MEDIDAS DE POSICION Y DISPERSION

Hemos visto cómo ordenar y manejar la información obtenida y cómo tener ciertas ideas descriptivas acerca de la distribución de frecuencias lograda.

Debemos insistir en la necesidad de definir, del modo más riguroso posible, la población sobre la que se va a hacer el estudio y, por supuesto y como consecuencia, la muestra que va a extraerse de dicha población.

Si bien es cierto que una cifra lleva en sí misma, un rigor que excluye toda duda, todavía es necesario saber exactamente a qué se aplica y qué representa. Una cifra siempre es exacta pero puede no servir para el caso en estudio.

Habíamos visto, que, una vez obtenidos los datos, es necesario ordenarlos de alguna manera para que resalte la información buscada.

Si la variable (o variables) en estudio es discreta, bastará tomar los valores distintos "x_i" de la variable dentro de su dominio y la tabla y diagramas se construyen fácilmente.

Si el rango de definición de esta variable discreta es muy amplio, también pueden construirse intervalos pero, por supuesto, sin perder de vista la característica de discontinuidad de la variable. La representación gráfica es simplemente un diagrama de puntos o barras separadas.

Si la variable es continua, la cantidad de valores posibles es teóricamente infinita. Pensando en que siempre se hace un número limitado de observaciones, podría trabajarse igual que en el caso discreto (con valores distintos de la variable). Sin embargo, cuando las mediciones son muchas, la tabla así construida resultará muy extensa y se corre el riesgo de no proporcionar un conocimiento claro sobre la ley de distribución. Por eso, y para dar una idea más concreta acerca de la continuidad de la variable, se agrupan los datos en intervalos, cuyo número y amplitud dependen del problema. Es sabido que en estos casos se suele perder información pero se gana economía de trabajo, y, a veces es útil observar la forma de la distribución resultante.

Si la variable es cualitativa, el trabajo es mucho más simple.

En todos los casos mucho depende de la práctica y habilidad del investigador.

Datos de todos los tipos son difíciles de interpretar, tan sólo con tener la información volcada en tablas y gráficos.

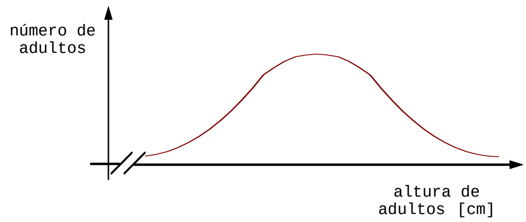
Las tablas y gráficos que contienen en general la totalidad de la información que se recoge, dan una idea a veces completa pero a menudo confusa del comportamiento de la población o muestra estudiada. Además, como generalmente el objetivo final es efectuar alguna comparación (inferencia) los datos obtenidos no son demasiado útiles o claros para tal fin.

Habíamos visto que tablas y gráficos proporcionan una o mas distribuciones de frecuencias. Estas deben ser examinadas cuidadosamente.

Existen distribuciones de distintas formas.

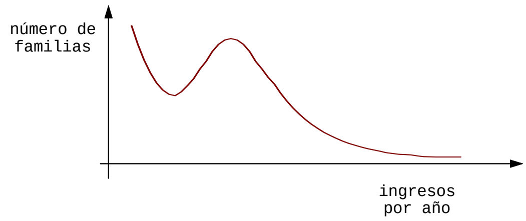
- muchas se caracterizan por ser "simétricas" y con un sólo "pico" y muestran una forma de campana con su altura máxima para los valores centrales de la variable disminuyendo gradualmente hacia los extremos, donde se observa un número reducido de valores.

Ejemplo:
distribución del número de adultos de acuerdo con sus alturas.

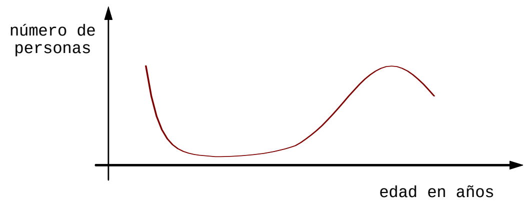


- otras distribuciones son "asimétricas" ya sea hacia la derecha o hacia la izquierda, otras presentan dos o más "picos" (son bimodales o plurimodales)

Ejemplo:
ingresos familiares por año.

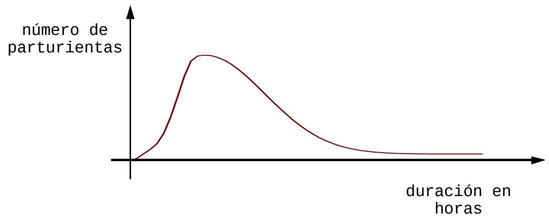


Ejemplo:
edad de las personas al morir.



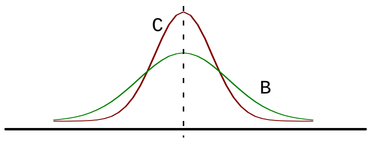
Ejemplo:

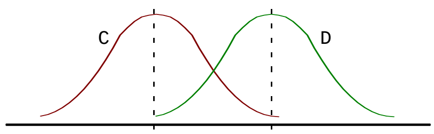
duración en horas del primer parto.



Retomemos el ejemplo de la distribución de alturas de adultos

Supongamos tener cuatro grupos de individuos.





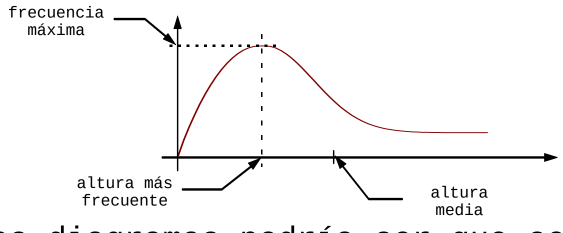
Observando los diagramas podemos decir que:

- Los grupos A y B presentan la misma altura "media" pero las alturas del grupo B se encuentran mas dispersas.(O lo que es lo mismo, menos concentradas)

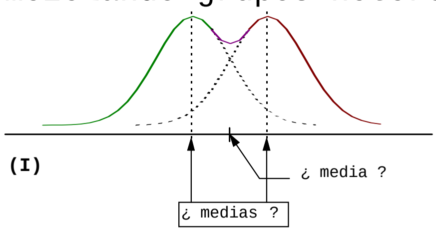
- Los grupos C y D tienen igual dispersión pero se "ubican" de distinta manera, es decir, las concentraciones de datos se corresponden con distintos valores de la variable.

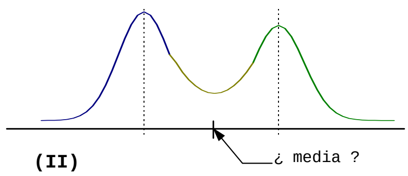
En general al decir "altura media" estamos usando una expresión convencional y no podemos asegurar aún, sin realmente comprobarlo, que esta "altura media" sea la "altura del mayor número de individuos".

Le sugerimos observar lo que sucede si la distribución es ligeramente asimétrica. Compare lo que observa en este caso con las situaciones anteriores.



En los siguientes diagramas podría ser que se estuvieran mezclando grupos heterogéneos.





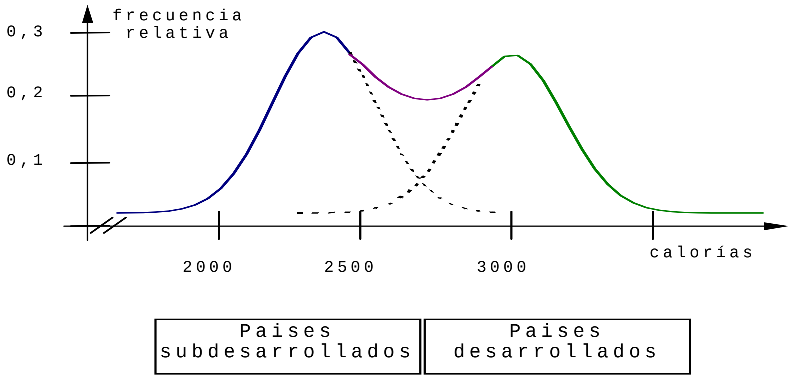
(1) ¿ Qué comentarios podría hacer a partir de los gráficos (I) y (II)?

Si en los esquemas tenemos representada una distribución de alturas, podremos observar que posiblemente en (I) se trate de dos distribuciones, con alturas medias diferentes y datos mezclados correspondientes a esas distribuciones escondidas; y en (II) se llega al extremo de que no hay individuos con altura media y es evidente que se trata de poblaciones heterogéneas, cada una con su propia **altura media**.

Por ejemplo:

Curva de distribución de las calorías disponibles por habitante y un día en 125 estados (1965-1967)

<u>Fuente:</u> anuario estadístico FAQ - 1969, pág.438 (obtenido de grupo Chedule - Pág. 67)



(2) Describa la forma de las distribuciones que obtendría si tuviera que representar gráficamente las siguientes situaciones:

a. el ingreso de las familias de Comodoro Rivadavia.
b. la mortalidad por cada 100 habitantes según la edad.
c. el diámetro de tornillos para determinado uso, producido por una máquina que funciona correctamente.
d. el tiempo de operación de determinada marca de filtro de arena para el tratamiento de residuos.
e. la temperatura de encendido de un interruptor controlado termostáticamente.

(3) Durante la última semana, los precios en australes de dos acciones negociadas en la bolsa de valores de Buenos Aires fueron

Día de la semana	**C**ompañía **A**	**C**ompañía **B**
Lunes	46	22
Martes	5	25
Miércoles	10	26
Jueves	34	24

Viernes	17	28

Compare las dos distribuciones. ¿Puede opinar respecto de ellas? ¿Cuál de las dos series de datos le parece mas confiable y por qué?

(4) Las longitudes en centímetros de determinados elementos de dos fabricantes distintos fueron

Longitudes (cm)	**F**abricantes	
	A	**B**
17 - 19	-	8.00
19 - 21	-	7.00
21 - 23	2.80	14.30
23 - 25	10.80	70.70
25 - 27	25.70	-
27 - 29	18.80	-
29 - 31	17.60	-
31 - 33	17.20	-
33 - 35	7.10	-
Total	100	100
Número de Individuos	642	1381

Compare ambas distribuciones y opine sobre su simetría, posición y dispersión.

REDUCCION DE LA INFORMACION

Será necesario y conveniente condensar, reducir o resumir los datos por medio de sólo unas cuantas medidas descriptivas para poder tener una información mas concreta sobre las distribuciones en estudio.

Debemos recordar:

- una medida descriptiva calculada a partir de los datos de una muestra recibe el nombre de **estadística** o **estadígrafo**
- una medida descriptiva calculada a partir de los datos de una población recibe el nombre de **parámetro**

Existen:

- medidas de posición o de tendencia central.
- medidas de dispersión.

La medida de posición se emplea para ubicar el **centro** de un conjunto de observaciones. Pero esto no es suficiente, pues los datos pueden estar diseminados, dispersos de cierta forma a ambos lados de ese **centro**. Esta diseminación es generalmente llamada **dispersión** o **variación**.

Si ésta es pequeña (o la concentración grande) los datos están muy concentrados y esto indica uniformidad en las observaciones.

Enumeraremos

- **medidas de posición**
  - media aritmética
  - media podada
  - media de Windsor
  - mediana
  - modo
  - cuartilos, decilos, centilos
    - semirango o rango medio

- **medidas de dispersión**
  - rango, recorrido o amplitud
  - rango intercuartil, rango interdecílico
  - desvío medio
  - desvío mediana
  - varianza
  - desvío estándar
  - coeficiente de variación
  - desviación intercuartílica

- **medidas de asimetría y apuntamiento**

Recordemos que:
Si a la variable aleatoria se la designa con letra mayúscula **X**, se tendrá que, para distinguir cada uno de los "n" valores de la variable que se designan con la letra minúscula **x** , habrá que agregar a la x un subíndice, un número natural, tal que su valor indica cuál es el valor de la variable.

Esto es x_1 es el primer valor observado de la variable, x_2 el segundo valor observado, ... , x_n el n-ésimo valor observado de la variable. En forma general, diremos que x_i es el i-ésimo valor observado de la variable.

Si no se tiene un número n de valores de la variable x sino que se tiene una población de tamaño N, el último valor será X_N.

Podemos también designar a la variable con la letra mayúscula Y, y entonces tendremos:

$$y_1,\ y_2,\ y_3,\ \dots\ ,\ y_i,\ \dots\ ,\ y_n \text{ (ó } Y_N)$$

Si una vez obtenidos los valores observados de la variable, se los ordena de menor a mayor se tiene:

$$x_{(1)},\ x_{(2)},\ x_{(3)},\ \dots\ ,\ x_{(N)}$$

Los subíndices de ambos listados pueden o no coincidir, pero recuerde que se interpreta por ejemplo, que x_2 es el segundo valor observado de la variable $x_{(2)}$ es el segundo valor de la variable luego de su ordenamiento de menor a mayor.

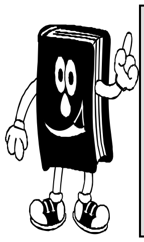
Procederemos ahora de la siguiente manera:

- Presentaremos las distintas medidas de posición y dispersión, definiendo, dando propiedades, haciendo comentarios que pueden resultar útiles para comprender mejor los conceptos, etc.

- Intercalaremos actividades, ejercitación que aumentará su complejidad gradualmente, ejercicios que se irán resolviendo en forma escalonada.

LEA detenidamente el material de consulta que requiera, para aclarar sus conceptos, a medida que lo vaya necesitando. Tiene bibliografía a su disposición.

No trate de memorizar fórmulas, puesto que no es nuestro objetivo, y sólo LÉALAS detenidamente, interpretándolas.



RECUERDE: El objeto de la **ESTADISTICA** no son las fórmulas, ni su objetivo es "adorarlas" u "odiarlas", sino tan solo (y nada más y nada menos) que **USARLAS** para resolver situaciones problemáticas. Esto es lo que pretendemos que usted aprenda a hacer.

MEDIDAS DE POSICION

Las medidas de posición vienen a resolver esos conceptos que aparecen intuitivamente cuando se trata de una característica de tendencia central que son la equi-repartición o centro de gravedad de la distribución, la posición central y la mayor frecuencia.

Las propiedades de una buena medida de posición son:

1. debe ser definida rigurosamente
2. si es posible, debe basarse en todas las observaciones
3. debe ser sencilla, clara
4. debe ser rápida para calcular
5. no debe estar afectada, dentro de lo posible, por fluctuaciones en el muestreo. Esto es, debe ser un **buen estimador** con valores **estables**

MEDIA ARITMÉTICA

Es el valor que tomaría la variable si estuviese uniformemente repartida entre todos los individuos que forman la muestra; se corresponde con el concepto de centro de gravedad en mecánica.

Se calcula simplemente sumando todos los valores de las observaciones y dividiendo por el número de datos considerados.

Si se tiene una muestra de tamaño n es **estimador.**

M(x) (se lee media de x (operador medio))

$$\boxed{M(x) = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{X}}$$

(se lee **x barra** o **x raya** o **x media**)

Si se tiene una población de tamaño N es **parámetro**

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

(se lee "**mu**")

Las dos expresiones anteriores corresponden a la media aritmética simple, es decir obtenida considerando a las observaciones individualmente, sin agrupar.

Si se tienen los datos agrupados en intervalos o por valores distintos de la variable, sea ésta discreta o continua, se calcula la media aritmética ponderada.

$$\boxed{\overline{X} = \frac{1}{\sum_{i=1}^{k} f_i} \cdot \sum_{i=1}^{k} x_i' \cdot f_i}$$

(muestral)

k : número de clases o categorías

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

$$\boxed{\mu = \frac{1}{\sum_{i=1}^{k} f_i} \cdot \sum_{i=1}^{k} x_i' \cdot f_i}$$

(poblacional)

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

MEDIA PODADA Y MEDIA DE WINDSOR

En ambos casos se procede de manera similar ordenando los datos de menor a mayor y encontrando los valores que separan grupos con el 25% de las observaciones en los extremos.

En la media podada se eliminan el 25% inferior y el 25% superior del listado ordenado y se calcula la media aritmética del 50% restante de las observaciones.

En la media de Windsor se reemplaza cada dato del 25% inferior por el valor de la menor observación no eliminada en la media podada y cada valor del 25% superior de los datos por el valor de la mayor observación no eliminada en la media podada.

En realidad, debemos aclarar que estas no son medidas demasiado usadas, sino que las mencionamos como una forma más de explicar características de las distribuciones, especialmente en los casos en que las distribuciones no son simétricas. No se mencionan, en general, en la bibliografía más frecuentemente usada.

MEDIANA

Es el valor central de la distribución, o sea, el valor del carácter a ambos lados del cual se reparten por mitades las observaciones.

MODO

Es el valor del carácter que se presenta con mayor frecuencia en la muestra o población, o sea, aquel al que le corresponde el mayor número de observaciones.

CUARTILOS, DECILOS, CENTILOS

Con el mismo sentido que la mediana, son los valores que dividen alas observaciones por cuartos, décimos y centésimos, respectivamente.(son 3, 9 y 99, respectivamente)

SEMI-RANGO

Es el valor "central" de los valores distintos de la variable, entre el valor mínimo y el valor máximo.

$$\boxed{\bar{R} = \frac{X_M + x_m}{2}}$$

Sigamos con las actividades.

(6) Analice las definiciones anteriores y recuerde los tipos de variable que conoce. Estará en condiciones de realizar la actividad que sigue. Con ésta información y su buen criterio, complete con una X en la siguiente tabla a doble entrada las celdas cuyo contenido considere posible. Cuando más adelante interprete ejercicios y realice otras actividades, tal vez le sea necesario retornar a esta tabla a fin de corregir y/o completar sus respuestas actuales.

Con variables de tipo se puede calcular	**M**edia **A**ritmética	**M**ediana	**M**odo
Cualitativa nominal			
Cualitativa ordinal			
Cuantitativa discreta			
Cuantitativa continua			

Tenga en cuenta que **las propiedades de la media aritmética** son:

1. **unicidad**: dado un conjunto de datos, existe una y sólo una media aritmética para él.
2. **simplicidad:** la media aritmética es fácil de comprender, calcular e interpretar.
3. intervienen en su cálculo **todos** los valores de la variable, por esta razón la media aritmética se ve afectada si existen valores extremos de la variable.

Tenga en cuenta que **las propiedades de la mediana** son:

1. **unicidad**.
2. **simplicidad**.
3. al no intervenir en su cálculo **todos** los valores de la variable, no es tan afectada por los valores extremos como lo es la media aritmética.

Como una relación entre las más importantes medidas de posición mencionadas, mencionamos que:

Karl Pearson, científico precursor de la estadística moderna en Inglaterra durante el siglo XIX encontró que:

$$\overline{X} - Mo = 3.(\overline{X} - M_{na})$$

LEER Recurra a la bibliografía, a fin de completar el desarrollo que sigue.

A continuación se enuncian las
propiedades de la media aritmética.

Esperamos que usted interprete cada una y luego...

(7) Verifique analíticamente estas propiedades y/o ejemplifique, según se indique en cada caso.

PROP.1.- La suma de los desvíos entre cualquier valor observado y el promedio es cero, cualquiera sea la distribución.

Es decir

$$\sum_{i=1}^{n}\left(x_i - \overline{x}\right) = 0 \qquad\qquad \sum_{i=1}^{k}\left(x_i - \overline{x}\right).f_i = 0$$

....Verifique y ejemplifique.

PROP.2.- La suma de los cuadrados de los desvíos entre cada valor observado y el promedio es menor o igual que la suma de los cuadrados de los desvíos con respecto a cualquier otro valor.

Es decir:

$$\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \leq \sum_{i=1}^{n}\left(x_i - A\right)^2$$

Se verifica la igualdad en el caso en que $\overline{x} = A$

....Ejemplifique.

PROP.3.- El promedio de un grupo de medias aritméticas, cada una de ellas ponderada por la cantidad de observaciones que le dio origen, coincide en el promedio de las observaciones individuales. Analicemos el siguiente caso. Si se tienen los siguientes grupos de datos:

x_{1i} = 3 - 4 - 8 - 3 - 7
x_{2i} = 10 - 12 - 11 - 13 - 14
x_{3i} = 7 - 3 - 9 - 8 - 3 - 1 - 4

y se quieren calcular las medias aritméticas, entonces se obtienen

$$\overline{x}_1 = 5 \quad n_1 = 5 \quad ; \quad \overline{x}_2 = 12 \quad n_2 = 5 \quad ; \quad \overline{x}_3 = 5 \quad n_3 = 7$$

Luego, designando con $\overline{\overline{x}}$ a la media de las medias tenemos:

$$\overline{\overline{x}} = \frac{\overline{x}_1.n_1 + \overline{x}_2.n_2 + \overline{x}_3.n_3}{.n_1 + n_2 + n_3} = \frac{25 + 60 + 35}{17} = 7{,}06$$

Por otra parte si consideramos ahora a los diecisiete (17) valores como provenientes de un sólo grupo, tendremos:

$$\overline{x}\ \text{grande} = \frac{\sum_{i=1}^{17} x_i}{17} = 7{,}06$$

con lo que se verifica que: $\overline{x}\ \text{grande} = \overline{\overline{x}} = \overline{X}$

Continuemos con el enunciado de las propiedades:

PROP.4.- La media aritmética de una constante es la misma constante.

PROP.5.- La media aritmética de una variable mas (o menos) una constante es igual a la media aritmética de la variable mas (o menos) la constante.

Si $y_i = x_i + c \quad \forall i$ desde 1 hasta n,

$\Rightarrow \quad M(y) = M(x \pm c) = \overline{x} \pm c$

PROP.6.- La media aritmética de una variable por una constante es igual a la media aritmética de la variable multiplicada por la constante.

Si $y_i = x_i \,.\, c \quad \forall i$ desde 1 hasta n,

$\Rightarrow \quad M(y) = M(c\,.\,x) = c\,.\,M(x)$

PROP.7.- La media aritmética de una variable que es suma (o resta) de otras variables originales es igual a la suma (o resta) de las medias aritméticas de las variables originales.

Si $y_i = x_{1i} \pm x_{2i}$, $\Rightarrow$ $M(y) = M(x_1) \pm M(x_2)$

es decir $\boxed{\overline{y} = \overline{x}_1 \pm \overline{x}_2}$

Por ejemplo
si se tiene que:

x_{1i} = 3, 9, 7, 5, 16, n_1 = 5 $\bar{x}_1 = 8$

x_{2i} = 13, 24, 32, 12, 19, n_2 = 5 $\bar{x}_2 = 20$

y se calculan los valores de una nueva variable

$\boxed{y_i = x_{1i} + x_{2i}}$, se tiene :

y_i = 16, 33, 39, 17, 35, n = 5, $\Rightarrow$ $\bar{y} = 28$

con lo que se confirma que: $\boxed{\bar{y} = \bar{x}_1 + \bar{x}_2}$

RECUERDE... Recurra a la bibliografía

Realizaremos ahora el análisis y discusión de ejemplos, y a medida que los desarrollemos completaremos algunos conceptos, aprendiendo a manejar la información numérica.

Tomaremos al comenzar ejemplos muy elementales.

(8)....a) Supongamos que se tiene una variable que toma los siguientes valores: 3, 5, 7, 4, 2 Se tiene entonces que el **tamaño de la población** es N = 5 . La **media aritmética** es:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{3+5+7+4+2}{5} = 4{,}2$$

Es decir, 4,2 es la media aritmética de los cinco valores de la variable.

Para calcular la **mediana** se debe, en primer lugar, ordenar los valores a fin de encontrar aquella observación que deja el mismo número de observaciones a su izquierda que a su derecha. En nuestro caso se tiene:

$$\underbrace{2 \qquad 3}_{\text{2 obs}} \quad \underset{\text{Mediana}}{\overset{4}{\uparrow}} \quad \underbrace{5 \qquad 7}_{\text{2 obs}}$$

3 obs   3 obs

La mediana es cuatro (4) y es el valor de la variable que deja dos observaciones tanto a su derecha como a su izquierda.

Es el valor de la variable que ocupa la posición central, en este caso la tercera posición. Notemos que esencialmente la mediana "divide" los datos en dos subconjuntos.

...¿Cómo determinar la posición de la mediana o el número de orden que ella ocupa?

Calculando lo que llamamos "**mediana de orden**":

M°=(N+1)/2,

donde 1 es la posición que ocupa el dato con el valor mínimo, y N es la posición que ocupa el dato con el valor máximo. Evidentemente la posición central está bien calculada.

En este ejemplo la mediana ocupa el tercer lugar, esto es **"el orden de la mediana"** es 3. O dicho de otra manera, "la mediana de orden" es 3 y se denota $Mna^0 = 3$

¿Cómo lo calculamos?

Haciendo $\frac{N+1}{2}$ en este caso $M^0 = 3$

y $\Rightarrow$ es $Mna = x_{(3)} = 4$

Al ser N un número impar será exactamente la mediana el valor central del conjunto. Para encontrarlo se puede comenzar a contar a partir del primero o del último de los datos ordenados.(Conviene contar a partir de los dos extremos, para verificar)

Recuerde que:

Cuando el subíndice se encuentra entre paréntesis se trata de valores ordenados de la variable.

Observando las frecuencias con que se dan los valores de la variable se ve que no existe **modo.**

....b) Supóngase que se tiene una **muestra de tamaño "n = 6"** de una población y que los valores observados de la variable aleatoria son: 41 , 36 , 17 , 28 , 25 , 39

Se tiene la **media aritmética** que es:

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{41+36+17+28+25+39}{6} = 31$$

La **mediana** no será ahora un valor observado de la variable. Es evidente que la posición de la mediana no es un entero si n es par.

Si ordenamos tendremos:

$$\underbrace{17 \quad 25 \quad 18}_{3 \text{ obs.}} \quad ; \quad \underbrace{36 \quad 39 \quad 41}_{3 \text{ obs.}}$$

$$\text{Mna} = \frac{28+36}{2} = 32$$

Mna = 32 es un valor no observado de la variable que separa las observaciones ordenadas dejando tres valores a su izquierda y tres a su derecha.

La mediana de orden en este caso es $\frac{n+1}{2} = \frac{7}{2} = 3{,}5$ lo que indica que la mediana se encuentra entre los datos ordenados tercero y cuarto, justamente en la mitad. Esto hace que se considere el valor de la mediana como la semi-suma entre la tercera y la cuarta observación:

$$\text{Mna}^{0} = 3{,}5 \qquad \Rightarrow \qquad \text{Mna} = \frac{x_{(3)} + x_{(4)}}{2} = 32$$

El **modo,** al igual que en el ejemplo anterior, no existe. ¿o sí?

....c)supongamos ahora tener una variable que toma los valores:

3, 5, 9, 16, 3, 5, 5, 5, 7, 9, 4, 5, 13

La **media aritmética** será:

$$\mu = \frac{3.2+4.1+5.5+7.1+9.2+13.1+16.1}{13} = \frac{89}{13} = 6{,}846$$

Para calcular la mediana de orden, tenemos en cuenta que:

$$N = 13, \quad \text{entonces} \quad Mna° = 7$$

luego la **mediana** será la séptima observación. Ordenando los valores observados tenemos:

$$\underbrace{3 \quad 3 \quad 4 \quad 5 \quad 5 \quad 5}_{\text{6 obs.}} \quad \underset{\text{Mediana}}{\boxed{5}} \quad \underbrace{5 \quad 7 \quad 9 \quad 9 \quad 13 \quad 16}_{\text{6 obs.}}$$

En este caso existe el **modo**; si se observan las frecuencias con que aparecen los valores de la variable se ve que el valor 5 es el que mas veces se repite, $f_{(5)} = 5$, entonces $M_0 = 5$

Además los valores 3 y 9 también son modos puesto que sus frecuencias son mayores que para los otros valores inmediatos de la variable. Consideraremos entonces $M_0 = 5$ como el **modo absoluto**.

En el ejemplo Nº. 10, se verá cómo calcular la mediana y el modo a través de fórmulas que no hacen mas que formalizar los conceptos empleados hasta aquí.

En este momento está en condiciones de realizar la actividad que le proponemos.

(9) Tome nuevamente los datos del ejercicio (8) c) y:

1. construya una tabla de frecuencias y complete el análisis utilizándola.
   Ponga columnas para x_i, f_i, $x_i f_i$.
   Compare con lo logrado en el item(8) c).
2. Represente gráficamente x_i versus f_i considerando a la variable como discreta.
3. Agregue en la tabla del item (1) una columna para la frecuencia acumulada (F_i).
4. Realice un gráfico de x_i versus F_i.
5. Compare los gráficos de (2) y(4) y encuentre en ellos la mediana y el modo.
6. Calcule μ empleando la expresión de la media aritmética ponderada, cuyo numerador logrará sumando los términos de la columna $x_i f_i$ que encontró en el punto (1).

EJEMPLO PARA ANALIZAR (resuelto)

Veamos un ejemplo más complejo y analicémoslo

(10)*Los siguientes datos son los tiempos de duración de ensayos de laboratorio (en centésimas de segundo)*

4,58	2,51	4,04	6,43	1,58	4,32	2,20	4,19
4,79	6,20	1,52	4,38	3,87	4,54	5,12	5,15
5,50	5,92	4,56	2,46	6,90	1,47	2,11	2,32
6,75	5,84	8,80	7,40	4,72	8,62	2,46	8,75
2,65	7,86	4,71	6,25	9,45	12,80	1,42	1,92
7,60	8,79	5,92	9,65	5,09	4,11	6,37	5,40
11,25	3,90	5,33	8,64	7,41	7,95	10,60	3,81
3,78	3,75	3,10	6,43	1,70	6,40	3,24	1,79
4,90	3,49	6,77	5,62	9,70	5,11	4,50	4,80
5,21	1,76	9,20	1,20	6,85	2,80	7,35	11,75

Dado que la variable es continua(pero no necesariamente por ello), hay gran cantidad de valores y estos son bastante diferentes; entonces se los agrupa en intervalos. Se podría trabajar con los datos y con el auxilio de una calculadora programable o una computadora, lo que haría al trabajo muy sencillo y nos permitiría en pocos minutos tener calculadas todas las medidas de posición, pero este no es el objetivo del ejercicio. Queremos analizarlo en detalle para comprender el mecanismo a seguir y los por qué de cada paso.

Tenga presente además que en muchos casos no se tiene acceso a los valores originales sino a tablas ya construidas.

Se construyen entonces los intervalos, sin olvidar que al hacerlo se pierde la información original.

Cómo se tienen n = 80 valores, uno puede preguntarse ¿cuántos intervalos se construyen?

Una idea de cuántos emplear la proporciona la fórmula de Sturges

$$2^{k-1} < n \leq 2^{k}$$

$$\underset{(64)}{2^6} < 80 < \underset{(128)}{2^7}$$

en consecuencia se construyen 7 intervalos (esta es una forma de encontrar el número de intervalos, existen otras)

Aceptemos en este caso sin discusión el hacer 7 intervalos.

Nos preguntamos entonces:
¿Cuáles podrían ser los extremos de los intervalos?

Se considera:

$$x_{min} = 1{,}20 \qquad x_{max} = 12{,}8$$

$$R = x_{max} - x_{min} \qquad R = 12{,}80 - 1{,}20 \qquad R = 11{,}60$$

(R = rango o diferencia entre valores máximo y mínimo)
Corregimos el rango de tal forma que al dividir por el número de intervalos obtengamos un valor que será usado como amplitud de cada intervalo, esto es:

Calculamos el rango corregido (R’) tal que R’/K = c

$$\frac{11{,}60+\overbrace{0{,}02}^{\text{Corrección}}}{7} = \frac{11{,}62}{7} = 1{,}66 \quad \text{Amplitud de cada intervalo}$$

Por lo tanto los intervalos pueden ser:

1,20	a menos de	2,86
2,86	a menos de	4,52
4,52	a menos de	6,18
6,18	a menos de	7,84
7,84	a menos de	9,50
9,50	a menos de	11,16
11,16	a menos de	12,82

Al completar la tabla con los f_i y F_i se sabrá cuántos valores cayeron en cada intervalo pero no cuáles son.

Intervalos	**C**onteo	f_i	F_i	x_i'	$x_i' f_i$	$x_i'^2 f_i$
1,20 - 2,86		18	18	2,03	36,54	4,1762
2,86 - 4,52		14	^{3} 32	3,69	51,66	190,6254
4,52 - 6,18		20	52	5,35	107,00	572,45
6,18 - 7,84		13	65	7,01	91,13	638,8213
7,84 - 9,50		9	74	8,67	78,03	676,5201
9,50 - 11,16		3	77	10,33	30,99	320,1267
11,16 - 12,82		3	80	11,99	35,97	431,2803
TOTAL		80			431,32	2904,0000

$$x_i' = \frac{x_{inf} + x_{sup}}{2}$$

Se pueden calcular las medidas de posición.

- **Media aritmética**:

$$\bar{x} = \frac{\sum x_i^{'}.f_i}{\sum f_i}$$

Donde $x_i^{'}$ es la **marca de clase** o **punto medio** del intervalo y se lo toma como representante del intervalo pues estamos bajo el supuesto de que los datos se distribuyen en forma homogénea dentro de cada intervalo.

Para este cálculo se construye la columna ($x_i^{'}.f_i$) y se tiene

$$\bar{x} = \frac{431,32}{80} \qquad \bar{x} = 5,3915$$ centésimos de segundo.

Lo que significa que el tiempo medio de duración de los ensayos de laboratorio muestreados, es de 5,39 centésimos de segundo.

- **Modo**:

Al tener la variable agrupada en intervalos se halla primeramente el intervalo modal, es decir, el intervalo de mayor frecuencia absoluta.

En este caso es el intervalo 4,52 a menos de 6,18 centésimos de segundo, cuya frecuencia absoluta es 20.

Si se quiere hallar un valor dentro de tal intervalo como valor modal se utiliza la fórmula:

$$M_0 = L_i + \left[\frac{f_{M_0} - f_{M_0-1}}{\left(f_{M_0} - f_{M_0-1}\right) + \left(f_{M_0} - f_{M_0+1}\right)} \right].c$$

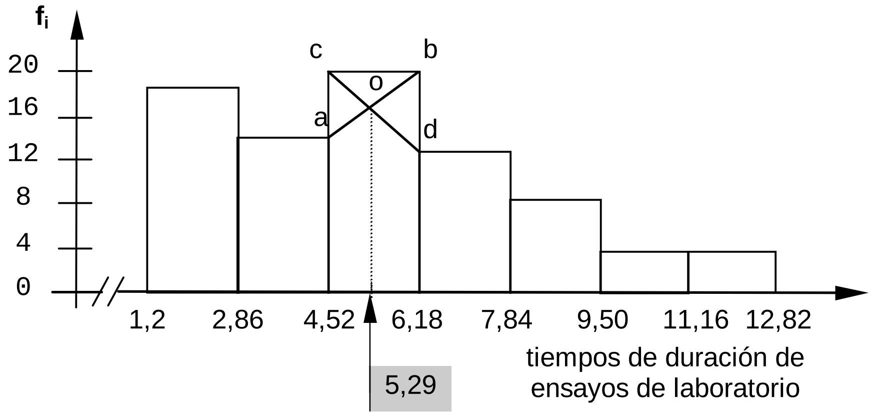
Donde:
- L_i límite inferior del intervalo modal
- f_{M_0} frecuencia absoluta del intervalo modal
- f_{M_0-1} frecuencia absoluta del intervalo anterior al modal
- f_{M_0+1} frecuencia absoluta del intervalo posterior al modal
- c amplitud de los intervalos

Se tiene entonces que:

$$M_0 = 4,52 + \left[\frac{20-14}{(20-14)+(20-13)} \right].1,66 = 5,285153846$$

Lo anterior se interpreta diciendo que el tiempo de duración que más se repite en las 80 observaciones para dichos ensayos de laboratorio considerados es 5,29 centésimos de segundo.

Si se representa la distribución de frecuencias absolutas en escala es posible gráficamente (y en forma aproximada) hallar este valor.



Para hallar el modo gráficamente se ubica el intervalo modal y luego...

1. Se traza un segmento desde la altura del extremo derecho del intervalo anterior al modal a la altura del extremo derecho del modal (segmento ab).
2. Se traza un segmento desde la altura del extremo izquierdo del intervalo posterior al modal a la altura del extremo izquierdo del intervalo modal (segmento cd).
3. Se traza una perpendicular por la intersección de dichos segmentos al eje de abscisas.
4. El punto de intersección de la perpendicular con el eje de la variable es el valor del modo.

Interrumpimos el desarrollo del ejemplo para plantearle una actividad.

(11) La forma explicada para hallar el modo gráficamente, no es mágica. Intente demostrarla aplicando el criterio de la semejanza entre los triángulos aoc y bod, o emplee algún otro criterio.

....Continuamos con el ejemplo (10).

Observando la distribución a ambos lados de la frecuencia modal, se advierte que podría haber otros intervalos en situación similar a este.

Si puede considerar que la frecuencia del anterior al primero es 0 y la frecuencia del posterior es 14 verá entonces que ese primer intervalo puede ser considerado intervalo modal. Se considera al mayor de los modos: **modo principal** y a los restantes **modos secundarios**.

Usted está en condiciones de hallar el valor aproximado del modo secundario. Hágalo.

(12) Encuentre el modo secundario en forma mas precisa usando la fórmula.

- **Mediana:**

En forma similar a la usada para hallar el modo se debe hallar primeramente el intervalo mediana hasta el cual se acumula, por lo menos, la mitad de las observaciones.

En este caso

$$Mna^0 = \frac{n+1}{2} \qquad\qquad Mna^0 = 40{,}5$$

Por lo tanto el intervalo mediana va desde 4,52 a menos de 6,18.

Para hallar el valor de la mediana dentro de dicho intervalo se aplica la fórmula:

Mna^0

$$Mna = L_i + \left[\frac{\left(\frac{n+1}{2}\right) - F_{i-1}}{\left(F_i - F_{i-1}\right)}\right].c$$

Donde: L_i límite inferior del intervalo mediana
F_i frecuencia acumulada hasta el intervalo mediana
F_{i-1} frecuencia acumulada hasta el intervalo anterior a la mediana
c amplitud del intervalo

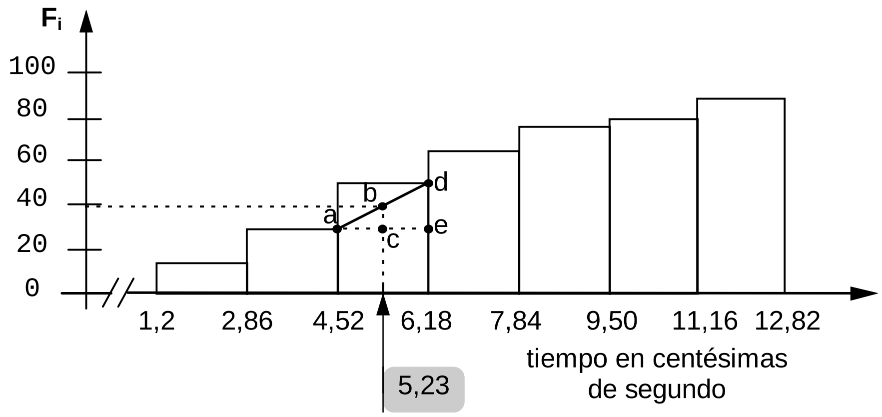
Se tiene:

$$\text{Mna} = 4{,}52 + \left[\frac{40{,}5 - 32}{20}\right].1{,}66$$

Mna = 5,2255 centésimos de segundo

Significa que antes y después de 5,2255 centésimos de segundos queda la misma cantidad de observaciones. Esto es que el área total encerrada por el polígono queda dividida en 2 partes iguales.

Dado que al trabajar con el concepto de mediana se utilizan las frecuencias acumuladas, si se quiere representar la mediana gráficamente, se debe utilizar igualmente el polígono de frecuencias acumuladas u ojivas.



Para hallar la mediana por el modo gráfico:

1. Se traza una paralela al eje horizontal por la mediana de orden hasta intersectarse con la ojiva.
2. Desde ese punto de intersección se traza una perpendicular y donde ésta corta al eje horizontal se encuentra la mediana.

¿ Puede usted demostrar la fórmula de la mediana comparando los triángulos acb y aed?

(13) Hágalo.

NOTA : recuerde lo que se señaló en la página 71

(14) Empleando el mismo procedimiento se pueden calcular los cuartilos Q_1 y Q_3

- ¿Cuánto valen?
- Verifíquelo gráficamente.
- Calcule el tercer decilo.

PROBLEMAS – PREGUNTAS – EJERCICIOS (a resolver)

(15) Encuentre media aritmética, mediana y modo en cada uno de los siguientes conjuntos de datos

13	6	4	8	5	9	10
4	3	4	4	5	6	6
7	7	7	7	7	7	7

a) Considere variable continua.
b) Considere variable discreta.

(16) a) Dados

4	6	2	3	5
3	2	4	22	5

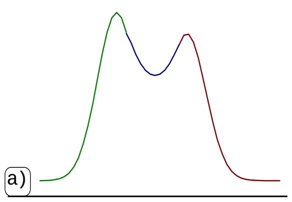
Sin calcular expresar cuál de los grupos tendrá una menor media aritmética y por qué. Calcule, compruebe e interprete.

b) Dados

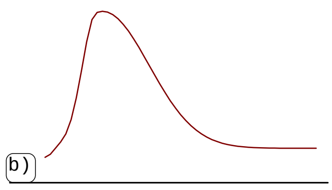
3	4	3	10	5	3	4
6	5	3	7	5	5	4

Si tuviera que elegir en cada caso entre la media aritmética y la mediana ¿Cuál elegiría ? ¿Por qué?

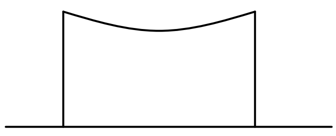
(17) Coloque en forma aproximada las posiciones de la media aritmética, la mediana y el modo en los siguientes gráficos.



a)

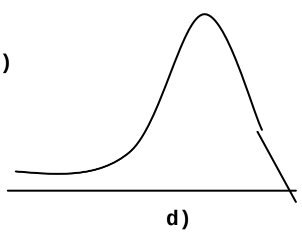


b)

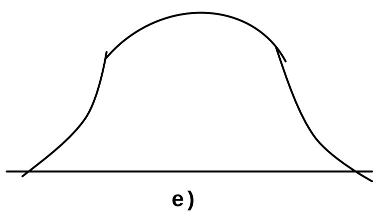


c)

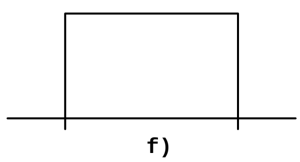
c)



d)



e)



f)

(18) ¿Cuál es el efecto que produce sobre la media aritmética

a) adicionarle 25 a cada valor
b) incrementar cada valor en un 25 %
c) dividir cada valor por 5

(19) Dados los siguientes datos

13	7	7	8	4	8	5	3	8	7	5	5
4	2	1	5	1	3	16	23	42	8	7	9
35	8	9	10	32	7	8	6	7			

que corresponden a los tiempos (en días) que han estado detenidas por reparaciones ciertas maquinarias.

a) Calcule media aritmética
b) Calcule media podada
c) Calcule media de Windsor
d) Compare e interprete la información brindada por las tres medias.

(20) Al informar sobre la vida útil de los neumáticos que se fabrican en las dos grandes fábricas de la ciudad, ambos empresarios coinciden en señalar que dicha vida útil es de 15000Km.

En realidad cuando se obtiene una muestra al azar de los neumáticos de cada fábrica se obtienen los siguientes datos:

Para un fabricante:

11 11 11 12 12 15 15 15 15 15 18 18 19 19 19

Para otro fabricante:

8 9 10 11 11 11 12 12 15 15 15 15 18 24 24

¿A qué promedio se referiría cada empresario ¿Cuál información cree usted que lo deja mas tranquilo, la de uno o la de otro?

(21) Dados los siguientes grupos de datos ordenados

3	5	6	9	10	10	13
3	5	6	9	10	10	41

Se tiene que la media aritmética del primer grupo es

x(gr1)=8

y la del segundo grupo es

x(gr2)=12.

Es evidente que éste segundo "promedio" no es un buen **informador** sobre la posición de la distribución ¿Por qué sucede ésto?

(22) En el ejercicio anterior se observa que tanto la mediana como el modo coinciden

$M_{na}(gr1)= 9$ $M_{O}(gr1)= 10$

$M_{na}(gr2)= 9$ $M_{O}(gr2)= 10$

¿Son lógicos éstos resultados? ¿Por qué?

(23) Presente un ejemplo en el que la medida mas adecuada sea
a) la media aritmética
b) la mediana
c) el modo

(24) Dados

3 8 30 10 12 5 5 4 5 9 14 10 2

a) Comprobar que la suma de los datos con respecto al promedio es cero y con respecto a la mediana y el modo no lo son.
b) ¿Por qué sucede esto?
c) ¿Cree usted qué esto es siempre así?

(25) A los 90 jóvenes del último año del Nacional "Chubut" se les pidió que elijan el atuendo para lucir en su fiesta de graduación.

¿Cuál medida de tendencia central resultaría mas adecuada para determinar la preferencia de atuendo del mayor número de jóvenes? ¿Por qué?

(26) Para cada una de las siguientes distribuciones diga si es posible calcular la media aritmética la mediana y/o el modo. Explique su respuesta

Longitud	Frecuencia
40 - 49	5
50 - 59	18
60 - 69	27
70 - 79	15
80 - 89	6

C.I.	Frecuencia
menos de 90	3
90 - 99	14
100 - 109	22
110 - 119	19
mas de 119	7

Peso	Frecuencia
100 o menos	41
101 - 110	13
111 - 120	8

121 - 130	3
131 - 140	1

(27) Retome las actividades 16) a), 16) b) y 20) .Calcule las medidas de posición

(28) En una máquina trabaja un elemento de corta vida en lo que respecta a su rendimiento dentro de las especificaciones fijadas. Para la operación se utilizan repuestos de la Fábrica **A** y de la Fábrica **B** .

De los registros se obtienen los datos del siguiente cuadro, donde encontramos la vida en horas de trabajo de 110 repuestos de **A** y de 77 repuestos de **B**.

Vida en horas	**A**	**B**
9 - 12	3	-
12 - 15	8	3
15 - 18	24	18
18 - 21	34	23
21 - 24	19	17
24 - 30	11	10
30 - 33	7	4
33 - 36	-	1
	110	77

Halle medidas de posición para cada tipo de repuestos **A** y **B**
Analice los resultados.

RECUERDE que estos problemas (o alguno de ellos) pueden resolverse con la PC.
En el "Bloque Anexo" encontrará algunas soluciones logradas con alguno de los paquetes con que se cuenta en el Laboratorio de Estadística, y que están a su disposición para que consulte o directamente emplee con las PC del Laboratorio en horarios a convenir.
Puede usted también trabajar con alguna calculadora. Pero tenga en cuenta que se espera que razone el mecanismo con el cual se trabaja y que conozca la teoría que fundamenta el procedimiento.