

CONCEPTO DE ESTADÍSTICA

El origen de la palabra “**estadística**” no es nada claro. Se dice que deriva del término “estadista” y éste a su vez del término “estado” (del latín status) refiriéndose a él como a la **situación en la que se encuentra una persona o cosa**.

El término estadística lo asociamos a los términos censo y recuento de población (de personas) ya que éstos se llevaban a cabo para conocer cómo una determinada población está conformada y cuántos en número la formaban, o sea su cualidad y cantidad, y esta preocupación de censar es de larga data. Como antecedentes podemos citar los siguientes:

- **Pentateuco**, engloba los 5 primeros libros del Antiguo Testamento. En el cuarto libro “Números” los primeros capítulos se referían al censo o numeración de las tribus israelitas.
- **Censo del emperador Yao** (China, 2.238 a. de J. C.). Primeras evidencias de recuentos de población.
- **Documentos asirios, egipcios y griegos**. La preocupación por la actividad censal de los individuos y bienes del Estado tenía una clara finalidad tributaria y militar.

Parece ser que la palabra “**ESTADÍSTICA**” fue utilizada por primera vez en Alemania en el siglo XVIII (18) como la “**Recolección de datos y documentos útiles para la administración del Estado**”. Es por todo este antecedente histórico que se la asocia con el estudio de hechos que se prestan a numeración o recuento y a la comparación de esos datos numéricos, o sea el análisis comparativo de los datos obtenidos.

La estadística incorpora técnicas de análisis de la información recopilada, esto es proporciona los métodos para organizar y resumir los datos para luego analizarlos y obtener conclusiones. Por lo tanto todo lo relacionado con la recolección, procesamiento, análisis e interpretación de los datos numéricos tienen que ver con el campo de la estadística.

La estadística tiene por objeto establecer los valores que son objeto de su estudio y determinar los valores medios o normales de los fenómenos expresados en números variables. Tengamos en cuenta que la estadística no sólo proporciona los métodos para analizar los resultados de los experimentos que se han efectuado sino también ofrece sugerencias sobre cómo se pueden llevar a cabo dichos experimentos para disminuir los efectos de la variación y tener mayor posibilidad de llegar a conclusiones correctas.

La **Estadística** es la rama de la investigación científica que proporciona métodos para organizar y resumir esa información o datos y luego obtener conclusiones.

La **Estadística** actúa como disciplina de interconexión entre los modelos matemáticos y los fenómenos reales. Un modelo matemático es una abstracción simplificada de una realidad más compleja. Por ello, siempre existirá cierta diferencia entre lo observado y lo previsto en el modelo. La estadística está considerada como una ciencia en sí misma.

¿Qué se entiende por Estadística?

Este término lo encontramos diario en nuestro lenguaje y tiene dos significados:

- a. **Uso más común:** se refiere a información numérica. Ejemplos, salario inicial promedio, venta promedio de automóviles, porcentaje de estudiantes que terminarán la carrera, número de muertes por alcoholismo en el último año, etc.
- b. **Otra menos común:** se refiere a la recolección de más de una cifra y la publicación de dicha información numérica.

Pero en realidad tiene un sentido más amplio y por eso definimos:

Estadística: es la ciencia de recolectar, organizar, presentar, analizar e interpretar datos para ayudar en una toma de decisión más efectiva.

Las técnicas estadísticas se usan ampliamente por personas en áreas de comercialización, contabilidad, control de calidad, consumidores, en los deportes, administración de hospitales, educación, política, medicina, etc.

Estadística descriptiva: métodos para organizar, resumir y presentar datos de manera informativa.

- Ejemplo 1: un sondeo de opinión encontró que 49% de las personas en una encuesta sabían el nombre del primer libro en la Biblia. La estadística “49” describe el número de personas cada 100 que saben la respuesta.
- Ejemplo 2: según el Consumer Reports, los dueños de lavadoras de ropa Whirlpool reportaron 9 problemas por cada 100 máquinas durante 1995. La estadística “9” describe el número de problemas por cada 100 máquinas.

Estadística inferencial: una decisión, estimación, predicción o generalización sobre una población, con base en una muestra.

- Ejemplo 1: las cadenas de TV monitorean la popularidad de sus programas contratando a Nielsen y otras organizaciones para muestrear las preferencias de televidentes.
- Ejemplo 2: el departamento de contabilidad de una empresa elegirá una muestra de facturas para verificar la exactitud de todas la facturación de la compañía.
- Ejemplo 3: los catadores de vino prueban unas cuantas gotas para tomar la decisión de liberar todo el lote de vino para la venta.

Fueron dos las áreas de interés que originaron, alrededor del siglo XVIII, la necesidad de la creación de esta ciencia:

- Los juegos de azar, mediante el estudio de las probabilidades de ganar.
- Las ciencias políticas. mediante la descripción de datos numéricos de las unidades políticas analizadas, tales como ciudades, provincias, distritos, etc.

O sea como resultado de la unión de dos disciplinas que evolucionaron independientemente hasta confluir en la Estadística:

- El cálculo de probabilidades desde el siglo XVII y
- La “Estadística descriptiva”

La combinación de ambas da lugar a una ciencia que busca cómo obtener conclusiones de la investigación empírica mediante la utilización de modelos matemáticos. Podemos decir que la estadística se divide en tres ramas:

- **Estadística descriptiva:** análisis de muestras, descripción cuantitativa y cualitativa de muestras o poblaciones.
- **Teoría de la probabilidad:** efectuar los cálculos y modelos para su posterior análisis.
- **Inferencia estadística:** efectuar las generalizaciones, estimaciones y predicciones en base al análisis de las muestras.

Las primeras tareas de la Estadística eran simplemente recopilar datos y llevar a cabo relevamientos, pero paulatinamente se le fueron incorporando técnicas de análisis de la información recopilada.

De la definición anterior, podemos notar que el primer paso en la investigación de un problema es la recolección de datos significativos. Estos deben organizarse de cierta forma y presentarse tal vez en gráficos, así podremos analizar e interpretar la información hasta que los datos hayan sido organizados.

En la actualidad la Estadística consiste entonces en las siguientes etapas:

1. **Recolección de los datos:** recuento, relevamiento, recopilación;
2. **Ordenamiento y presentación de los datos:** se lleva cabo el procesamiento de los datos y con ellos se hacen tabulaciones, agrupaciones de datos, graficación;
3. **Cálculo de medidas y parámetros estadísticos**
4. **Análisis y predicción de los resultados:** la medición e interpretación de los datos para, basándose en el principio de *Inferencia estadística*, hacer predicciones sobre el problema analizado.

Recolección de los Datos

En la **recolección de datos**, se toma la población de estudio y se obtienen los datos. Se efectúa un relevamiento de cada individuo de la población, o de la muestra. Para unificar criterios, antes de continuar, definamos algunos conceptos:

Población: es el conjunto sobre el cuál se llevarán a cabo las observaciones, sobre el cual se realizará el estudio. La población debe estar muy bien determinada al inicio del estudio y puede estar formada por personas, animales, cosas, períodos temporales, etc. Una población es un conjunto de todos los posibles individuos, objetos o medidas de interés. Es el conjunto de **todas** las observaciones o datos acerca de algo que nos interesa.

Individuo: es cada una de las unidades elementales de la población.

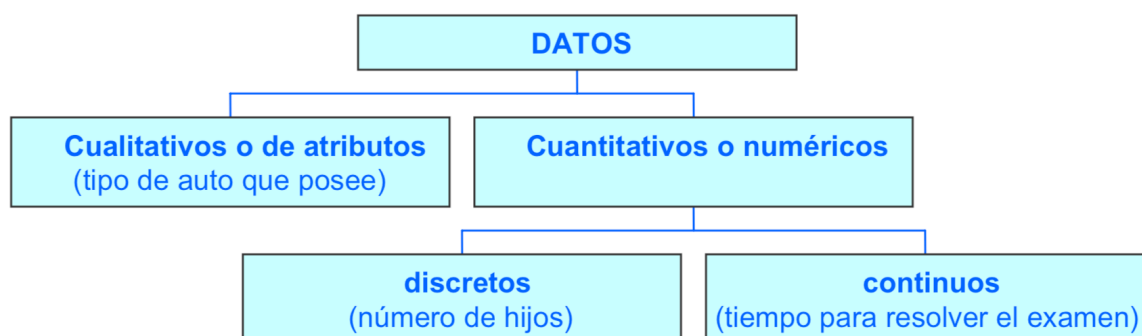
Muestra: cuando las poblaciones cuentan con muchos individuos, se hace imposible el estudio de todos y cada uno de los individuos de la población. O bien, cuando no disponemos de las herramientas necesarias para efectuar la medición de todos o tiempo suficiente para hacerlo, trabajamos con muestras. Las muestras son un subconjunto de individuos de la población sobre el cual se hace el estudio para poder generalizar las conclusiones a toda la población. Para poder lograr hacer una extrapolación de resultados y no estar muy alejados de la realidad, las muestras tienen que ser significativas de la población en número y sus miembros seleccionados al azar. Se dice que una muestra es representativa, cuando reproduce con la máxima exactitud posible las características de toda población, al menos en los aspectos que nos interesan estudiar. Una muestra es una porción, o parte, de la población de interés. O sea es un subconjunto representativo de la población que debe reflejar las características esenciales de la misma; para ello hay que asegurarse de que cada objeto de la población tenga igual oportunidad de ser elegido que el resto, con independencia de lo que ocurre con los restantes. Una muestra así elegida, se llama **muestra aleatoria**.

O sea, de la **población** se obtendrían todos los datos posibles del estudio, ya que es el conjunto o colección bien definido de todos los objetos de interés de la investigación y nos brinda una realidad hacia donde se enfoca el estudio. Las observaciones que se realicen se refieren a diferentes características o atributos de los individuos que la conforman. En cambio, una **muestra**

aleatoria es una parte de la población, pero al analizarla obtendré los mismos resultados, pues conserva las características del conjunto total.

A estas características cuyo valor puede variar, cambiar de un objeto al siguiente de la población, se las llama **variables**. Estas características pueden ser:

- **Cualitativas** o de atributos: la característica o variable que se estudia no es numérica. Por ejemplo: sexo, afiliación religiosa, tipo de automóvil que se posee, lugar de nacimiento, color de los ojos, nacionalidad, nivel cultural, etc. Estos atributos no se pueden determinar numéricamente, y a su vez se pueden clasificar en:
 - **Nominales:** no se les puede dar un orden (Por ejemplo, sexo, color de ojos).
 - **Ordinales:** se pueden ordenar (Por ejemplo, calificaciones: suficiente, insuficiente, bueno, malo, regular, etc.).
- **Cuantitativas:** la variable se puede registrar numéricamente, si se las puede medir numéricamente en cantidades **discretas** (números enteros) o **continuas** (números reales). A estas se las conoce como variables estadísticas.
 - **Variable estadística discreta:** cuando los valores que toma la variable son aislados. El caso más normal es aquel en que los valores que toma la variable son números enteros (por ejemplo: cantidad de hijos, cantidad de trabajadores en una empresa).
 - **Variable estadística continua:** cuando la variable estadística puede tomar cualquier valor de un determinado intervalo. En general, son variables estadísticas continuas todas las variables relacionadas con el espacio, el tiempo o la masa (por ejemplo, la altura, el peso, la distancia, la duración de procesos).



Los problemas de investigación (tales como crímenes, salud, importaciones y exportaciones, producción y costos por hora) suelen realizarse a partir de datos publicados. Se pueden encontrar estadísticas relacionadas en artículos publicados, revistas y periódicos u otras fuentes confiables.

No todos los temas disponen de datos publicados. En esos casos, la información deberá recolectarse y analizarse. Una manera habitual de recolectar datos es mediante encuestas.

Niveles de Medición

El nivel de medición de un dato marca los cálculos que pueden realizarse para resumir y presentar la información y las pruebas estadísticas que pueden desarrollarse. Existen cuatro niveles de medición: nominal, ordinal, de intervalo y de razón.

- **Nivel nominal:** los datos sólo se pueden clasificar en categorías, no se pueden ordenar. Ejemplos: color de los ojos, sexo, afiliación religiosa.

- **Nivel ordinal:** involucra datos que se pueden ordenar, pero no es posible determinar las diferencias entre los valores de los datos o, no tienen significado. Por ejemplo, en una prueba de sabor de 4 refrescos, al C se asignó el número 1, al B el número 2, al A el 3 y al D el 4.
- **Nivel de intervalo:** similar al nivel ordinal, con la propiedad adicional de que se pueden determinar cantidades significativas de las diferencias entre los valores. No existe un punto cero natural. Ejemplo: temperatura en la escala de grados Fahrenheit.
- **Nivel de razón:** el nivel de intervalo con un punto cero inicial inherente. Las diferencias y razones son significativas para este nivel de medición. Ejemplos: dinero en el patrimonio de distintos contribuyentes, altura de los jugadores de básquetbol de la NBA, etc.

Dentro de cada uno de los niveles de medición deben darse las siguientes dos categorías:

- **Mutuamente excluyente:** un individuo, objeto o artículo, al ser incluido en una categoría, debe excluirse de todas las demás. Ejemplo: color de los ojos.
- **Exhaustivo:** cada persona, objeto o artículo debe clasificarse en al menos una categoría. Ejemplo: afiliación religiosa.

Ordenación y Presentación de los Datos

Tabulaciones y Graficaciones

Los datos recopilados se anotan en planillas o fichas, luego se escriben en tablas en forma ordenada, constituyendo una serie simple de datos. La diferencia entre los valores extremos de la variable se llama **amplitud o rango recorrido**.

Una vez recogidos los datos lo primero que hay que observar es el número de repeticiones de cada valor, o sea se toma el valor y se cuentan las ocurrencias del mismo. El primer método que se utiliza para describir un conjunto de datos es la **distribución de frecuencias**. La distribución de frecuencias es el agrupamiento de datos en categorías que muestran el número de observaciones en cada categoría mutuamente excluyente.

Sea **X** una variable estadística y sea x_i uno de los valores que puede tomar, se denomina **frecuencia absoluta** del valor x_i (o simplemente frecuencia) al número de veces que se presenta este valor, el número de veces que la variable toma un determinado valor. Se representa por f_i .

La frecuencia relativa de la variable x_i es el cociente entre la frecuencia absoluta y la cantidad total de individuos, o sea el número de observaciones N .

$$fr_i = \frac{f_i}{N}$$

La **frecuencia relativa de un intervalo de clase** es el cociente de la frecuencia absoluta del intervalo dividido por el número total de observaciones.

A veces, es útil conocer cuántas observaciones (o qué proporción de las observaciones) se encuentran por debajo o por encima de cada valor. Esta información viene dada por lo que denominamos **frecuencia acumulada**.

Dada la variable estadística X , y siendo x_i uno de los valores que puede tomar, se denomina **frecuencia absoluta acumulada** de x_i a la suma de las frecuencias absolutas de los valores anteriores o iguales a x_i . La representamos por F_i .

Y la **frecuencia relativa acumulada** de x_i es la suma de las frecuencias relativas de los valores anteriores o iguales a x_i . Si construimos una tabla con las frecuencias de cada valor x_i , a este conjunto se lo llama **serie de frecuencias**.

Cuando los valores numéricos de las variables estadísticas son muchos, o cuando las frecuencias absolutas de la mayoría de los valores son unitarias, o cuando los datos son continuos, a los valores se los agrupa en **intervalos**, llamados **clases**.

Esta agrupación, de hacerla, debe ser tal que quede claramente definida, que muestre las necesidades e interés del estudio, ser exhaustiva (no puede haber ningún valor que no pertenezca a una clase) y mutuamente excluyente (ningún valor puede pertenecer simultáneamente a dos clases).

Normalmente, se arman entre 5 y 20 clases (grupos). En función de la cantidad de clases que se desean armar, se toma el rango de la muestra y ese valor se lo divide por la cantidad de grupos buscados, menos uno. Esto nos dará el **tamaño o ancho de cada clase**.

Por ejemplo, para la clase [150, 170), el 150 es el extremo inferior, es el menor elemento del intervalo y pertenece al intervalo. En cambio, el 170 es el extremo superior de la clase, es el mayor de los elementos del intervalo, pero **no** pertenece al intervalo. La cantidad de veces que una variable toma valores de un intervalo es la **frecuencia del intervalo de clase**.

A la hora de hacer cálculos tomaremos como representante de cada intervalo a su punto medio, valor que se denomina **marca de clase**. En la clase mencionada [150, 170) su marca de clase es 160, o sea $(150+170)/2$.

Trabajemos ahora con una muestra dada a fin de obtener las estadísticas necesarias para un análisis sobre la población.

Ejemplo 1. La cantidad de hermanos de un grupo de 25 alumnos de la Tecnicatura en Programación de la Regional Haedo, UTN, es la siguiente:

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

La tabla estadística que corresponde a estos datos es la siguiente:

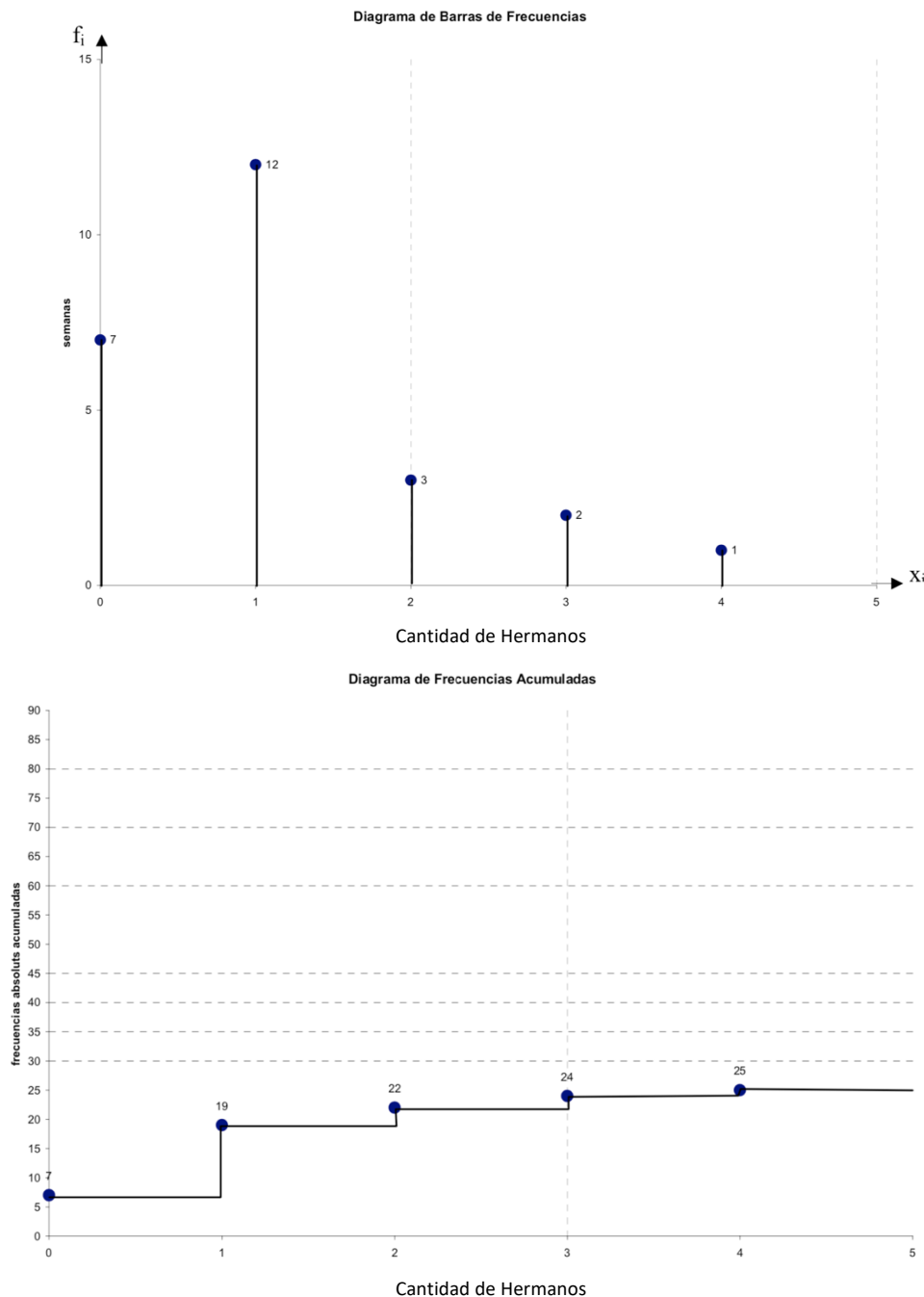
x_i	f_i	F_i	fr_i	Fr_i
0	7	7	0.28	0.28
1	12	19	0.48	0.76
2	3	22	0.12	0.88
3	2	24	0.08	0.96
4	1	25	0.04	1.00

Efectuados la recopilación de los datos y la creación de tablas de análisis se procede a la graficación o construcción de gráficos que permitan a través de ellos analizar los datos obtenidos. Esos gráficos permiten interpretar y visualizar mejor el fenómeno que se estudia, hacerlo en forma más simple y clara para poder comprender rápidamente el significado de las observaciones realizadas y cuál es la relación existente entre ellos.

Hay diferentes tipos de gráficos y en función de lo que se desea analizar se elige uno u otro: barras, sectores circulares ("pie charts" o gráficos de torta), mapas, curvas, histogramas, etc.

Diagrama de Barras

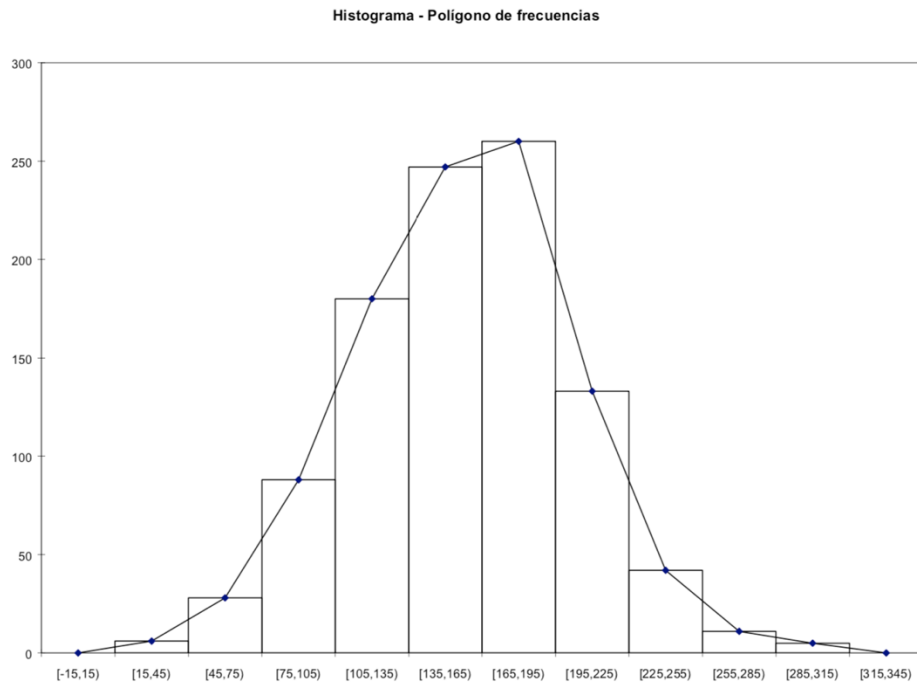
Se utiliza para variables estadísticas discretas con datos no agrupados. Consta de dos semiejes perpendiculares que se cortan en un punto, cartesianos ortogonales. Sobre uno de los semiejes (normalmente el horizontal), situamos los valores de la variable y sobre el otro eje las frecuencias. Finalmente, sobre cada valor se sitúa una barra de altura proporcional a la frecuencia del valor.



Histograma

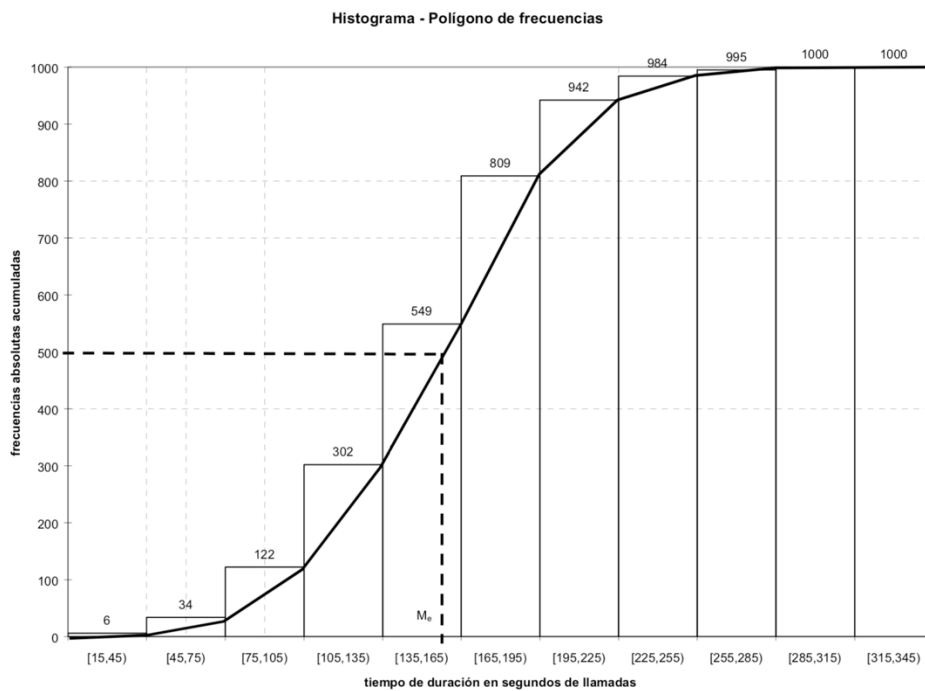
Se utiliza en distribuciones de frecuencias con los datos agrupados en intervalos de clase. Se construye situando en el eje de abscisas los intervalos y en el de ordenadas las frecuencias.

Después se dibuja para cada clase, un rectángulo con base la amplitud del intervalo y área proporcional a la frecuencia.



Polígono de Frecuencias

En el diagrama de barras o en el histograma, si en vez de dibujar barras, unimos con una poligonal los puntos que tienen como abscisa el punto medio de las barras y como ordenada la frecuencia correspondiente a esa barra, obtenemos el **polígono de frecuencias**.



Diagramas de Sectores

Es una representación en un círculo que se construye dibujando sobre éste un sector circular de área proporcional a la frecuencia absoluta de cada modalidad de la variable estadística. Se

utilizan cuando se quiere comparar entre sí las frecuencias correspondientes a los diferentes valores de una variable.

Mapas

Mediante la utilización de colores sobre mapas cartográficos se identifica las variables y la distribución que ellas tienen.

Cálculo de Medidas o Parámetros Estadísticos

Como estamos en el camino del análisis de los datos deberíamos tener algunos parámetros de comparación. Generalmente los valores tienden a concentrarse en ciertos valores llamados medidas o **parámetros de posición**, entre ellos se encuentran:

- **media aritmética (promedio)**
- **mediana**
- **moda**

Media aritmética: cuando los valores no están agrupados, es la suma de todos los valores dividida la cantidad o **tamaño de la muestra**. Si tenemos los datos ya tabulados, con sus frecuencias, calculamos la media aritmética ponderada o sea la suma de cada valor por su frecuencia dividida la cantidad total de datos. En cambio, si los valores están agrupados haremos la suma de la marca de cada clase ponderada con su frecuencia dividida el tamaño de la muestra.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

En el Ejemplo 1, la media aritmética es $\bar{x} = \frac{28}{25} = 1.12$

Mediana: una vez ordenados los datos de menor a mayor, la mediana es el valor que separa las observaciones en dos grupos de igual cantidad de datos.

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

En el Ejemplo 1, la mediana es el valor que está en la posición 13 o sea 1 ($25+1=26$, $26/2=13$).

Modo o Moda: es la observación de mayor frecuencia. En el Ejemplo 1, el 1 tiene 12 apariciones y es la mayor frecuencia, por lo tanto la moda es 1.

Cuartiles, Deciles, Percentiles: dividen las observaciones en 4, 10 y 100 partes iguales, respectivamente y se calculan con idéntico criterio que la mediana.

Por ejemplo, los cuartiles son 3 y los indicamos con q_1 , q_2 y q_3

q_1 : deja el 25% de las observaciones a su izquierda

q_2 : deja el 50% de los datos a su izquierda (luego coincide con la mediana)

q_3 : deja el 75% de los datos a su izquierda

Notar que el primer cuartil es la mediana de la primera mitad de los datos. Además, hay 9 deciles y 99 percentiles. En el Ejemplo 1, $q_1=0$, $q_2=1$, $q_3=1.5$

Rango o Recorrido: es la diferencia entre la mayor y la menor observación de una muestra o población. En el Ejemplo 1, el rango es $4-0 = 4$.

Rango Intercuartil: es la diferencia entre el tercer y primer cuartil. En el Ejemplo 1, siendo $q_1 = 0$, y $q_3 = 1.5$, el rango intercuartil es 1.5

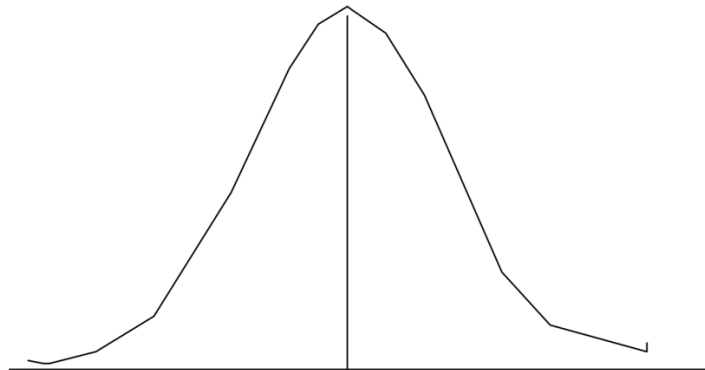
Se puede obtener una representación gráfica de una distribución de frecuencias en el caso que los datos estén agrupados construyendo así un **histograma**.

Se ubican en el eje de abscisas los intervalos de clase y se representa sobre cada intervalo un rectángulo cuya área sea proporcional a la frecuencia de dicho intervalo, de modo que si un intervalo tiene el doble de datos que otro, el área del rectángulo sobre ese intervalo también resulte el doble que el área del otro. En particular, si los intervalos son de igual longitud se pueden construir rectángulos cuya altura resulte ser la frecuencia.

Los histogramas presentan distintas formas, en función del comportamiento de las frecuencias. Si presentan un solo pico, se los llama **unimodales** (tiene una sola moda), si presentan dos se dice que son **bimodales** y con más de dos se los llama **multimodales** (o **polimodales**).

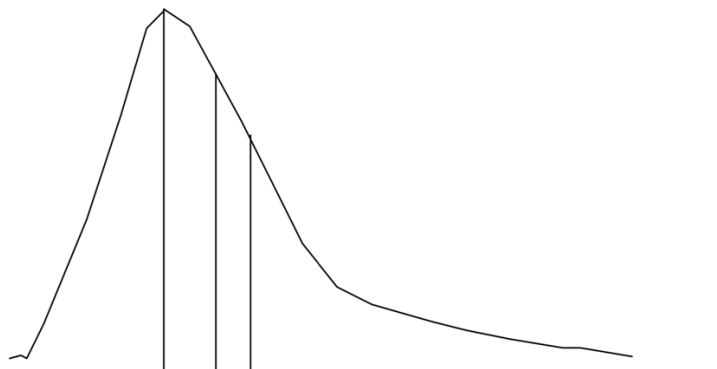
Sesgo cero: moda = mediana = media

Un histograma se dice que es simétrico si la mitad izquierda es una imagen reflejada de la otra mitad. En este caso no tiene sesgo o se dice que el sesgo es cero.



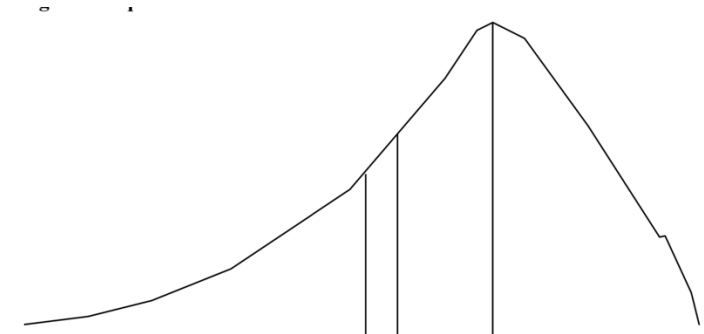
Sesgo a la derecha moda < mediana < media

Si un histograma unimodal tiene la cola derecha más alargada que la izquierda se dice que tiene sesgo positivo o que está sesgado positivamente.



Sesgo a la izquierda media < mediana < moda

Si un histograma unimodal tiene la cola izquierda más alargada que la derecha se dice que tiene sesgo negativo o que está sesgado negativamente.



El sesgo o la simetría se nota mejor cuando construimos el **polígono de frecuencias**, que consiste en unir los puntos medios de la base superior de cada rectángulo mediante una poligonal, la cual debe empezar y terminar sobre el eje de abscisas, para ello en nuestra tabla deberemos agregar dos intervalos de clase (uno al principio y otro al final) de frecuencias 0.

Si se conocen dos promedios de una distribución de frecuencias con sesgo moderado, el tercero se puede aproximar mediante las siguientes ecuaciones:

- $\text{moda} = \text{media} - 3 \times (\text{media} - \text{mediana})$
- $\text{media} = [3 \times (\text{mediana}) - \text{moda}] / 2$
- $\text{mediana} = [2 \times \text{media} + \text{moda}] / 3$

Datos agrupados

En ocasiones los datos se pueden presentar agrupados en lo que llamamos intervalos de clase, (entre 5 y 20 es lo habitual) y de ser posible de igual longitud. Se toman como intervalos semicerrados a izquierda $[;)$ y el punto medio del mismo se llama **marca de clase** y será el representante de todos los valores del intervalo.

Veamos a través de un ejemplo cómo tabular, representar y calcular las medidas.

Ejemplo 2: La siguiente tabla da la distribución del tiempo de 1.000 llamadas telefónicas, medidas en segundos. Vemos que las marcas de clase están igualmente espaciadas, por tanto se tomaron intervalos de igual longitud. Construyamos una tabla:

Clase	[15:45)	[45:75)	[75:105)	[105:135)	[135:165)	[165:195)	[195:225)	[225:255)	[255:285)	[285:315)
Marca	30	60	90	120	150	180	210	240	270	300
Frecuencia	6	28	88	180	247	26	133	42	11	5

Se puede obtener una representación gráfica de una distribución de frecuencias al construir un histograma, ubicando en el eje de abscisas los intervalos de clase y representando sobre cada intervalo un rectángulo cuya área sea proporcional a la frecuencia de dicho intervalo, de modo que si un intervalo tiene el doble de datos que otro, el área del rectángulo sobre ese intervalo también resulte el doble que el área del otro. En particular, si los intervalos son de igual longitud se pueden construir rectángulos cuya altura resulte ser la frecuencia.

Número	Intervalo	Marca (x_i)	f_i	F_i
1	[15,45)	30	6	6
2	[45,75)	60	28	34
3	[75,105)	90	88	122
4	[105,135)	120	180	302
5	[135,165)	150	247	549
6	[165,195)	180	260	809
7	[195,225)	210	133	942
8	[225,255)	240	42	984
9	[255,285)	270	11	995
10	[285,315)	300	5	1000

Para el cálculo de la **media** usamos la fórmula:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k f_i x_i$$

Siendo:

k : cantidad de intervalos.

x_i : la marca de la clase i.

f_i : la frecuencia de la clase i.

n: cantidad de valores de la muestra.

Para la **varianza muestral**:

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

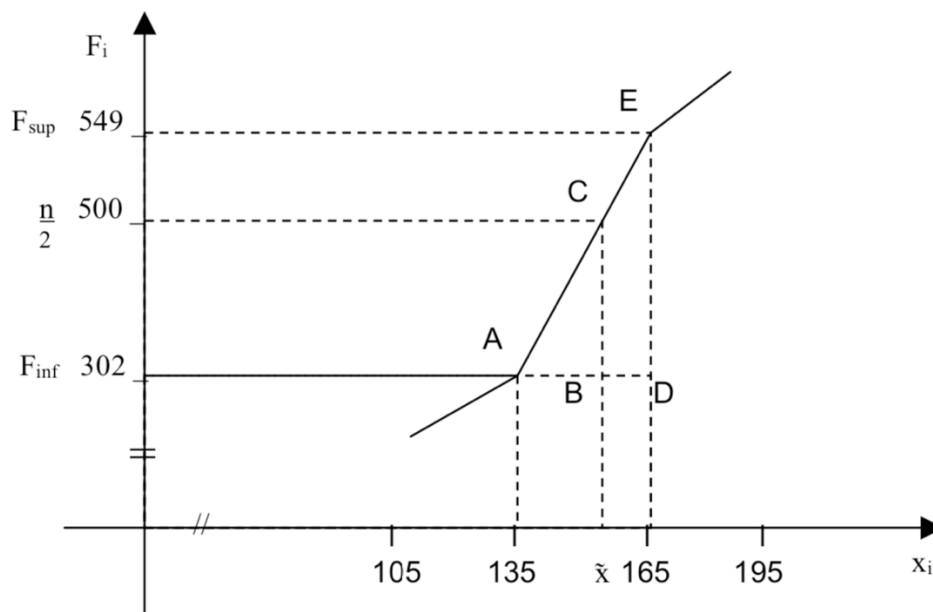
Siendo $\bar{x}$ la media calculada con anterioridad. Para el **desvío estándar**:

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

En nuestro Ejemplo 2, la media resultó ser $\bar{x} \cong 157,71$ y el desvío $s \cong 45,37$

La obtención de la mediana en ese ejemplo se efectúa de la siguiente forma: Localizamos en qué intervalo de clase se encuentran las $n/2$ observaciones. En nuestro ejemplo, las 500 observaciones están en el 5to. Intervalo (el anterior tiene acumuladas sólo 302).

Luego la mediana pertenece al intervalo [135;165). En el polígono de frecuencias acumuladas, individualizamos los triángulos



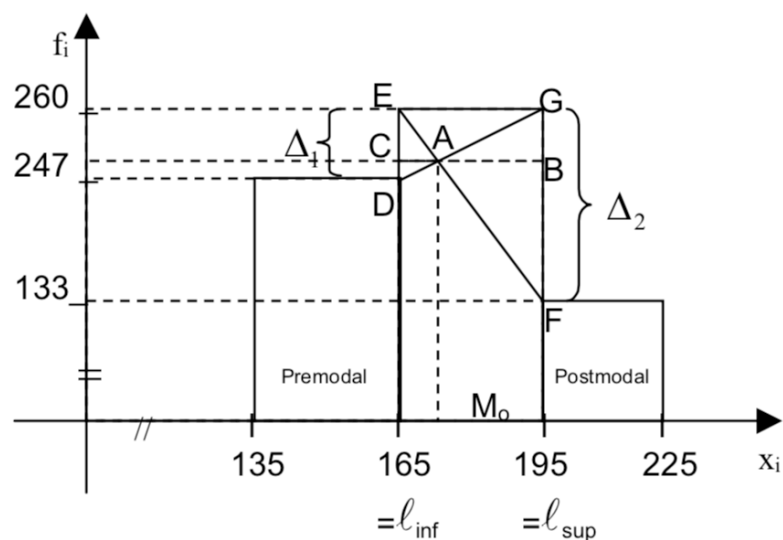
Los triángulos ABC y ADE son semejantes y por tanto tienen sus lados proporcionales.

$$\frac{AB}{BC} = \frac{AD}{DE} \Rightarrow \frac{\bar{x} - l_{\inf}}{\frac{n}{2} - F_{\inf}} = \frac{l_{\sup} - l_{\inf}}{F_{\sup} - F_{\inf}}$$

y de allí despejamos la **mediana**. En el ejemplo, vemos que

$$\frac{\bar{x} - 135}{500 - 302} = \frac{165 - 135}{549 - 302} \Rightarrow \bar{x} - 135 = \frac{30 \times 198}{247} \Rightarrow \bar{x} \cong 159,05$$

Para determinar el **modo** conviene ir al histograma de frecuencias absolutas y localizar el intervalo de mayor frecuencia.



En el ejemplo, es el intervalo [165;195) con 260 de frecuencia. Los triángulos DEA y FGA, que son también semejantes, nos permitirán calcular por proporcionalidad de lados la moda:

$$\frac{AB}{FG} = \frac{AC}{DE} \Rightarrow \frac{l_{\text{sup}} - Mo}{\Delta_{\text{post}}} = \frac{Mo - l_{\text{inf}}}{\Delta_{\text{pre}}}$$

Y se obtiene Mo

$$Mo = \frac{\Delta_{\text{post}} \cdot l_{\text{inf}} + \Delta_{\text{pre}} \cdot l_{\text{sup}}}{\Delta_{\text{post}} + \Delta_{\text{pre}}}$$

Con los datos del ejemplo

$$\frac{195 - Mo}{260 - 133} = \frac{Mo - 165}{260 - 247}$$

Y despejando se obtiene $Mo \cong 167,79$

Los cuartiles se encuentran de manera similar a la mediana. Para el primer cuartil, ubicamos en qué intervalo de clase se tiene acumulado las $n/4$ primeras observaciones y nuevamente hacemos proporcionalidad de lados, así en el ejemplo, $n/4=250$ que está en el $[105;135)$ y hacemos

$$\frac{q_1 - 105}{250 - 122} = \frac{135 - 105}{302 - 122}$$

$\Rightarrow q_1 \cong 126,33$. De igual forma para $q_3 \cong 188,19$