz cuadrada positiva de la varianza:

Desviación estándar =

$$S = \sqrt{s^2}$$

**Ejemplo:**

Hállese la desviación estándar del siguiente conjunto de datos:

X	f	f.X	f.X <sup>2</sup>
5	1	5	25
10	2	20	200
15	4	60	900
20	2	40	800
25	1	25	625
	<b>n = 10</b>	<b>150</b>	<b>2.550</b>

$$S^2 = \frac{\sum f \cdot x^2}{n} - (\bar{X})^2$$
$$S^2 = \frac{2550}{10} - (15)^2$$
$$S^2 = 30 u^2$$

$$S = \sqrt{S^2}$$
$$S = \sqrt{30 u^2}$$
$$S = 5,48 u$$

$$\bar{X} = \frac{\sum x}{n}$$
$$\bar{X} = \frac{150}{10}$$
$$\bar{X} = 15$$

## MEDIDAS DE TENDENCIA CENTRAL

En el análisis de grandes conjuntos de datos, las medidas de tendencia central son las mismas que para los pequeños conjuntos de datos.

### 1. MEDIA ARITMETICA

Cuando los datos se presentan en tablas de frecuencias, los valores de las variables son desconocidos y han sido reemplazados por categorías de datos, los intervalos de clase, de los cuales el punto medio es el valor más representativo de todos aquellos valores de la variable que caen dentro de dicho intervalo. Utilizando tales puntos medios ( $X$ ) se halla una buena aproximación en el cálculo de la media de datos agrupados.

La fórmula para calcular dicha media es la misma que utilizamos para la media de una distribución de frecuencia simple:

f.x: producto de cada punto medio por su frecuencia;  
n = número total de casos.

Ejemplo:

Hállese la media de los datos consignados en la siguiente distribución:

#### CALIFICACIONES DE QUÍMICA

INTERVALOS	FRECUENCIAS
30-39	4
40-49	6
50-59	8
60-69	12
70-79	9
80-89	7
90-99	4

#### CALIFICACIONES DE QUÍMICA

CALIFICACIONES INTERVALO	FRECUENCIA f	PUNTO MED. x	fx fx
30-39	4	34.5	138
40-49	6	44.5	267
50-59	8	54.5	436
60-69	12	64.5	774
70-79	9	74.5	670.5
80-89	7	84.5	591.5
90-99	4	94.5	378
<b>TOTAL</b>	<b>n = 50</b>		<b><math>\Sigma fx = 3255</math></b>

## METODO BREVE PARA EL CÁLCULO DE LA MEDIA ARITMÉTICA

- A = media supuesta (generalmente el punto medio del intervalo de clase de mayor frecuencia)  
 d = Desviación de cada punto medio desde A  
 $n = (x - A)$   
 n = número total de casos

Utilizamos la siguiente ecuación:

Ejemplos:

Hállese la media de los siguientes datos:

INTERVALO	f	x	d	f.d
30-39	4	34.5	-30	-120
40-49	6	44.5	-20	-120
50-59	8	54.5	-10	- 80 (-320)
60-69	12	64.5	0	0
70-79	9	74.5	10	90
80-89	7	84.5	20	140
90-99	4	94.5	30	120 (+350)
<b>TOTAL</b>	n = 50			$\Sigma fd = 30$

## 2. MEDIANA

Es el punto de distribución debajo del cuál queda el 50 % de los casos.

Su valor se obtiene con la siguiente ecuación.

$l_1$  = Límite real inferior del intervalo de clase en el cual se cumplen la mitad de los datos ( $n/2$ )

$(\Sigma f)_1$  = Frecuencia acumulada que se encuentra inmediatamente por debajo de la mediana

$n = (x - A)$

f = Frecuencia del intervalo de clase mediano

i = Longitud del intervalo de clase.

**Ejemplo:**

Hállese la mediana con los siguientes datos

INTERVALO	f	Fa
30-39	4	4
40-49	6	10
50-59	8	18
60-69	12	<b>30</b>
70-79	9	39
80-89	7	46
90-99	4	50
<b>TOTAL</b>	n = 50	



Cuando el resultado de  $n/2$  coincide con la  $Fa$  de uno de los intervalos de la distribución, la mediana es igual al límite real superior de dicho intervalo

### Ejemplo

Hállese la mediana con los siguientes datos:

INTERVALO	f	Fa
4-8	1	1
9-13	2	3
14-18	6	9
19-23	4	<b>13</b>
24-28	8	21
29-33	3	24
34-38	0	24
39-43	2	26
<b>TOTAL</b>	n = 26	

$$Mdna = 23.5$$

## MEDIDAS DE POSICIÓN

### 1. LOS CUARTILES

Son los puntos que dividen a la distribución en cuatro partes iguales.

**El primer cuartil.-** ( $Q_1$ ) es el punto de la distribución debajo del cual queda el 25 % de los datos

**El segundo cuartil.-** ( $Q_2$ ) es la mediana

**El tercer cuartil.-** ( $Q_3$ ) es el punto de la distribución debajo del cual queda el 75 % de los datos

Los cuartiles se calculan de la misma manera que la mediana, pero entonces, la fórmula de la mediana se modifica, reemplazando  $n/2$  por  $n/4$  y  $3n/4$ , respectivamente

$$Q_1 = l_1 + \left( \frac{\frac{n}{4} - (\sum f)_1}{f} \right) i$$

$$Q_3 = l_1 + \left( \frac{\frac{3n}{4} - (\sum f)_1}{f} \right) i$$

## GUIA DE ESTUDIO

1. Leer detenidamente el documento guía
2. Subrayar las ideas principales
3. Desarrollar la presente guía

### ESTADÍSTICA DESCRIPTIVA

1. ¿Defina lo qué es estadística?
2. ¿De que se ocupa la estadística descriptiva?
3. ¿De qué se ocupa la estadística Inferencial?
4. ¿A qué llamamos población?
5. ¿A qué llamamos muestra?
6. ¿Cuáles son las condiciones necesarias para determinar una muestra?
7. ¿Cite cuatro ejemplos de aplicación de la estadística descriptiva en su carrera?
8. ¿A que llamamos datos estadísticos?
9. ¿A qué llamamos caracteres?
10. ¿A que llamamos variables?
11. ¿Cite cuatro ejemplos de variables cuantitativas en educación?
12. ¿Cite cuatro ejemplos de variables cualitativas en educación?
13. ¿Para qué utilizamos la notación sigma?
14. Con un ejemplo calcule la media aritmética de un grupo pequeño de datos
15. Con un ejemplo calcule la media aritmética de un grupo grande de datos
16. ¿Con un ejemplo determine la mediana?
17. ¿Con un ejemplo determine la desviación media?
18. ¿Con un ejemplo determine la varianza?
19. ¿Con un ejemplo construya la distribución de frecuencias?
20. ¿Con un ejemplo determine la frecuencia acumulada?
21. ¿Con un ejemplo determine la frecuencia relativa?
22. ¿Con un ejemplo elabore un polígono de frecuencias e interprete?
23. ¿Con un ejemplo elabore una curva de frecuencias acumuladas e interprete?
24. ¿Con un ejemplo elabore un histograma e interprete?
25. Con gráficos explique la diferencia entre un histograma y un gráfico de barras
26. ¿Con un ejemplo elabore un grafico de sectores circulares e interprete?

COMPILADOR: Dr. JORGE ARMIMJOS Mg. Sc <b>COORDINADOR DE LA          CARRERA DE          FÍSICO-MATEMÁTICAS</b>
---