



ESTADÍSTICA

Objetivos:

Conocer y trabajar conceptos básicos de la estadística descriptiva.

Analizar situaciones representadas en los gráficos.

Adquirir habilidades para conseguir una tabla de frecuencias, un diagrama de barras y algunos índices descriptivos

Saber interpretar una tabla de frecuencias, un diagrama de barras y algunos índices descriptivos

Observar la ganancia en términos de facilidad de interpretación y de disposición de la información original

Relacionar la información contenida en la tabla con el tipo de variable que se está tabulando

Cuestionar la validez y generalidad de las afirmaciones en relación a los conocimientos del método estadístico.

Un poco de historia...

La estadística tiene antecedentes históricos en los censos, que consistían en observaciones sistemáticas y periódicas sobre datos de la población para fines de guerra y finanzas realizados años antes de Cristo.

En la actualidad la estadística no limita solamente a la recolección de datos, sino a la organización, recopilación y análisis de los mismos con un determinado objetivo.

La palabra "estadística" suele utilizarse bajo dos significados distintos, a saber:

1° **Como colección de datos numéricos.** Este es el significado más vulgar de la palabra estadística. Se sobrentiende que estos datos numéricos han de estar presentados de manera ordenada y sistemática. Una información numérica cualquiera puede no constituir una estadística, para merecer este apelativo, los datos han de constituir un conjunto coherente, establecido de forma sistemática y siguiendo un criterio de ordenación.

Hay numerosas situaciones en las cuales es necesario organizar una gran cantidad de datos para que estén disponibles y se pueda leer toda la información que ellos proveen. En otras oportunidades hace falta hacer un resumen de esa colección de datos para tomar decisiones.

Tenemos muchos ejemplos de este tipo de estadísticas. El Anuario Estadístico publicado por el Instituto Nacional de Estadística, El Anuario de Estadísticas del Trabajo,...

2° **Como ciencia.** En este significado, la Estadística estudia el comportamiento de los fenómenos de masas. Como todas las ciencias, busca las características generales de un colectivo y prescinde de las particulares de cada elemento.

Así por ejemplo al investigar el sexo de los nacimientos, se inicia el trabajo tomando un grupo numeroso de nacimientos y obteniendo después la proporción de varones. Es frecuente enfrentarse con fenómenos en los que es muy difícil predecir el resultado; así, no se puede dar una lista, con las personas que van a morir con una cierta edad, o el sexo de un nuevo ser hasta que transcurra un determinado tiempo de embarazo,...

Por tanto, el objetivo de la estadística es hallar las regularidades que se encuentran en los fenómenos de masa.

También es útil usar datos obtenidos a partir de una parte de un grupo para predecir qué sucederá con el grupo entero.

Sus aplicaciones se dan en todos los campos de la investigación, siendo utilizada como medio auxiliar, entre otros por economistas, médicos, físicos y técnicos.

Vamos a ver cómo una vez recolectados los datos cómo se organiza la información para sacar conclusiones válidas y cómo transmitir claramente esta información.



ESTADÍSTICA

Estadística descriptiva y estadística inductiva.

La estadística puede dividirse en dos partes:

- ✓ *Estadística descriptiva*: es la parte de la estadística que trata de analizar un grupo dado sin sacar conclusiones de un grupo mayor que lo contenga.
- ✓ *Estadística inductiva*: es la parte de la estadística que partiendo de los datos de una muestra deduce o infiere conclusiones de la población de la que procede. Al no ser absoluta la veracidad de las conclusiones, con frecuencia se utiliza en ellas el término de probabilidad.

Población, muestra, individuos y caracteres

Es obvio que todo estudio estadístico ha de estar referido a un conjunto o colección de personas o cosas. Este conjunto de personas o cosas es lo que denominaremos **población ó universo**.

Las personas o cosas que forman parte de la población se denominan **individuos**. En sentido estadístico un individuo puede ser algo con existencia real, como personas, animales, plantas, artículos elaborados, consumidores, bacterias, estudiantes, votantes, un automóvil o una casa, o algo más abstracto como la temperatura, un voto, niveles de venta o un intervalo de tiempo.

A su vez, cada individuo de la población tiene una serie de características que pueden ser objeto del estudio estadístico. Así por ejemplo si consideramos como individuo a una persona, podemos distinguir en ella los siguientes **caracteres**:

Sexo, Edad, Nivel de estudios, Profesión, Peso, Altura, Color de pelo, etc.

Luego por tanto de cada elemento de la población podremos estudiar uno o más aspectos cualidades o caracteres.

Por ejemplo:

Si se efectúa un censo sobre viviendas:

Población ----- el total de viviendas

Individuo ----- cada vivienda

Si se efectúa un censo sobre las notas de los alumnos:

Población ----- el total de notas de escritos de los alumnos

Individuo ----- la nota de cada uno de los alumnos

Muchas veces es muy difícil estudiar el número total (N) de los individuos de una población debido al costo económico o el tiempo que ocasiona o por ser muy difícil de delimitar el número total.

Por ejemplo, si se desea saber la diferencia de presión la presión sanguínea promedio de la población masculina argentina, sería imposible determinar la presión de todos los hombres que habitan nuestro país. El problema se resuelve recurriendo a las muestras.

Se llama **muestra** al conjunto de n individuos ($n < N$) elegidos al azar entre los N de una población dada.

Luego, si en el ejemplo anterior se toma la presión a 200 hombres elegidos al azar:

Población ----- todos los hombres que habitan nuestro país.

Muestra ----- los 200 hombres escogidos.

Individuo ----- cada hombre.

Para que los estudios realizados sobre la muestra sean válidos, la muestra debe ser representativa de la población.

Por ejemplo, si se quiere saber la audiencia de un programa televisivo y se interrogan al azar una cantidad de abonados telefónicos, la muestra de la población no es tomada al azar, ya que sólo representa a la gente que tiene teléfono.



ESTADÍSTICA

En general, se toman muestras sucesivas al azar de una misma población y sobre ellas se trabaja para estimar medidas de la población.

La población puede ser según su tamaño de dos tipos:

Población finita: cuando el número de elementos que la forman es finito, por ejemplo el número de alumnos de un centro de enseñanza, o grupo clase.

Población infinita: cuando el número de elementos que la forman es infinito, o tan grande que pudiesen considerarse infinitos. Como por ejemplo si se realizase un estudio sobre los productos que hay en el mercado. Hay tantos y de tantas calidades que esta población podría considerarse infinita.

Cuando en un estudio estadístico no se puede trabajar con todos los elementos de la población sino que se realiza sobre un subconjunto de la misma. Este subconjunto puede ser una **muestra**, cuando se toman un determinado número de elementos de la población, sin que en principio tengan nada en común; o una **subpoblación**, que es el subconjunto de la población formado por los elementos de la población que comparten una determinada característica, por ejemplo de los alumnos del Polimodal Universitario, la subpoblación formada por los alumnos de 3º, o la subpoblación de los varones.

Variables

Se denomina variable al aspecto que se pretende estudiar de una población, ya se considerando todos los individuos que la componen o una muestra representativa.

Por ejemplo las variables estudiadas en la situación pueden ser, sexo, edad, peso, estatura (por edad), etc.

Las variables se pueden clasificar, de acuerdo con el tipo de característica a la que se refieren:

✓ **Cualitativas:** se refieren a características **no medibles** o **atributos**. Por ejemplo, sexo, estado civil, escolaridad, candidato, etc.

✓ **Cuantitativas:** se refieren a características **medibles**. Por ejemplo, peso, estura, notas de escritos, etc.

Entre las variables cuantitativas se distinguen, a su vez, dos categorías diferentes:

✓ **Variable discreta:** cuando toma valores pertenecientes a conjuntos numéricos discretos (finito o numerable). Aquellas que por su naturaleza no admiten un fraccionamiento de la unidad, por ejemplo número de hermanos, páginas de un libro, número de libros, población de bacterias, número de plantas, edad, etc.

✓ **Variable continua:** que toma valores pertenecientes a conjuntos numéricos continuos, son aquellas variables que por su naturaleza admiten que entre dos valores cualesquiera pueda tomar cualquier valor intermedio, por ejemplo peso, tiempo, estatura, etc.

No obstante en muchos casos el tratamiento estadístico hace que a variables discretas las trabajemos como si fuesen continuas y viceversa.

Los **atributos** son aquellos caracteres que para su definición precisan de palabras, es decir, no le podemos asignar un número. Por ejemplo Sexo Profesión, Estado Civil, etc.

A su vez las podemos clasificar en:

- **Ordenables:** Aquellas que sugieren una ordenación, por ejemplo la graduación militar, el nivel de estudios, etc.
- **No ordenables:** Aquellas que sólo admiten una mera ordenación alfabética, pero no establece orden por su naturaleza, por ejemplo el color de pelo, sexo, estado civil, etc.



ESTADÍSTICA

Nota: cuando las observaciones se realizan sobre caracteres que no se pueden expresar mediante un valor numérico (escolaridad, estado civil, etc.), se suele extender el concepto de variable asignándole valores numéricos.

Por ejemplo,

Escolaridad	Sin escolaridad	EGB	Polimodal	Terciaria
Se asigna	0	1	2	3

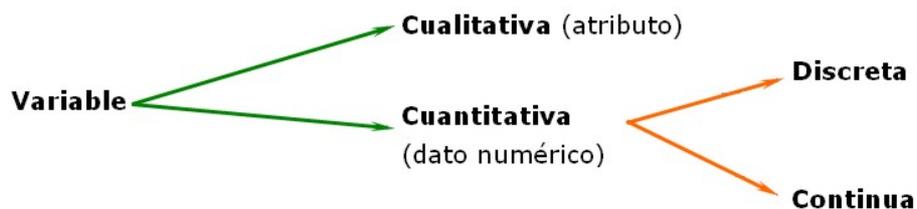
Las variables también se pueden clasificar en:

Variables unidimensionales: sólo recogen información sobre una característica (por ejemplo: edad de los alumnos de una clase).

Variables bidimensionales: recogen información sobre dos características de la población (por ejemplo: edad y altura de los alumnos de una clase).

Variables pluridimensionales: recogen información sobre tres o más características (por ejemplo: edad, altura y peso de los alumnos de una clase).

Resumiendo:



ORGANIZACIÓN DE LOS DATOS

La organización de los datos se realiza de diferentes maneras, mediante tabulaciones y representaciones gráficas.

Una de los primeros pasos que se realizan en cualquier estudio estadístico es la tabulación de resultados, es decir, recoger la información de la muestra resumida en una tabla.

Representación de tallo y hojas

Un método para iniciar el análisis exploratorio de los datos, previo al uso de los métodos estadísticos tradicionales, y que además proporciona información rápida, visual y es relativamente nueva, es la representación gráfica de tallo y hojas. Esta representación se basa en la ordenación de los datos a manera de gráfico, pero sin llegar a ello, utilizando las decenas y las unidades.

☞ Ejemplo que contiene las calificaciones obtenidas en una prueba de matemáticas:

78	93	61	100	70	83	88	74	97	72
66	73	76	81	83	64	91	70	77	86



ESTADÍSTICA

Se usa cada uno de los datos separando las decenas de las unidades, es decir, el número 51 se verá como 5 | 1.

6		1 6 4
7		8 0 4 2 3 6 0 7
8		3 8 1 3 6
9		3 7 1
10		0

El procedimiento para realizarla es primero empezar con el tallo, es decir la columna de la izquierda, y después dato por dato ir llenando las hojas a la derecha de la línea vertical, en el tronco correspondiente.

Además, si se desean tener los datos ordenados, y hay gente que lo prefiere así, se pueden ordenar las hojas en cada renglón para que la representación quede como sigue:

6		1 4 6
7		0 0 2 3 4 6 7 8
8		1 3 3 6 8
9		1 3 7
10		0

En realidad una representación de tallo y hojas presenta la misma información que la lista original de datos, pero de una manera mucho más compacta (especialmente si la lista de datos es más grande) y manejable.

Sin embargo, información más compleja resulta un poco más difícil de manejar, por lo que en ocasiones conviene redondear los datos, ignorar sus partes decimales o utilizar las centenas.

Variables cuantitativas discretas. Distribución de frecuencias

Los datos obtenidos a partir de un relevamiento se ordenan en tablas, llamadas de **distribución de frecuencias**.

Se tienen los siguientes tipos de frecuencia:

1. Frecuencia absoluta
2. Frecuencia relativa
3. Porcentaje
4. Frecuencia absoluta acumulada
5. Frecuencia relativa acumulada
6. Porcentaje acumulado

La cantidad de veces que se repite determinado valor de la variable se llama **frecuencia absoluta (f_i)**.

La frecuencia absoluta, es una medida que está influida por el tamaño de la muestra, al aumentar el tamaño de la muestra aumentará también el tamaño de la frecuencia absoluta. Esto hace que no sea una medida útil para poder comparar. Para esto es necesario introducir el concepto de **frecuencia relativa (f_r)**, que es el cociente entre la frecuencia absoluta y el tamaño de la muestra.

$$f_r = \frac{f_i}{n} \quad \text{donde } n = \text{tamaño de la muestra}$$

Significa qué parte del total representa cada valor de la variable.



ESTADÍSTICA

La frecuencia relativa es un *tanto por uno*, sin embargo, hoy día es bastante frecuente hablar siempre en términos de *tantos por ciento* o **porcentajes**, por lo que esta medida resulta de multiplicar la frecuencia relativa por 100. La denotaremos (**f%**).

$$f_{\%} = p_i = \frac{f_i}{n} \cdot 100$$

Para poder calcular este tipo de frecuencias hay que tener en cuenta que la variable estadística ha de ser cuantitativa o cualitativa ordenable. En otro caso no tiene mucho sentido el cálculo de esta frecuencia.

Muchas veces interesa conocer cuántos datos se acumulan hasta cierto valor, para lo cual habrá que sumar a la frecuencia de ese valor, la frecuencia de los valores anteriores. A esta suma parcial se la llama **frecuencia acumulada**.

Frecuencia absoluta acumulada (Fi): cada una de las frecuencias absolutas de la columna es igual a la suma de las frecuencias absolutas anteriores. Puede calcularse en forma creciente o en forma decreciente.

Frecuencia relativa acumulada (Fr): cada una de las frecuencias relativas de la columna es igual a la suma de las frecuencias relativas anteriores. Puede calcularse en forma creciente o en forma decreciente.

Ejemplo:

- Se considera un grupo de 34 alumnos de 1° año de la Escuela Polimodal UNCPBA, de los cuales se enumeran las notas obtenidas en el 1° escrito de matemática del año 2006.

Toma de datos: 9, 5, 7, 9, 8, 7, 9, 9, 9, 8, 7, 10, 9, 8, 3, 9, 8, 3, 7, 9, 10, 8, 5, 6, 3, 7, 9, 9, 8, 8, 9, 7, 9, 7.

Nota, x_i	f_i	fr	$f\%$	F_i crec	F_i dec	Fr crec	Fr dec
3	3	0,088	8,82	3	34	0,088	1,000
5	2	0,059	5,88	5	31	0,147	0,912
6	1	0,029	2,94	6	29	0,176	0,853
7	7	0,206	20,59	13	28	0,382	0,824
8	7	0,206	20,59	20	21	0,588	0,618
9	12	0,353	35,29	32	14	0,941	0,412
10	2	0,059	5,88	34	2	1,000	0,059
Suma	34	1,00	100				

Observaciones:

- La suma de las frecuencias absolutas es igual al número de individuos de la población.
- La suma de las frecuencias relativas es siempre igual a 1.

Cuando se tienen variables con frecuencia absoluta cero no se colocan en la tabla.

Variables cuantitativas continuas. Intervalos de clase

Se denomina **intervalo de clase** a cada uno de los intervalos en que pueden agruparse los datos de una variable cuantitativa continua.

¿Cómo se determinan los intervalos de clase?

Para tabular los datos agrupados en intervalos de clase hay que tener en cuenta algunas cuestiones:



ESTADÍSTICA

- ☞ Generalmente se toman todos los intervalos con la misma amplitud. Depende de qué se esté midiendo.
- ☞ Cada dato debe pertenecer a un único intervalo.
- ☞ La cantidad de intervalos depende de la cantidad de datos y de la precisión que se necesita.

Muchas veces se toman intervalos que son abiertos en un extremo y cerrados en el otro, ya que cada dato debe pertenecer sólo a un intervalo. Otras veces, se consideran como extremos de los intervalos, valores que seguro no pueden ser valores de la variable en cuestión.

¿Cómo se calcula la amplitud de cada intervalo?

Para determinar la amplitud de cada intervalo hay que tener en cuenta el mayor y el menor de los datos. Se realiza la diferencia entre el mayor y el menor y se divide por el número de intervalos que se quiere armar.

Es muy importante determinar el *número adecuado de intervalos de clase* que conforman la distribución, el cual dependerá de la cantidad y de la naturaleza de los datos. Debe tenerse en cuenta que si el número de intervalos se elige muy grande, pierde sentido trabajar con este método; por otro lado, si el número de intervalos se elige demasiado pequeño, puede distorsionarse el comportamiento de la distribución. Se sugiere no tener menos de cinco intervalos.

Se denomina **marca de clase** (x_{mi}) al punto medio del intervalo. Se calcula de la siguiente manera:

$$x_{mi} = \frac{\text{límite superior} + \text{límite inferior}}{2}$$

Representación gráfica

Variables cuantitativas discretas.

Los gráficos permiten visualizar el comportamiento de la distribución de frecuencias.

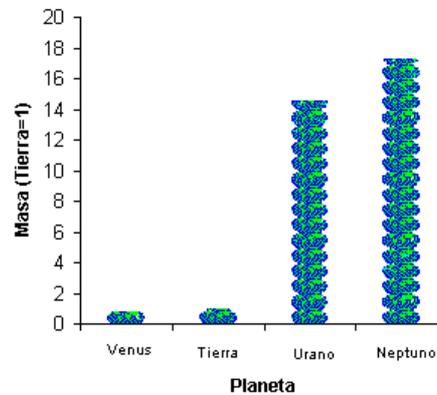
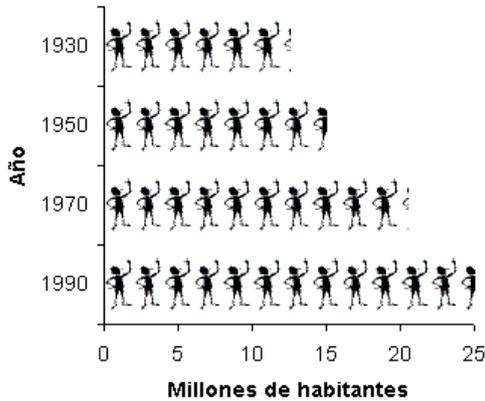
Se presentan los **gráficos de barras**. En el eje horizontal del gráfico, en la base de las barras, se colocan el atributo que adquiere la variable cuantitativa o el valor de la variable cuantitativa discreta. En el eje vertical, se pueden indicar las frecuencias absolutas o las frecuencias relativas. La altura de cada barra representa la frecuencia de la categoría correspondiente.

Actualmente, debido al desarrollo que adquirió el diseño con computadoras, se suelen reemplazar los diagramas de barras por **pictogramas**, para que al lector le resulte más atractivo. En el pictograma, la frecuencia está representada por dibujos alegóricos al tema en estudio. Estos diagramas son usuales en medios masivos de comunicación.

Ejemplos:



ESTADÍSTICA



El de la izquierda representa la población de los Estados Unidos (cada hombrecillo representa a dos millones de habitantes), el de la derecha representa la masa de tres planetas de nuestro sistema solar tomando como unidad a la masa de la Tierra (cada planeta representa la masa de nuestro planeta: Venus tiene masa menor y Neptuno tiene más 17 veces más masa que la Tierra).

Las versiones del *Excel 7.0* y anteriores **no** tienen opciones para realizar este tipo de gráficas, las posteriores sí. Otros programas contemporáneos (como el *Corel Draw* o el *Harvard Graphics*) sí son capaces.

Otra forma de mostrar el comportamiento de una distribución de variable cualitativa o de variable cuantitativa discreta es por medio de los **gráficos de sectores o circulares**. Cada zona corresponde a un atributo o a un valor de la variable. El área de cada sector es proporcional a la frecuencia correspondiente.

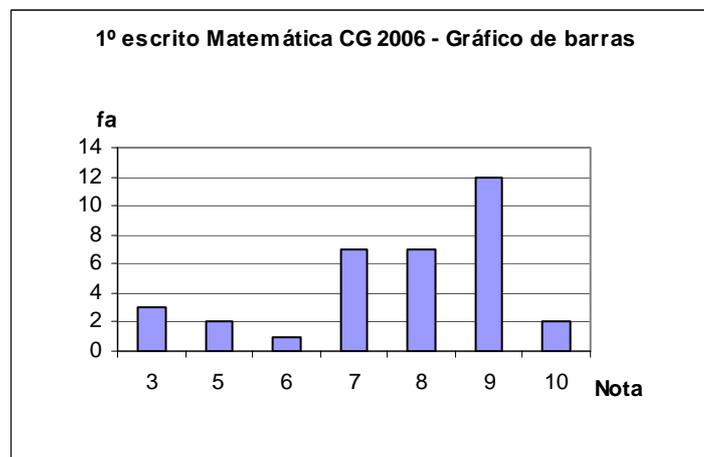
Para determinar en forma manual el ángulo central correspondiente a cada sector circular se calcula de la siguiente manera:

$$\frac{\text{ángulo central}}{360^\circ} = \frac{\text{frecuencia absoluta}}{\text{total de individuos u observaciones}}$$

ó bien,

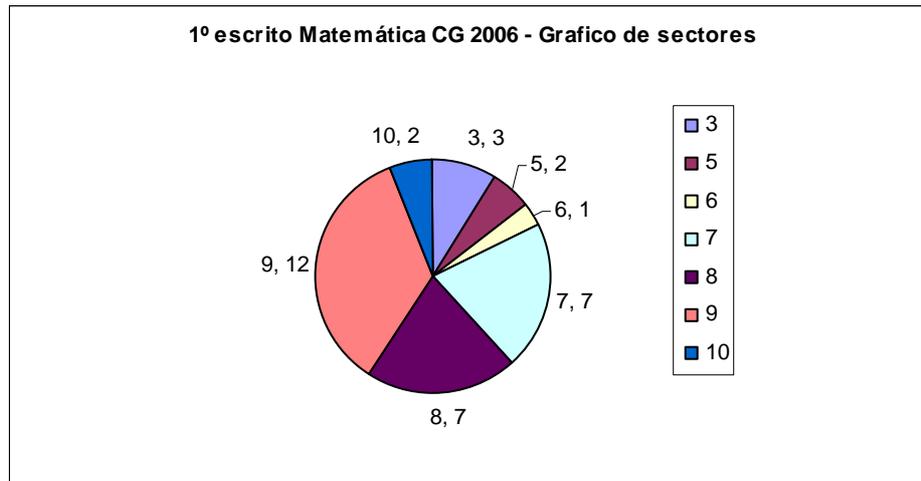
$$\text{ángulo central} = \text{frecuencia relativa} \times 360^\circ$$

A continuación, se presentan los gráficos de barras y de sectores para el ejemplo considerado.





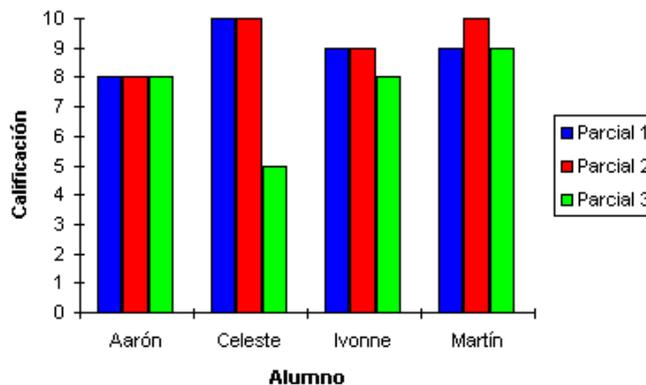
ESTADÍSTICA



En la actualidad, los gráficos se pueden realizar en forma sencilla con la aplicación del Microsoft Excel, sin necesidad de calcular escalas para los gráficos de barras o sectores circulares para el gráfico de sectores.

Otra observación pertinente es que se pueden representar en la misma gráfica, utilizando las mismas escalas horizontales y verticales, varios datos correspondientes a las mismas variables producto de varias observaciones. Esto produce una gráfica con varias **series**, correspondiendo cada una de ellas a cada observación de la muestra (o población).

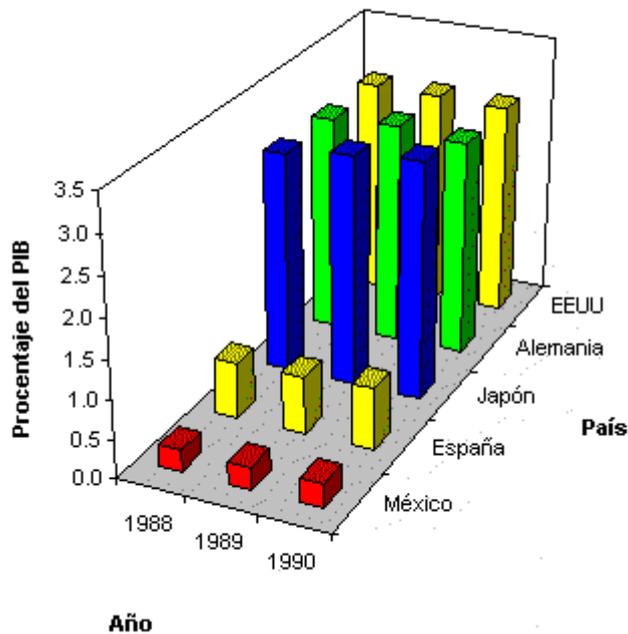
El ejemplo que sigue pertenece al comportamiento de las calificaciones parciales de tres alumnos de Polimodal. Las series (cada una de las calificaciones parciales) están coloreadas con diferente color para mostrar el comportamiento tanto individual, como de cada uno de los alumnos con respecto a los demás. Es interesante observar que la escala horizontal no es continua (es nominal).



Existe la posibilidad, y si los recursos lo permiten, de representar gráficos compuestos de una manera "tridimensional", es decir, con gráficos que posean no sólo dos ejes, sino tres; y en los que los rectángulos son sustituidos por prismas de base rectangular. Un ejemplo es el siguiente:



ESTADÍSTICA



donde se representa el porcentaje del PIB gastado en docencia e investigación por cinco países en el lapso de 1988 a 1999 (fuente: Revista "Ciencia y Desarrollo", 1994, XIX(114):12).

Es importante considerar que este tipo de gráficos puede complicarse mucho, haciendo que la información sea menos legible.

Una representación gráfica correctamente hecha permite muchas veces visualizar rápidamente el conjunto de datos y las relaciones entre ellos. Se debe tener mucho cuidado con las escalas que se utilizan pues muchas veces los gráficos resultan engañosos.

Variables cuantitativas continuas.

El gráfico que se utiliza para variables continuas es el **histograma**, muy similar al gráfico de barras, pero tiene algunas diferencias importantes.

En el eje horizontal (o de las abscisas) se representan los intervalos de los datos, marcándose de manera continua las fronteras entre cada uno de los éstos. De esta manera, el histograma está compuesto por rectángulos, cuyo número coincide con la cantidad de intervalos considerados, el ancho de la base de cada uno de esos rectángulos es la misma siempre y coincide con las fronteras de los intervalos, y la altura corresponde a la frecuencia de cada intervalo.

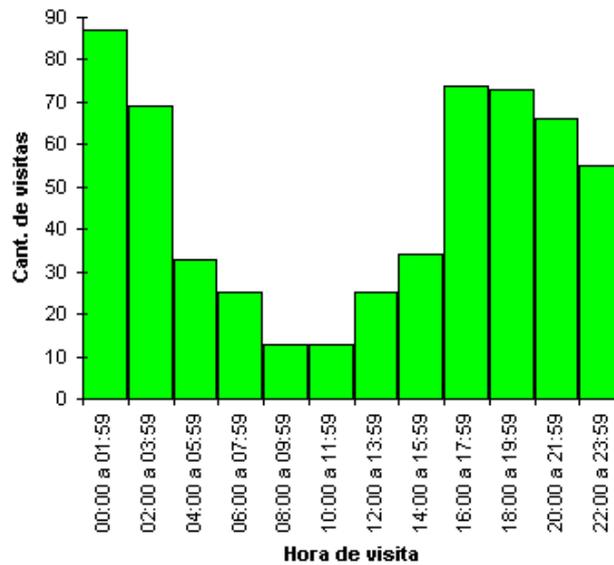
Es importante observar que resulta difícil utilizar este tipo de representación cuando existen intervalos abiertos o cuando los intervalos no son iguales entre sí.

Para construir un histograma, en el eje horizontal se marcan los intervalos de clase y se toma cada uno de ellos como la base de un rectángulo, cuya altura es la frecuencia (o bien absoluta o bien relativa) de cada clase.

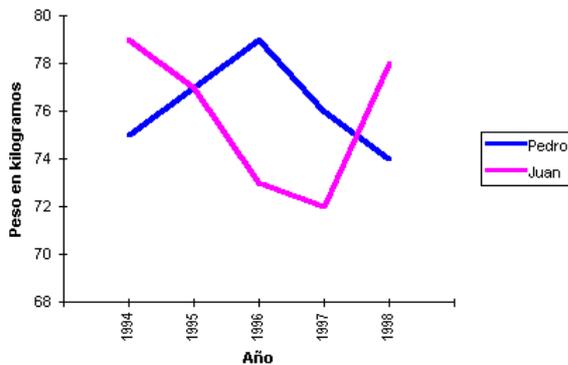
En el ejemplo que se presenta a continuación, que representa el número de "visitas" que ha tenido esta una página web de acuerdo a la hora de la visita.



ESTADÍSTICA



Quando los datos se relacionan entre sí, es decir, cuando podemos decir que existe cierta continuidad entre las observaciones (como por ejemplo el crecimiento poblacional, la evolución del peso o estatura de una persona a través del tiempo, el desempeño académico de un estudiante a lo largo de su instrucción escolar, las variaciones presentadas en la medición realizada en algún experimento cada segundo o minuto) se pueden utilizar las **gráficas de líneas**, que consisten en una serie de puntos trazados en las intersecciones de las marcas de clase y las frecuencias de cada una, uniéndose consecutivamente con líneas:



Este ejemplo se muestra el comportamiento del peso corporal (en kilogramos) de dos individuos a lo largo de cinco observaciones anuales.

Es posible presentar varias series de observaciones (en este caso cada serie de observaciones son los pesos de un individuo).

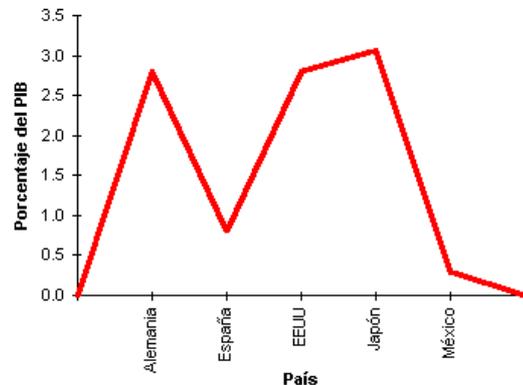
Otra forma de representación de un uso menos común, y muy parecida a las gráficas de líneas, es el **polígono de frecuencias**. La diferencia fundamental entre ambas es que en el polígono de frecuencias se añaden dos clases con frecuencias cero: una antes de la primera clase con datos y otra después de la última. El resultado es que se "sujeta" la línea por ambos extremos al eje horizontal y lo que podría ser una línea separada del eje se convierte, junto con éste, en un polígono.



ESTADÍSTICA

El siguiente ejemplo corresponde al porcentaje del PIB gastado en docencia e investigación durante el año de 1990 en cinco países

(fuente: Revista "Ciencia y Desarrollo", 1994, XIX(114):12)



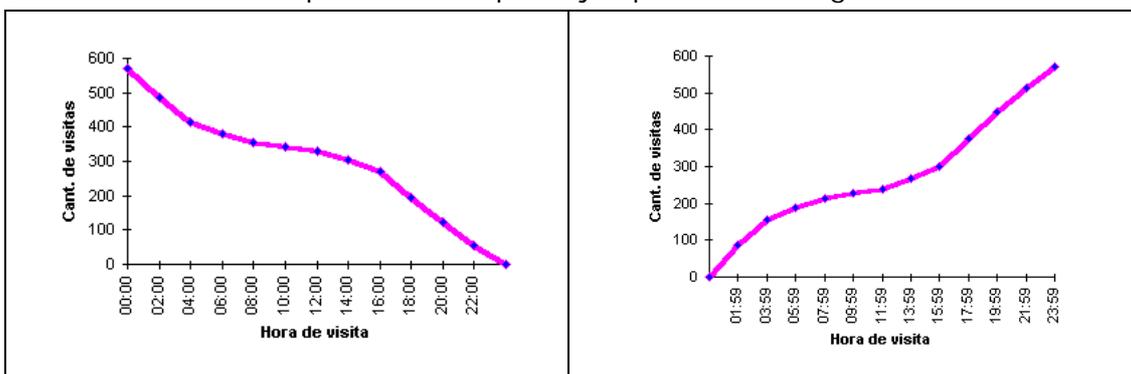
El *Excel* no crea automáticamente polígonos de frecuencias, sino que produce gráficas de líneas. Sin embargo, es posible arreglárselas para hacerlas.

Una gráfica similar al polígono de frecuencias es la **ojiva**, pero ésta se obtiene de aplicar parcialmente la misma técnica a una distribución acumulativa y de igual manera que éstas, existen **ojivas mayor que** y **ojivas menor que**.

Existen dos diferencias fundamentales entre las ojivas y los polígonos de frecuencias (y por ésto la aplicación de la técnica es parcial):

1. Un extremo de la ojiva no se "amarrar" al eje horizontal, para la *ojiva mayor que* sucede con el extremo izquierdo; para la *ojiva menor que*, con el derecho.
2. En el eje horizontal en lugar de colocar las marcas de clase se colocan las fronteras de clase. Para el caso de la *ojiva mayor que* es la frontera menor; para la *ojiva menor que*, la mayor.

Las siguientes son ejemplos de ojivas, a la izquierda la *mayor que*, a la derecha la *menor que*, utilizando los datos que se usaron para ejemplificar el histograma:



La ojiva mayor que (izquierda) se le denomina de esta manera porque viendo el punto que está sobre la frontera de clase "4:00" se ven las visitas que se realizaron en una hora *mayor que* las 4:00 horas (en cuestiones temporales se diría: *después de las 4:00 horas*). De forma análoga, en la ojiva menor que la frecuencia que se representa en cada frontera de clase son el número de observaciones *menores que* la frontera señalada (en caso de tiempos sería el número de observaciones *antes* de la hora que señala la frontera).



ESTADÍSTICA

Resumen de tablas y gráficos.

Cuadro a modo de síntesis donde se muestran las formas de tabular y representar los datos correspondientes a distintos tipos de variables.

Tabla de distribución de frecuencias	de Variables cualitativas; variables cuantitativas discretas.
Tabla de distribución de frecuencias con intervalos de clase	de Variables cuantitativas continuas; variables cuantitativas discretas cuando presentan gran cantidad y diversidad de valores *.
Gráfico de sectores o circular	Para presentar cualquier tipo de distribución.
Gráfico de barras	Para presentar distribuciones de variables cualitativas y de variables cuantitativas discretas.
Pictogramas	Reemplaza a los diagramas de barras, para variables cualitativas y también a los gráficos de sectores.
Histograma	Para representar distribuciones de variables cuantitativas continuas con agrupamiento de datos.

(*) Si bien las distribuciones con datos agrupados son utilizadas, en general para casos de variable continua, muchas veces se adaptan a distribuciones de variable discreta. Esto sucede cuando la distribución presenta mucha diversidad en los datos.

Parámetros estadísticos

En el resto del tema nos ocupamos exclusivamente de las variables cuantitativas, puesto que con los atributos no se pueden realizar operaciones aritméticas. Como hemos estudiado.

Los métodos gráficos permiten percibir rápidamente ciertas características de una distribución de datos. No obstante, en algunos casos, para describir una distribución es necesario definir algunas medidas numéricas, llamadas **medidas numéricas descriptivas**.

Las medidas estadísticas pretenden "resumir" la información de la "muestra" para poder tener así un mejor conocimiento de la población. Se clasifican en

1. Medidas de centralización o de posición:
 - o Que sirven para determinar los valores centrales o medios de la distribución
2. Medidas de variabilidad o de dispersión:
 - o Nos van a dar una idea sobre la representatividad de las medidas centrales, a mayor dispersión menor representatividad.
3. Medidas de localización:
 - o Útiles para encontrar determinados valores importantes, para una "clasificación" de los elementos de la muestra o población. (se describirán brevemente)
4. Medidas de la simetría:



ESTADÍSTICA

- Sirven para ver si la distribución tiene el mismo comportamiento por encima y por debajo de los valores centrales. (sólo se mencionan)

Cuando las medidas se calculan sobre la base de datos de una muestra, se denominan estadísticas. Las estadísticas se usan como base para hacer inferencias acerca de ciertas características numéricas de la población que reciben el nombre de parámetro estadístico.

Medidas de tendencia central, de posición o de centralización

Las medidas de tendencia central permiten conocer cómo se concentran los datos de una distribución alrededor de cierto valor, dicho de otro modo, las medidas de posición se utilizan para encontrar un valor que represente a todos los datos. Las más empleadas son:

- ✓ Media
 - aritmética (promedio)
 - geométrica
 - armónica
- ✓ Modo o moda.
- ✓ Mediana.

Media o media aritmética

La media aritmética de una variable se define como la suma ponderada de los valores de la variable por sus frecuencias relativas. Es el promedio de las observaciones. Esta medida se simboliza \bar{x} .

Para calcularla se pueden usar las siguientes expresiones.

- Para *distribuciones de variable discreta*. Se calcula como el cociente de la suma de todos los productos de las observaciones por la frecuencia absoluta dividido por el número total de observaciones.

$$\bar{x} = \frac{\sum x_i \cdot f_i}{n}$$

n = número total de observaciones; x_i cada uno de los distintos valores que toma la variable.

- Para *distribuciones de variable continua separada en intervalos de clase o por datos agrupados*. En primer lugar se calculan las marcas de clase o los puntos medios de cada uno de los intervalos. Luego, la media aritmética se calcula de la siguiente manera:

$$\bar{x} = \frac{\sum x_{mi} \cdot f_i}{n}$$

En general, el promedio es una medida de tendencia central no confiable.

Propiedades:

1. Si multiplicamos o dividimos todas las observaciones por un mismo número, la media queda multiplicada o dividida por dicho número.
2. Si le sumamos a todas las observaciones un mismo número, la media aumentará en dicha cantidad.

La media geométrica y la media armónica son de uso poco frecuente.



ESTADÍSTICA

Moda o modo

La moda es el valor de la variable que tenga mayor frecuencia absoluta, la que más se repite, es la única medida de centralización que tiene sentido estudiar en una variable cualitativa, pues no precisa la realización de ningún cálculo.

Por su propia definición, la moda no es única, pues puede haber dos o más valores de la variable que tengan la misma frecuencia siendo esta máxima. En cuyo caso tendremos una distribución bimodal o polimodal según el caso.

Puede suceder que no haya ninguna moda (si todos los valores tienen la misma frecuencia).

Por lo tanto el cálculo de la moda en distribuciones discretas o cualitativas no precisa de una explicación mayor; sin embargo, debemos detenernos un poco en el cálculo de la moda para distribuciones cuantitativas continuas.

Es el valor que se repite mayor cantidad de veces. Se simboliza **Mo**.

Se utiliza como medida de centralización cuando se están utilizando variables cualitativas.

- Para una *distribución de variable discreta* es el valor de la variable al cual le corresponde mayor frecuencia absoluta.
- Para una *distribución de variable continua separada en intervalos de clase o por datos agrupados*, la moda es la marca de clase correspondiente al intervalo de mayor frecuencia absoluta.

El intervalo al cual pertenece la moda, se denomina **intervalo modal** o **clase modal**.

Mediana

Es el valor de la variable que ocupa el lugar central al ordenar los datos en forma creciente o decreciente, es el valor que divide a la distribución en dos partes de igual cantidad de observaciones. Si la cantidad de datos es par, la mediana es el promedio entre los dos valores centrales. Se simboliza **Me**.

Para calcular la mediana debemos tener en cuenta si la variable es discreta o continua.

- Para una *distribución de variable discreta* se determina de la siguiente manera:

Si n es impar, es el dato numérico central;

Si n es par, es el promedio de los dos datos centrales.

- Para las *distribuciones de variable continua separada en intervalos de clase o por datos agrupados*, se calcula el orden de la mediana:

$$\text{orden de la mediana} = \frac{n+1}{2}$$

Luego se va a la columna de las frecuencias absolutas acumuladas (F_i), se ubica el orden de la mediana y se lee la mediana en la columna de las variables. La clase que tiene la posición central es la clase de la mediana.



ESTADÍSTICA

A continuación, se presentan los cálculos de las medidas de posición para el ejemplo considerado:

Nota, xi	fi	xi . fi
3	3	9
5	2	10
6	1	6
7	7	49
8	7	56
9	12	108
10	2	20
	34	258

Media aritmética:

$$\bar{x} = \frac{258}{34} = 7,59$$

Moda:

$$Mo = 9$$

Mediana:

Datos ordenados en forma creciente:

3 3 3 5 6 7 7 7 7 7 7 8 8 8 8 8 8 8 9 9 9 9 9 9 9 9 9 9 10 10

$$Me = 8$$

Conclusión: Para este caso es más significativa la medida de la mediana debido a que el promedio se halla afectado por valores extremos como el 3 y el 10. En este caso la moda no se debiera tener en cuenta. Podemos decir que el rendimiento académico del curso en su 1° escrito fue de 8 puntos, pues la mediana es la medida más adecuada para este tipo de problemas.

Población, muestra, individuos y caracteres

Medidas de dispersión o de variabilidad.

Para conocer el comportamiento general de la distribución de frecuencias no alcanza con determinar su valor central. Es necesario, además, analizar la variabilidad de los datos, lo que implica estudiar cómo se encuentran esparcidos dentro de la distribución. Dan idea del alejamiento de los datos respecto de la media aritmética.

Si hay poca dispersión significa que las medidas de tendencia central son representativas. Las más empleadas son:

- ✓ Rango
- ✓ Concepto de desviación
- ✓ Desviación Media
- ✓ Varianza o variancia
- ✓ Desviación Típica o desvío estándar
- ✓ Cuasivarianza
- ✓ Cuasi Desviación típica
- ✓ Coeficiente de Variación

Rango o amplitud

Se define como la diferencia existente entre el valor mayor y el menor de la distribución. Lo notaremos como R. Realmente no es una medida muy significativa en la mayoría de los casos, pero indudablemente es muy fácil de calcular.

$$V_{\text{máx}} - V_{\text{mín}}$$



ESTADÍSTICA

Hemos estudiado varias medidas de centralización, por lo que podemos hablar de desviación con respecto a cualquiera de ellas, sin embargo, la más utilizada es con respecto a la media.

Desviación

Es la diferencia que se observa entre el valor de la variable y la media aritmética ($x_i - \bar{x}$)

No es una medida, son muchas medidas, pues cada valor de la variable lleva asociada su correspondiente desviación, por lo que precisaremos una medida que resuma dicha información.

Varianza

Es el promedio (o media) de los cuadrados de las desviaciones o desvíos. Se simboliza σ^2 .

Para calcular la varianza se siguen los siguientes pasos:

Se calcula la diferencia entre cada dato y el promedio.

Se eleva al cuadrado cada una de las diferencias anteriores.

Se suman todos los valores hallados en el paso anterior y se divide el resultado por la cantidad de datos obteniéndose σ^2 .

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2 \cdot f_i}{n}$$

Este estadístico tiene el inconveniente de ser poco significativo, pues se mide en el cuadrado de la unidad de la variable, por ejemplo, si la variable viene dada en cm. La varianza vendrá en cm^2 .

Desviación típica o Desvío estándar

Mide la dispersión de los datos respecto al promedio. Cuanto menor es el desvío estándar, menos dispersos están los datos con respecto al promedio. Se simboliza σ .

Para calcular el desvío estándar se saca la raíz cuadrada de la varianza.

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2 \cdot f_i}{n}}$$

Este estadístico se mide en la misma unidad que la variable por lo que se puede interpretar mejor.

Otros dos estadísticos importantes son la cuasivarianza y la cuasidesviación típica, que en el tema de estimación estadística, son los estimadores de la varianza y desviación típica poblacionales respectivamente.



ESTADÍSTICA

Para el ejemplo que se venía desarrollando se obtienen los siguientes valores:

Nota, x_i	f_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
3	3	-4,59	21,0681
5	2	-2,59	6,7081
6	1	-1,59	2,5281
7	7	-0,59	0,3481
8	7	0,41	0,1681
9	12	1,41	1,9881
10	2	2,41	5,8081
Σ	34		38,6167
Varianza			1,1358

Varianza:

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{34} = \frac{38,6167}{34} = 1,1358$$

Desvío estándar:

$$\sigma = \sqrt{\sigma^2} = \sqrt{1,1358} = 1,07$$

Conclusión: No se presenta gran dispersión de los datos, aunque se continúan teniendo en cuenta lo dicho anteriormente.

Todos estos cálculos se pueden realizar más sencillamente con la utilización de la computadora a través del **Microsoft Excel** o de un programa especialmente diseñado para cálculos estadísticos llamado **InfoStat**.

Ejemplo: Se considera un grupo de 34 alumnos de 1° año Serú Girán de la Escuela Polimodal UNCPBA, de los cuales se enumeran las notas obtenidas en el 1° escrito de TICs del año 2006.

Datos: 6; 6,25; 6,50; 6,75; 7; 7,25; 7,50; 8; 8; 8; 8,25; 8,50; 8,75; 9; 9; 9; 9,25; 9,25; 9,25; 9,25; 9,50; 9,50; 9,50; 9,50; 9,50; 9,50; 9,50; 9,50; 9,50; 9,75; 9,75; 10; 10; 10; 10; 10

Este problema a diferencia del ejemplo anterior se considera de variable continua, por lo tanto se designarán intervalos de clase para obtener las tablas de frecuencias y marcas de clase para hacer el cálculo de las medidas de tendencia central y las de dispersión.

A continuación se presentan la tabla con los datos y cálculos correspondientes, como así también los gráficos.

IC	x_{mi}	f_i	fr	$f\%$	Fi crec	Fi dec	Fr crec	Fr dec	$x_i \cdot f_i$	$x_{mi} - \bar{x}$	$(x_{mi} - \bar{x})^2$
[6, 7)	6,50	4	0,118	11,76	4	34	0,118	1,000	26	-2,29	5,26297578
[7, 8)	7,50	3	0,088	8,82	7	30	0,206	0,882	22,5	-1,29	1,67474048
[8, 9)	8,50	6	0,176	17,65	13	27	0,382	0,794	51	-0,29	0,08650519
[9; 10]	9,50	21	0,618	61,76	34	21	1,000	0,618	199,5	0,71	0,4982699
Σ		34	1,00	100					299		7,52249135



ESTADÍSTICA

Estadísticos de centralización

Media aritmética:

$$\bar{x} = \frac{299}{34} = 8,79$$

Moda:

$$Mo = 9,50$$

Mediana:

Datos ordenados en forma creciente:

6; 6,25; 6,50; 6,75; 7; 7,25; 7,50; 8; 8; 8; 8,25; 8,50; 8,75; 9; 9; 9; 9,25; 9,25; 9,25; 9,25; 9,50; 9,50; 9,50; 9,50; 9,50; 9,50; 9,75; 9,75; 10; 10; 10; 10; 10.

Varianza:

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{34} = \frac{7,52249}{34} = 0,2212$$

Desvío estándar:

$$\sigma = \sqrt{\sigma^2} = \sqrt{1,1358} = 0,47$$

Medidas de Localización: Cuartiles, deciles y percentiles.

Las medidas de localización dividen la distribución en partes iguales, sirven para clasificar a un individuo o elemento dentro de una determinada población o muestra.

- ✓ Cuartiles.
- ✓ Deciles.
- ✓ Percentiles.

Cuartiles

Medida de localización que divide la población o muestra en cuatro partes iguales.

- Q_1 = Valor de la variable que deja a la izquierda el 25% de la distribución.
- Q_2 = Valor de la variable que deja a la izquierda el 50% de la distribución (mediana).
- Q_3 = Valor de la variable que deja a la izquierda el 75% de la distribución.

Al igual que ocurre con el cálculo de la mediana, el cálculo de estos estadísticos, depende del tipo de variable.

Deciles

Medida de localización que divide la población o muestra en 10 partes iguales

No tiene mucho sentido calcularlas para variables cualitativas discretas. Por lo que lo vamos a ver sólo para las variables continuas.

Percentiles

Medida de localización que divide la población o muestra en 100 partes iguales

No tiene mucho sentido calcularlas para variables cualitativas discretas. Por lo que lo vamos a ver sólo para las variables continuas.