ESTADÍSTICA DESCRIPTIVA

Métodos Estadísticos Aplicados a las Auditorías Sociolaborales

Francisco Álvarez González
francisco.alvarez@uca.es

Bajo el término "Estadística Descriptiva" se engloban las técnicas que nos permitirán realizar un análisis elemental de las observaciones experimentales observadas.

Se subdivide en dos bloques :

1º **Estadística primaria :** Obtenido un grupo de observaciones experimentales, este apartado nos enseña a ordenarlas adecuadamente, de modo que se ofrezca una información lo más clara posible.

2º **Estadística derivada o secundaria :** Con los datos observados realizaremos ciertos cálculos, obteniendo así unas medidas. Este bloque temático nos enseña a interpretarlas.

PROCEDIMIENTO A SEGUIR EN UN ESTUDIO ESTADÍSTICO.

El proceso seguido en el estudio estadístico de una cierta característica o variable, puede subdividirse en tres pasos sucesivos :

A **RECOGIDA DE DATOS :**
Planteado el test o encuesta oportuno y recogidos los datos que correspondan, el primer análisis que realizaremos es el del tipo de variable que pretendemos estudiar (Cualitativa o Cuantitativa ; Discreta o Continua). Esto condicionará en gran medida su posterior tratamiento.

B **ORGANIZACIÓN DE LOS DATOS :**
Determinado el modo de agrupamiento de las observaciones, procedemos a su recuento, construyendo la tabla de frecuencias. Posteriormente podremos visualizar tales frecuencias de forma gráfica con el diagrama estadístico apropiado.

C **ANÁLISIS FINAL :**
La obtención de muy diversas conclusiones respecto de la variable estudiada, se podrá realizar con auxilio de los diferentes parámetros estadísticos (de centralización , posición , dispersión , etc.)

VARIABLES ESTADÍSTICAS. CLASIFICACIÓN.

El aspecto que deseamos estudiar (edad, sexo, peso, ...) recibe el nombre de VARIABLE ESTADÍSTICA. A lo largo de esta unidad observaremos, que las técnicas estadísticas a seguir serán diferentes según el tipo de variable objeto de estudio.

La clasificación más tradicional de las variables estadísticas es la siguiente :

CUALITATIVAS

Los valores de las observaciones quedan expresados por características o atributos.

Por ejemplo : Estado civil ; Color preferido ; Nivel de estudios ; Raza ; ...

Dentro de ellas podremos subdividirlas en función de que puedan ser ordenadas (Nivel de estudios) o no tenga sentido una determinada ordenación que se establezca (Color preferido, Razas, ...).

CUANTITATIVAS

Los valores de las observaciones son numéricos (cuantificables) y, en consecuencia, ordenables.
A su vez las variables cuantitativas se subdividen en dos tipos :

DISCRETAS :
Toman valores concretos (Nº de hijos : 0, 1, 2, ...)

CONTINUAS :
Pueden tomar cualquier valor de un cierto intervalo (Peso ; Estatura ; ...).

TABLAS DE FRECUENCIAS.

Si la variable es **Cualitativa**, observamos los valores diferentes de la misma.
Si es **Cuantitativa** buscaremos los valores ***mínimo*** y ***máximo*** obtenidos. En función del número de observaciones, decidiremos si se realiza su estudio de forma **individual** o **agrupando en intervalos**.

CONSTRUCCIÓN DE INTERVALOS :

Teniendo en cuenta la amplitud total de las observaciones (Valor máximo menos valor mínimo observados), tomaremos una decisión sobre el número total de intervalos, o bien sobre la amplitud o tamaño de los mismos.

EJEMPLO :
Supuesto : Valor máximo = 87 , Valor mínimo = 11 . Luego : AMPLITUD = 87 - 11 = 76.

Si decidimos construir 8 intervalos, la amplitud de cada uno será de 10 unidades (valor aproximado de 76/8). El primer intervalo no tiene porqué iniciarse en 11 (mínimo); es más, se aconseja tomar siempre valores "visualmente agradables" (5, 10, 15 ,...).
Con esto los intervalos serían :
[10,20) [20,30) [30,40) [40,50) [50,60) [60,70) [70,80) [80,90]

Si partimos de la decisión de que los intervalos tengan 15 unidades de amplitud, simplemente iniciaremos su construcción hasta llegar a un intervalo que contenga al valor máximo observado.
[10,25) [25,40) [40,55) [55,70) [70,85) [85,90]

Teóricamente se establece que el número ideal de intervalos debe ser la raíz cuadrada del número de observaciones disponibles :

Para **N** observaciones : Criterio de Kaiser **Nº de intervalos** $\approx \sqrt{N}$

Criterio de Sturges **Nº de intervalos** $\approx E\left(1'5 + 3'3\ln(N)\right)$ (E = parte entera)

NOTACIÓN

Al establecer dos intervalos consecutivos, por ejemplo de 10 a 20 y de 20 a 30, hemos de decidir si el valor 20 (final de uno e inicio del siguiente) pertenece al primer intervalo o al segundo. Para ello empleamos los símbolos **[** y **(** .

[o] el valor situado junto a él pertenece al intervalo
(o) el valor situado junto a él no pertenece al intervalo

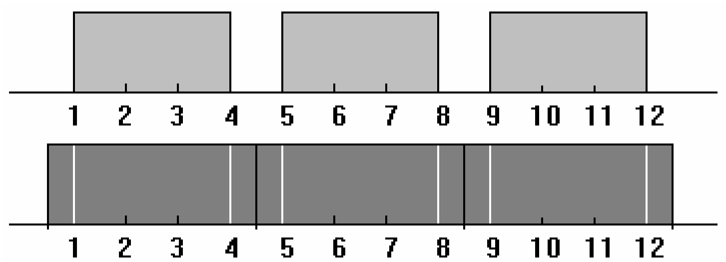
NOTACIONES PARA REPRESENTAR INTERVALOS

EXTREMOS REALES

Desde 0 hasta menos de 10	**[0 , 10)**
De 10 a menos de 20	**[10 , 20)**
De 20 a menos de 30	**[20 , 30)**
De 30 a menos de 40	**[30 , 40)**
Desde 40 hasta 50	**[40 , 50]**

EXTREMOS APARENTES

1 - 4	Valores : 1, 2, 3 y 4	**[0'5 , 4'5)**
5 - 8	Valores : 5, 6, 7 y 8	**[4'5 , 8'5)**
9 - 12	Valores : 9, 10, 11 y 12	**[8'5 , 12'5]**



RECUENTO. TABLA DE FRECUENCIAS ABSOLUTAS.

Situados en una tabla los valores de la variable (desde el mínimo al máximo) o los intervalos que los contienen, procedemos a contar las veces que se repiten. Construimos así una tabla como la de la izquierda. En ella podrá observarse que, en el supuesto de datos agrupados en intervalos, se ha incluido una columna encabezada por **x** . Tal valor de **x** se denomina **marca de clase** y es el valor central de cada intervalo.

Intervalos	**x**	**Recuento**	**n**	**N**
$[e_1, e_2)$	x_1	///	n_1	n_1
$[e_2, e_3)$	x_2	///// ///// /	n_2	n_1+n_2
. . .	. . .	. . .	. . .	. . .
$[e_i, e_{i+1})$	x_i	///// ///	n_i	$n_1+n_2+ ... +n_i$
. . .	. . .	. . .	. . .	. . .
			$\Sigma n_i = N$	

FRECUENCIAS.

FRECUENCIA ABSOLUTA (n) :

Para datos no agrupados en intervalos, es el número de veces que se presenta cada valor de la variable. Si los datos se agrupan en intervalos, es el número de observaciones que pertenecen a dicho intervalo.

FRECUENCIA ABSOLUTA ACUMULADA (N) :

Para un cierto valor de la variable, la frecuencia absoluta acumulada nos da el número de observaciones menores o iguales que dicho valor.

OTRAS FRECUENCIAS :

FRECUENCIA RELATIVA (r) :

Cociente entre la frecuencia absoluta y el número total de observaciones (N).

PROPORCIÓN o PORCENTAJE (p) :

Frecuencia relativa multiplicada por 100 (es la expresión de las frecuencias en %).

De igual modo que se definió para las frecuencias absolutas, se definen las **FRECUENCIAS RELATIVAS ACUMULADAS (R)** y los **PORCENTAJES ACUMULADOS (P)**.

TABLA COMPLETA DE FRECUENCIAS :

x	n	r	p	N	R	P
x_1	n_1	$r_1 = n_1 / N$	$p_1 = r_1 \cdot 100$	n_1	r_1	p_1
x_2	n_2	$r_2 = n_2 / N$	$p_2 = r_2 \cdot 100$	n_1+n_2	r_1+r_2	p_1+p_2
...	...	...	...	...	...	...
x_i	n_i	$r_i = n_i / N$	$p_i = r_i \cdot 100$	$n_1+n_2+ ... +n_i$	$r_1+r_2+ ... +r_i$	$p_1+p_2+ ... +p_i$
...	...	...	...	...	...	...
	$\Sigma n_i = N$	$\Sigma r_i = 1$	$\Sigma p_i = 100$			

EJEMPLO :

x	n	r	p	N	R	P
2	5	0'125	12'5	5	0'125	12'5
3	10	0'250	25	15	0'375	37'5
4	16	0'400	40	31	0'775	77'5
5	6	0'150	15	37	0'925	92'5
6	3	0'075	7'5	40	1'000	100
	40	**1**	**100**			

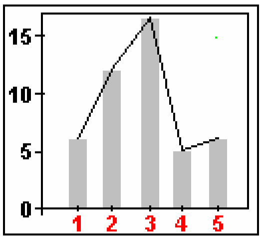
GRÁFICOS ESTADÍSTICOS.

La norma que hemos de seguir en la construcción de un gráfico estadístico es siempre : **"La zona que identifica a cada valor será proporcional a su frecuencia"**

Los diagramas usuales son los que se describen a continuación.

A Diagramas de barras

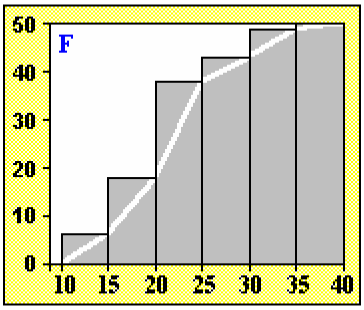
Para variables cualitativas o cuantitativas no agrupadas en intervalos.



FUNDAMENTO : Sobre un eje (normalmente el horizontal) marcamos los valores de la variable, dibujando sobre cada uno de ellos una ***barra*** cuya longitud sea proporcional a la frecuencia que se esté visualizando.
Si la variable representada **es cuantitativa**, enlazando los extremos de las barras obtendremos el **POLÍGONO DE FRECUENCIAS,** denominado **PERFIL ORTOGONAL** para **cualitativas ordenables** .

B Histogramas

Representativo de las variables agrupadas en intervalos.

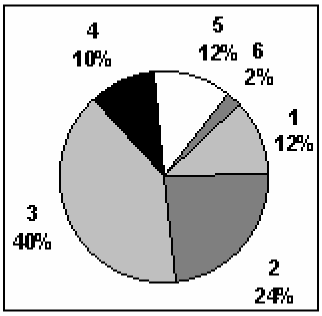


FUNDAMENTO : Sobre el eje horizontal marcamos los distintos ***intervalos***, dibujando sobre cada uno de ellos un ***rectángulo*** cuya área sea proporcional a la frecuencia que se esté visualizando (Si todos los intervalos tienen la misma amplitud, nos bastará con que la **altura** de los rectángulos sea proporcional a las frecuencias).

POLÍGONOS DE FRECUENCIAS :

Si la frecuencia representada **no es acumulada**, enlazamos los puntos medios de los extremos superiores de los rectángulos.
Para **frecuencias acumuladas**, el polígono de frecuencias se obtiene de la forma indicada en el gráfico.

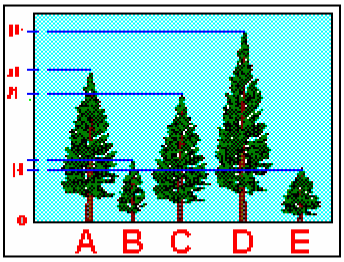
C Diagramas de sectores



Utilizable en cualquier tipo de variable.

FUNDAMENTO : Dividimos el círculo en **sectores circulares**, de modo que la amplitud de cada sector, sea proporcional a la frecuencia. Junto a cada sector, se suele indicar el valor representado. Es aconsejable la expresión de las amplitudes de los sectores en % (porcentajes **p**).

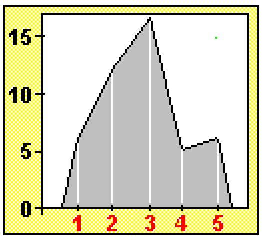
D Pictogramas



Utilizable en todo tipo de variables, especialmente con las cualitativas.

FUNDAMENTO : Es el mismo que se sigue para la construcción de los **diagramas de barras** y **histogramas**. La diferencia estriba en que, en lugar de dibujar una barra o un rectángulo, **se dibuja una figura** que hace referencia al problema objeto de estudio.

E Diagramas de áreas



Representativo de las variables cuantitativas, equivale a la representación independiente de los **polígonos de frecuencias** (descritos en los diagramas de barras y histogramas).

FUNDAMENTO : Indica la evolución de los valores de la variable, consistiendo en la visualización del área encerrada bajo el polígono de frecuencias. Para ello, se conecta dicho polígono con el eje de la variable *(el horizontal en el gráfico)*, tanto a la izquierda del primer valor como a la derecha del último.

Los diagramas de barras , histogramas , pictogramas y de áreas , admiten la representación correspondiente a sus frecuencias acumuladas.

MEDIDAS DE CENTRALIZACIÓN.

MEDIA ARITMÉTICA :

$$\bar{x} = \frac{\sum n_i . x_i}{N}$$

Es el resultado de dividir la suma de todas las observaciones entre el número de ellas.

MODA :

$$Mo = e_i + \frac{n_{i+1}}{n_{i+1} + n_{i-1}} . a_i$$

Es el valor que más se repite. Será pues el valor (o valores) cuya frecuencia absoluta sea la mayor de las observadas.

Si los datos se encuentran agrupados en intervalos, obtendremos el intervalo en el que se encuentra la moda (INTERVALO MODAL). Para determinar su valor concreto, aplicamos la expresión de la izquierda.

NOTACIONES

Los subíndices indican :
- **i** intervalo donde se encuentra la moda.
- **i-1** intervalo anterior al que contiene la moda.
- **i+1** intervalo siguiente al que contiene la moda.

e extremo inferior del intervalo en el que se encuentra la moda.
a amplitud del intervalo en el que está la moda.
n frecuencia absoluta.

MEDIANA :

Supuestas ordenadas las observaciones, MEDIANA es el valor de la variable que está en el centro de las mismas. Deja pues a la mitad (el 50%) de las observaciones por debajo de dicho valor.

Para obtener el valor de la mediana, seguimos los pasos siguientes :
1º Calculamos la tabla de frecuencias absolutas acumuladas.
2º La mediana será el valor de la variable cuya frecuencia absoluta acumulada primero iguale o supere a N/2.

$$Me = e_i + \frac{\frac{N}{2} - N_{i-1}}{n_i} . a_i$$

Si los datos se encuentran agrupados en intervalos, el punto 2º nos dará el intervalo en el que se encuentra la mediana. Para determinar su valor concreto, aplicamos la expresión de la izquierda.

NOTA : En el caso de variables continuas no agrupadas en intervalos, suelen considerarse previamente los intervalos reales que esos valores representan, procediendo a aplicar la expresión superior.

Así, los valores 1 , 2 ,3 , ... representan a los intervalos de valores [0'5 , 1'5) , [1'5 , 2'5) , [2'5 , 3'5) , ...

NOTACIONES

Los subíndices indican : **i** intervalo donde se encuentra la mediana.
i-1 intervalo anterior al que contiene la mediana.

e extremo inferior del intervalo en el que se encuentra la mediana.
a amplitud del intervalo en el que está la mediana.
n frecuencia absoluta.
N frecuencia absoluta acumulada.

OTRAS MEDIDAS DE CENTRALIZACIÓN.

MEDIA PONDERADA :	MEDIA GEOMÉTRICA :	MEDIA ARMÓNICA :
Aplicable cuando a cada valor (X_i) se le asigna un peso (p_i) : $\overline{x}_p = \dfrac{\sum p_i . X_i}{\sum p_i}$	$\overline{x}_G = \sqrt[N]{x_1 . x_2 . \ \ldots \ . x_N}$ Con frecuencias f_i para cada x_i : ($N = \Sigma f_i$) $\overline{x}_G = \sqrt[N]{x_1^{n_1} . x_2^{n_2} ..... x_n^{n_n}}$	$\overline{x}_A = \dfrac{N}{\sum\left(\dfrac{1}{x_i}\right)}$ Con frecuencias f_i para cada x_i : ($N = \Sigma f_i$) $\overline{x}_A = \dfrac{N}{\sum\left(\dfrac{n_i}{x_i}\right)}$

MEDIDAS DE POSICIÓN.

CONCEPTO : Permiten el cálculo del valor de la variable que ocupa una cierta posición relativa respecto del conjunto total de los valores observados.

PERCENTIL DE ORDEN K : Es el **valor de la variable** que deja por debajo de él el **K%** de las observaciones.

PROCESO DE CALCULO :

Para obtener el valor del percentil de orden K, seguimos los pasos siguientes :

1º Calculamos la tabla de frecuencias absolutas acumuladas.
2º Obtenemos el LUGAR que ocupa : Lugar = N . K / 100
3º El percentil de orden K será el valor de la variable cuya frecuencia absoluta acumulada primero iguale o supere a dicho lugar.

$$P_k = e_i + \frac{\frac{k.N}{100} - N_{i-1}}{n_i} . a_i$$

Si los datos se encuentran agrupados en intervalos, el punto 3º nos dará el intervalo en el que se encuentra el percentil de orden K. Para determinar el valor concreto del percentil, aplicamos la expresión de la izquierda.

NOTA : En el caso de variables continuas no agrupadas en intervalos, suelen considerarse previamente los intervalos reales que esos valores representan, procediendo a aplicar la expresión anterior.

Así, los valores 1 , 2 ,3 , ... representan a los intervalos de valores [0'5 , 1'5) , [1'5 , 2'5) , [2'5 , 3'5) , ...

NOTACIONES

Los subíndices indican : **i** intervalo donde se encuentra el percentil.
i-1 intervalo anterior al que contiene el percentil.

e extremo inferior del intervalo en el que se encuentra el percentil.
a amplitud del intervalo en el que está el percentil.
n frecuencia absoluta.
N frecuencia absoluta acumulada.

PERCENTILES ESPECIALES

MEDIANA Percentil de orden 50.
CUARTILES Percentiles de órdenes 25 (Cuartil 1º), 50 (Cuartil 2º) y 75 (Cuartil 3º).
DECILES Percentiles de órdenes 10, 20, ... , 90 (Deciles 1º, 2º, ... , 9º).

MEDIDAS DE DISPERSIÓN.

RANGO , RECORRIDO O AMPLITUD TOTAL :

$$\mathbf{R = Máx - Mín}$$

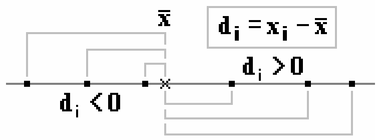
Con el fin de medir el mayor o menor grado de separación de las observaciones, en una primera instancia se define el RANGO (también denominado recorrido o amplitud total), como la diferencia existente entre los valores máximo y mínimo observados.

AMPLITUD SEMI-INTERCUARTÍLICA :

$$Q = \frac{Q_3 - Q_1}{2}$$

Esta medida de dispersión se basa en medidas de posición (Cuartiles),.Su empleo tendrá sentido en el supuesto de imposibilidad de cálculo de la media.

El no tomar en consideración a la totalidad de las observaciones, hace pensar que esta medida es poco representativa. Por ello se intenta definir las medidas de dispersión, de modo que sean el promedio de las separaciones de cada valor respecto de uno tomado como referencia (la MEDIA).



Observando la figura apreciamos que las desviaciones **d** antes definidas tienen como media cero (las positivas compensan con las negativas), lo cuál obliga a subsanar este inconveniente **tomándolas en valor absoluto o elevándolas al cuadrado**.

DESVIACIÓN MEDIA :

$$D_x = \frac{\sum n_i . |x_i - \bar{x}|}{N}$$

Es la media de las desviaciones o separaciones de cada una de las observaciones, respecto a la media aritmética, consideradas en valor absoluto. Sustituyendo la media por la moda o la mediana, definiremos **las desviaciones medias respecto de la moda** y de **la mediana**.

VARIANZA :

$$s^2 = \sigma^2 = \frac{\sum n_i . (x_i - \bar{x})^2}{N} = \frac{\sum n_i . x_i^{\ 2}}{N} - \bar{x}^2$$

Es la media de los cuadrados de las desviaciones o separaciones de cada una de las observaciones, respecto a la media aritmética.

DESVIACIÓN TÍPICA :

$$s = \sigma = \sqrt{\text{var}\,ianza} = \sqrt{\frac{\sum n_i . x_i^{\ 2}}{N} - \bar{x}^2}$$

Es la raíz cuadrada de la varianza. Con ello corregimos el haber tomado cuadrados de separaciones en el cálculo de la varianza. Esta medida de dispersión es la más característica.

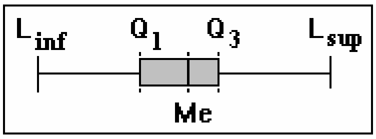
COEFICIENTE DE VARIACIÓN :

$$CV = \frac{\sigma_x}{\bar{x}} .100$$

Mide la representatividad de la media. Valores extremos del mismo nos llevarán a concluir que la media no es representativa, es decir, existirán valores entre las observaciones que se separan significativamente de las demás.
Sólo puede ser utilizado cuando los valores de la variable toman valores "normales". Es decir, no son muy elevados ni muy pequeños, ya que una media próxima a cero o muy alta darían valores nulos o infinitos al coeficiente.
Si la media es representativa de las observaciones (no existen valores extremos exageradamente distanciados de la mayoría), el coeficiente de variación permite **comparar la dispersión de dos series estadísticas** : mayor coeficiente indica menor homogeneidad, o lo que es lo mismo, mayor dispersión o variabilidad.

GRÁFICO DE VARIABILIDAD :

Basado en los cuartiles, adopta la forma del gráfico de la derecha. En él se reflejan los cuartiles 1º y 3º y la mediana, junto a los extremos inferior y superior :



$$L_{inf} = Q_1 - 3.\frac{Q_3 - Q_1}{2} = Q_1 - 3.Q \quad ; \quad L_{sup} = Q_3 + 3.Q$$

Se consideran observaciones atípicas aquellas que quedan fuera del intervalo :

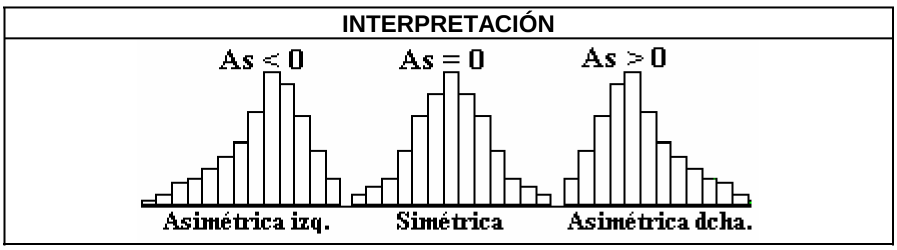
(L_{inf} , L_{sup})

OTRAS MEDIDAS ESTADÍSTICAS.

COEFICIENTE DE ASIMETRÍA DE FISHER :

Permite interpretar la forma de la distribución, respecto a ser o no simétrica.

$$As_1 = \frac{\dfrac{\sum n_i . (x_i - \bar{x})^3}{N}}{\sigma^3}$$



Basados en al relación existente entre media, mediana y moda : $\bar{x} - Mo = 3.(\bar{x} - Md)$

se definen dos nuevos coeficientes de asimetría (de Pearson):

$$As_2 = \frac{\bar{x} - Mo}{\sigma} \qquad As_3 = \frac{3.(\bar{x} - Md)}{\sigma}$$

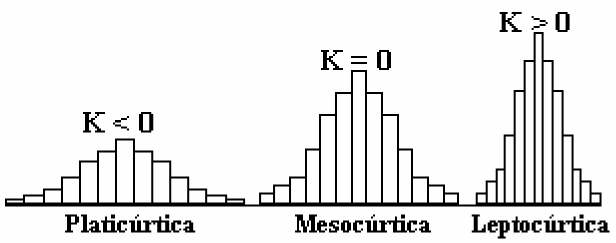
COEFICIENTE DE CURTOSIS :

Recibe también el nombre de coeficiente de ***concentración central***, midiendo el grado de aplastamiento o apuntamiento de la gráfica de la distribución de la variable estadística. Una mayor concentración de datos en torno al promedio harán que la forma sea alargad, siendo tanto más plana (o aplastada) cuanto mayor sea la dispersión de los mismos.

Determina la forma de la distribución, en relación con su grado de aplastamiento.

$$K = \frac{\dfrac{\sum n_i.(x_i - \bar{x})^4}{N}}{\sigma^4} - 3$$

INTERPRETACIÓN



Basados en medidas de posición, se definen los nuevos coeficientes :

Coeficiente de asimetría de Bowley-Yule, o intercuartílico : $Y = \frac{Q_3 - 2.Me + Q_1}{Q_3 - Q_1}$ **Coeficiente absoluto de asimetría:** $A = \frac{Q_3 - 2.Me + Q_1}{\sigma}$	**Coeficiente de curtosis de Kelley :** $K = \frac{Q}{P_{90} - P_{10}} - 0'263 \quad \left(\text{con:} \quad Q = \frac{Q_3 - Q_1}{2}\right)$

ANÁLISIS CONJUNTO DE VARIOS GRUPOS.

Si disponemos de **k** grupos con $\mathbf{n_i}$ elementos, medias $\bar{x}_i$, y varianzas S_i^2, podemos obtener :

Media conjunta de los k grupos	**Varianza conjunta de los k grupos**
$\bar{X} = \frac{\sum n_i.\bar{x}_i}{\sum n_i}$	$S^2 = \frac{\sum n_i.S_i^2}{\sum n_i}$ **, o, con mayor rigor :** $S^2 = \frac{\sum n_i.S_i^2}{\sum n_i} + \frac{\sum n_i.(\bar{x}_i - \bar{X})^2}{\sum n_i}$

PROPIEDADES DE LAS MEDIDAS ESTADÍSTICAS.

TABLA PARA CÁLCULOS :

La tabla siguiente nos muestra una disposición práctica de los cálculos necesarios para la obtención de los parámetros estadísticos usuales: **Media , Moda, Mediana , Percentiles , Varianza y Desviación típica**.

Intervalos	x	n	n.x	n.x²	N	P
$[e_1, e_2)$	x_1	n_1	$n_1 . x_1$	$(n_1 . x_1).x_1$	$N_1=n_1$	$P_1 = (N_1 / N) . 100$
$[e_2, e_3)$	x_2	n_2	$n_2 . x_2$	$(n_2 . x_2).x_2$	$N_2=n_1+n_2$	$P_2 = (N_2 / N) . 100$
. . .	. . .	. . .	. . .	. . .	. . .	. . .
$[e_i, e_{i+1})$	x_i	n_i	$n_i . x_i$	$(n_i . x_i).x_i$	$N_i=n_1+n_2+ ... +n_i$	$P_i = (N_i / N) . 100$
. . .	. . .	. . .	. . .	. . .	. . .	. . .
		Σn_i	$\Sigma n_i . x_i$	$\Sigma n_i . x_i^2$	Cálculo de percentiles	
		N	A	B		
		Cálculo de media y varianza				

La media y la varianza serían el resultado de calcular :**Cálculo de media y varianza**

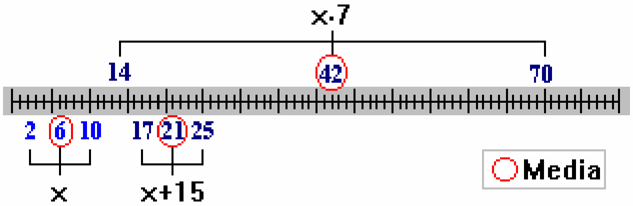
$$\bar{x} = \frac{A}{N} \qquad \sigma^2 = \frac{B}{N} - \bar{x}^2$$

PROPIEDADES :

A) Si a todos los valores de una variable **x** les sumamos una cantidad constante, la media queda incrementada en dicha constante, mientras que la desviación típica (y la varianza) no varía.

B) Si multiplicamos todos los valores de una variable **x** por una constante, la media y la desviación típica quedan también multiplicadas por dicha constante (la varianza quedará multiplicada por el cuadrado de la constante).

EJEMPLO :



CAMBIO DE VARIABLE. TIPIFICACIÓN.

Haciendo uso de las propiedades de las medidas estadísticas ,podremos facilitar y simplificar los cálculos de parámetros estadísticos, realizando un **cambio de variable**.
Así, si todos los valores son muy altos, podremos restarles una cantidad (normalmente la Moda) y, si poseen cifras decimales o son múltiplos de un mismo número, podremos multiplicarlos o dividirlos por el valor adecuado.
Una vez calculados los parámetros estadísticos, en virtud de las **propiedades** descritas, obtendremos el valor final real de tales parámetros.

Mención especial merecen dos cambios de variables particulares :

A) **Diferenciales :** partiendo de la variable inicial **x** *(puntuaciones directas)*, si a todos los valores les **restamos la media**, obtenemos una nueva variable **d** *(puntuaciones diferenciales)* cuya **media es cero** (la desviación típica no se modifica).

B) **Tipificadas :** Si a todos los valores de la variable inicial **x** les **restamos la media** y el resultado lo **dividimos por la desviación típica**, obtenemos una nueva variable **z** *(puntuaciones tipificadas)* cuya **media es cero** , teniendo siempre como **desviación típica la unidad**.

Este último cambio de variable recibe el nombre de **TIPIFICACIÓN**.

SUMA Y DIFERENCIA DE VARIABLES.

Partiendo de dos variables X , Y, podemos definir las nuevas variables :

- **S = X + Y** obtenida sumando cada valor de X con el correspondiente de Y.
- **D = X - Y** obtenida restando a cada valor de X el valor correspondiente de Y.

Esto supone la existencia de tantas observaciones de X como de Y, así como el emparejamiento de ellas; es decir, a cada valor de X queda asociado un valor de Y. Esto constituirá la base de estudio del siguiente tema .

Veamos como se comporta la media de las dos nuevas variables **S** y **D** definidas.

$\overline{S} = \overline{X} + \overline{Y}$ En efecto : $\overline{S} = \dfrac{\sum (X_i + Y_i)}{N} = \dfrac{\sum X_i + \sum Y_i}{N} = \dfrac{\sum X_i}{N} + \dfrac{\sum Y_i}{N} = \overline{X} + \overline{Y}$

Análogamente se verifica que : $\overline{D} = \overline{X} - \overline{Y}$

Calculemos la varianza de la suma S :

$$S_S^2 = \frac{\sum\left((X_i + Y_i) - \overline{S}\right)^2}{N} = \frac{\sum\left((X_i + Y_i) - (\overline{X} + \overline{Y})\right)^2}{N} = \frac{\sum\left((X_i - \overline{X}) + (Y_i - \overline{Y})\right)^2}{N} =$$

$$= \frac{\sum\left((X_i - \overline{X})^2 + (Y_i - \overline{Y})^2 + 2.(X_i - \overline{X}).(Y_i - \overline{Y})\right)}{N} =$$

$$= \frac{\sum (X_i - \overline{X})^2}{N} + \frac{\sum (Y_i - \overline{Y})^2}{N} + 2.\frac{\sum (X_i - \overline{X}).(Y_i - \overline{Y})}{N} = S_X^2 + S_Y^2 + 2.S_{XY}$$

La expresión $\dfrac{\sum (X_i - \overline{X}).(Y_i - \overline{Y})}{N}$, representada por $\mathbf{S_{XY}}$, recibe el nombre de covarianza, justificándose que es igual también a :

$$S_{XY} = \frac{\sum (X_i - \overline{X}).(Y_i - \overline{Y})}{N} = \frac{\sum X_i.Y_i}{N} - \overline{X}.\overline{Y}$$

Análogamente se verifica que : $S_D^2 = S_X^2 + S_Y^2 - 2.S_{XY}$

Si las variables X , Y son independientes, la covarianza (medida de variación conjunta) es igual a **cero**.

Resumiendo :	**Medias**	**Varianzas**	
		Dependientes ($S_{XY} \neq 0$)	**Independientes ($S_{XY} = 0$)**
S = X + Y	$\overline{S} = \overline{X} + \overline{Y}$	$S_S^2 = S_X^2 + S_Y^2 + 2.S_{XY}$	$S_S^2 = S_X^2 + S_Y^2$
D = X - Y	$\overline{D} = \overline{X} - \overline{Y}$	$S_D^2 = S_X^2 + S_Y^2 - 2.S_{XY}$	$S_D^2 = S_X^2 + S_Y^2$

MOMENTOS ORDINARIOS Y CENTRALES

Momento ordinario de orden k :

$$a_k = \sum \frac{n}{N}.x^k$$

Momento central de orden k :

$$m_k = \sum \frac{n}{N}.(x - \bar{x})^k$$

Se verifica que :

$$m_1 = 0 \qquad m_2 = a_2 - a_1^2$$

$$m_3 = a_3 - 3.a_2.a_1 + 2.a_1^3$$

$$m_4 = a_4 - 4.a_3.a_1 + 6.a_2.a_1^2 - 3.a_1^4$$

Algunos parámetros estudiados, pueden expresarse :

$$\mu = \bar{x} = a_1 \qquad \sigma^2 = s_x^2 = m_2 = a_2 - a_1^2$$

$$As = \frac{m_3}{\sigma^3} = \frac{m_3}{\left(\sqrt{m_2}\right)^3} \qquad K = \frac{m_4}{\sigma^4} - 3 = \frac{m_4}{m_2^2} - 3$$

MEDIDAS DE CONCENTRACIÓN.

Estas medidas, de aplicación económica fundamentalmente, determinan el nivel de igualdad en el reparto total de las observaciones de la variable.

Su determinación se realizará a partir de la siguiente tabla de cálculos :

A	B	C	D	E	N	G	H
x_i	**n_i**	**$N_i = \Sigma\, n_i.$**	**$P_i = (N_i.. /N).100$**	**$t_i = n_i. x_i$**	**$T_i = \Sigma\, t_i.$**	**$Q_i = (T_i.. /T).100$**	**$P_i - Q_i$**
x_1	n_1	N_1	P_1	t_1	T_1	Q_1	$P_1 - Q_1$
x_2	n_2	N_2	P_2	t_2	T_2	Q_2	$P_2 - Q_2$
...	...	...	...	...	...	...	...
x_k	n_k	N_k	P_k (= 100)	t_k	T_k	Q_k (= 100)	$P_k - Q_k$ (= 0)
	$N = \Sigma\, n_i.$		**$TP = \Sigma\, P_i$**	**$T = \Sigma\, n_i. x_i$**			**$TD = \Sigma\, (P_i - Q_i)$**

Siendo :

A) Valores de la variable (marca de clase si está agrupada en intervalos).
B) Frecuencias absolutas (N = total de observaciones).
C) Frecuencias absolutas acumuladas.
D) Porcentajes acumulados (totalizando - TP).
E) Productos de cada frecuencia por su correspondiente valor (T = suma total de estos productos).
F) Productos anteriores acumulados (de igual modo que se realiza con frecuencias).
G) Expresión en porcentaje del contenido de la columna anterior.
H) Diferencias de los valores de las columnas D y G (totalizando - TD).

MEDIALA :

Su definición tiene un fundamento similar al de la **mediana**.

- Para distribuciones discretas (no agrupadas en intervalos), la **mediala** es el valor de la variable cuyo **Q_i** primero iguala o supera el **50%**.
- Para distribuciones continuas (agrupadas en intervalos), el **intervalo que contiene la mediala** es aquel cuyo **Q_i** primero iguala o supera el **50%**. De aquí obtenemos el valor de la **mediala** del modo siguiente :

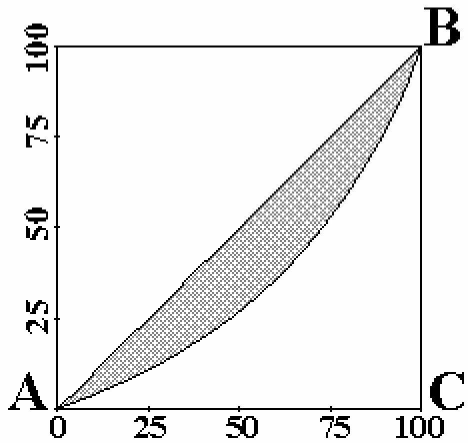
$$Ml = e_i + \frac{50 - Q_{i-1}}{Q_i - Q_{i-1}}.a_i$$

Los subíndices indican :

i intervalo donde se encuentra la mediala.
i-1 intervalo anterior al que contiene la mediala.

e extremo inferior del intervalo en el que se encuentra la mediala.
a amplitud del intervalo en el que está la mediala.

CURVA DE LORENZ :



Sobre un rectángulo de 100 unidades de lado, se dibuja la poligonal que resulta de unir los puntos (**P_i , Q_i**).

Esta poligonal (curva de Lorenz) determina con la diagonal AB un recinto (sombreado en la figura) que mide el grado de concentración.

Cuando el área sombreada es muy pequeña (la curva de Lorenz se aproxima a la diagonal AB) se presenta una baja concentración, o lo que es lo mismo, indica uniformidad en el reparto de los valores de la variable.

La mayor concentración se producirá cuando la zona sombreada coincide con el triángulo ABC.

ÍNDICE DE CONCENTRACIÓN DE GINI :

Haciendo uso de la tabla de cálculos anterior, necesaria para la obtención de la curva de Lorenz, definiremos el presente estadístico. Otros, como el índice de Dalton, el de paridad, etc. , pueden ser empleados con idéntica interpretación a la que tratamos con el de Gini, si bien omitimos su estudio.

$$G = \frac{\sum_{i=1}^{k-1} \left(P_i - Q_i \right)}{\sum_{i=1}^{k-1} P_i} = \frac{TD}{TP - 100}$$

El índice de Gini (expresión de la izquierda) coincide geométricamente con el cociente entre el área sombreada (definida por la curva de Lorenz) y la del triángulo ABC.

- Concentración mínima : G = 0
- Concentración máxima : G = 1

1

La tabla siguiente nos muestra el resultado de una encuesta entre los alumnos de primer curso, analizando el número de suspensos en la primera evaluación :

0	**2**	**2**	**4**	**0**	**3**	**3**	**2**	**5**	**2**	**3**	**2**	**4**	**3**	**4**
3	**1**	**4**	**1**	**1**	**0**	**4**	**1**	**1**	**4**	**2**	**4**	**2**	**0**	**3**
1	**3**	**0**	**5**	**2**	**2**	**3**	**0**	**3**	**0**	**5**	**1**	**1**	**4**	**0**
3	**2**	**3**	**2**	**3**	**3**	**1**	**2**	**4**	**2**	**3**	**1**	**3**	**1**	**4**

Realicemos un estudio estadístico completo.

Se trata de una variable **cuantitativa discreta**. Esto condicionará algunos procesos del cálculo estadístico.

RECUENTO Y TABLA DE FRECUENCIAS

x	recuento	n	r	p	N	R	P
0	///// ///	**8**	0'1333	13'33	8	0'1333	13'33
1	///// ///// /	**11**	0'1833	18'33	19	0'3167	31'67
2	///// ///// ///	**13**	0'2167	21'67	32	0'5333	53'33
3	///// ///// /////	**15**	0'2500	25'00	47	0'7833	78'33
4	///// /////	**10**	0'1667	16'67	57	0'9500	95'00
5	///	**3**	0'0500	5'00	60	1'0000	100'00
	Totales :	**N = 60**	1'0000	100'00			

GRÁFICOS ESTADÍSTICOS APROPIADOS PARA ESTE TIPO DE VARIABLE

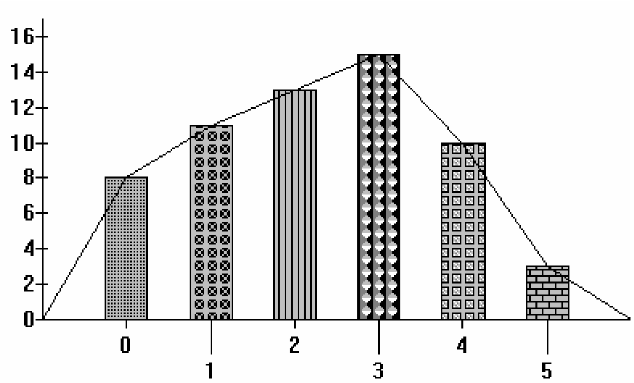
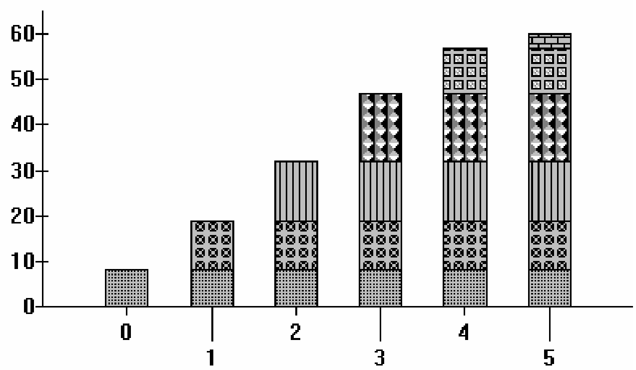


DIAGRAMA DE BARRAS :
Sobre el valor de cada variable dibujamos una barra con altura igual a la frecuencia que deseamos representar (en este caso las absolutas n).

POLÍGONO DE FRECUENCIAS :
Obtenidos enlazando los extremos superiores de las barras.

NOTA :Siendo la variable discreta, no tiene sentido dibujar el polígono de frecuencias.



DIAGRAMAS ACUMULADOS :

Construidos como los anteriores, son los representativos de las distintas frecuencias acumuladas.

El ejemplo representa las frecuencias absolutas acumuladas (N).

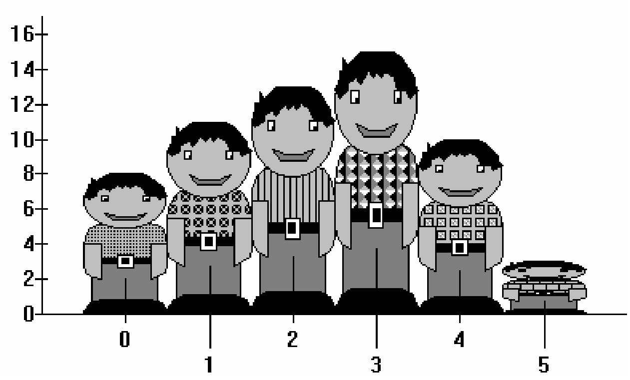
El polígono de frecuencias se construiría enlazando los extremos superiores de las barras.

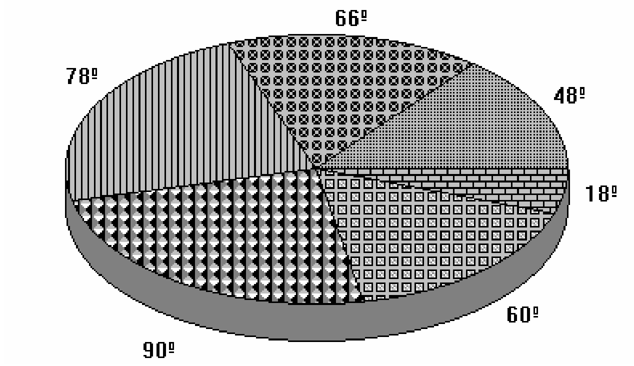
PICTOGRAMAS:

Con el mismo principio seguido para la construcción de los diagramas de barras, sustituimos dichas barras por dibujos alusivos a la variable estadística estudiada.

DIAGRAMAS DE SECTORES :
Resultan de la división de un círculo en sectores cuya amplitud es proporcional a la frecuencia.

La amplitud de cada sector será : $\alpha = \frac{n}{N}.360^{\circ} = r.360^{\circ}$





MEDIA, VARIANZA Y DESVIACIÓN TÍPICA

x	n	n.x	n.x^2
0	**8**	0	0
1	**11**	11	11
2	**13**	26	52
3	**15**	45	135
4	**10**	40	160
5	**3**	15	75
	N = 60	**137**	**433**

Este tipo de tabla facilita los cálculos.

Media = 137 / 60 = 2,283
Varianza = (433 / 60) - media al cuadrado = 2'005
Desviación típica = raíz cuadrada de la varianza = 1'416

$$\overline{x} = \frac{\sum n_i . x_i}{N} = \frac{137}{60} = 2'283 \quad s_x^2 = \frac{\sum n_i . x_i^2}{N} - \overline{x}^2 = \frac{433}{60} - 2'283^2 = 2'0(\quad \mathbf{s_x = \sqrt{s_x^2} = \sqrt{2'005} = 1'416}$$

MODA = Valor de mayor frecuencia **= 3**

PERCENTILES

Para la determinación de medidas de posición (percentiles), podemos seguir dos procedimientos de cálculo :

1º) Basado en las frecuencias absolutas acumuladas N :

Determinamos el lugar que ocupa : L = k.N / 100

El percentil será el valor cuya frecuencia N primero iguale o supere al lugar L.

2º) Basado en porcentajes acumulados P :

El percentil será el valor cuyo porcentaje P primero iguale o supere al orden k del percentil.

Apliquemos el primer procedimiento para calcular la **mediana** y el **9º decil** :

La mediana (percentil 50) ocupará el lugar : L = 50 . 60 / 100 = 30

El 9º decil (percentil 90) ocupará el lugar : L = 90 . 60 / 100 = 54

x	n
0	**8**
1	**11**
2	**13**
3	**15**
4	**10**
5	**3**
	N = 60

N		
8		
19		
32	⇐	Mediana = 2
47		
57	⇐	9º decil = 4
60		

Aplicando el segundo procedimiento descrito, determinemos los **cuartiles 1º y 3º**, así como la **amplitud semi-intercuartílica** :

x	n	r	p	P		
0	**8**	0'1333	13'33	13'33		
1	**11**	0'1833	18'33	31'67	⇐	Cuartil 1º (percentil 25) = 1
2	**13**	0'2167	21'67	53'33		
3	**15**	0'2500	25'00	78'33	⇐	Cuartil 3º (percentil 75) = 3
4	**10**	0'1667	16'67	95'00		
5	**3**	0'0500	5'00	100'00		
	N = 60	1'0000	100'00			

Amplitud semi-intercuartílica = $\mathbf{\frac{Q_3 - Q_1}{2} = \frac{3-1}{2} = 1}$

2

Trabajamos ahora con las edades de 50 jóvenes de nuestro barrio :

1	**11**	**20**	**15**	**10**	**4**	**12**	**20**	**5**	**23**	**9**	**12**	**13**	**14**	**15**
24	**15**	**7**	**8**	**12**	**9**	**9**	**5**	**2**	**20**	**13**	**15**	**7**	**11**	**22**
20	**6**	**12**	**4**	**7**	**1**	**18**	**20**	**11**	**10**	**14**	**20**	**11**	**13**	**15**
21	**25**	**20**	**22**	**10**										

Como en el ejemplo anterior, realicemos un estudio estadístico completo.

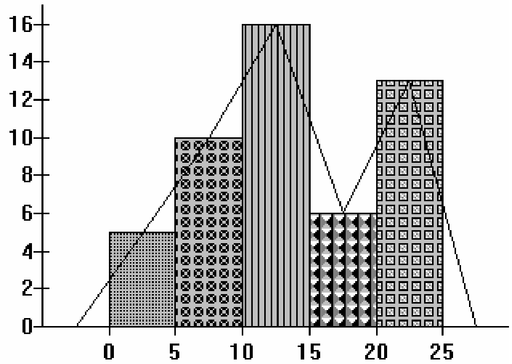
Nos encontramos ante una variable estadística cuantitativa continua. Agruparemos o no las observaciones en intervalos en función de los diferentes valores observados.

TABLA DE FRECUENCIAS

Observado el valor mínimo (1) y máximo (24), decidimos agrupar los datos en intervalos de 5 años de amplitud, empezando por 0.

Intervalos	recuento	n	r	p	N	R	P
[0 , 5)	/////	**5**	0'10	10	5	0'10	10
[5, 10)	///// /////	**10**	0'20	20	15	0'30	30
[10 , 15)	///// ///// ///// /	**16**	0'32	32	31	0'62	62
[15 , 20)	///// /	**6**	0'12	12	37	0'74	74
[20 , 25]	///// ///// ///	**13**	0'26	26	50	1'00	100
	Totales :	**N = 50**	1'00	100			

GRÁFICOS ESTADÍSTICOS

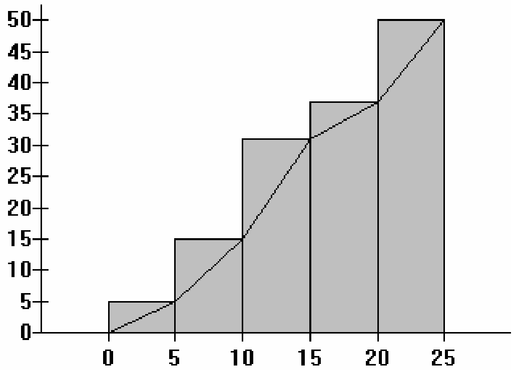


HISTOGRAMA :

Sobre el valor de cada variable dibujamos una franja con altura igual a la frecuencia que deseamos representar (en este caso las absolutas n).

POLÍGONO DE FRECUENCIAS :

Obtenido enlazando los puntos medios de los extremos superiores de las franjas.



HISTOGRAMAS ACUMULADOS :

Construidos como los anteriores, son los representativos de las distintas frecuencias acumuladas.

El ejemplo representa las frecuencias absolutas acumuladas (N).

En este caso, el polígono de frecuencias **NO** se construiría enlazando los puntos medios de los extremos superiores de las franjas, sino como se indica en la figura.

Cálculo de Moda, Media, Varianza y Desviación típica :

Para el cálculo de la media y la varianza utilizamos la tabla auxiliar siguiente. En ella se incorpora la columna **x** , que contiene la marca de clase (valor central) de cada intervalo.

La **MODA** (valor de mayor frecuencia) se encuentra en el intervalo [10 , 15) . Determinemos su valor concreto :

$$Mo = e_i + \frac{n_{i+1}}{n_{i+1} + n_{i-1}}.a_i = 10 + \frac{6}{6+10}.5 = 11'875$$

Intervalos	n	x	n.x	n.x²
[0 , 5)	**5**	2'5	12'5	31'25
[5, 10)	**10**	7'5	75'0	562'50
[10 , 15)	**16**	12'5	200'0	2500'00
[15 , 20)	**6**	17'5	105'0	1837'50
[20 , 25]	**13**	22'5	292'5	6581'25
	N = 50		685'0	11512'50

$$\bar{x} = \frac{\sum n_i.x_i}{N} = \frac{685}{50} = 13'7 \qquad s_x^2 = \frac{\sum n_i.x_i^2}{N} - \bar{x}^2 = \frac{11512'5}{50} - 13'7^2 = 42 \quad \mathbf{s_x = \sqrt{s_x^2} = \sqrt{42'56} = 6'524}$$

Utilizando las frecuencias absolutas acumuladas, calculemos el **decil 2º** y el **percentil 62** :

Lugar que ocupa el decil 2º (percentil 20) = 20 . 50 / 100 = 10

Lugar que ocupa el percentil 62 = 62 . 50 / 100 = 31

Intervalos	n	N			
[0 , 5)	**5**	5			
[5, 10)	**10**	15	⇐	Decil 2º (percentil 20) en [5,10)	Lugar = 10
[10 , 15)	**16**	31	⇐	Percentil 62 en [10,15)	Lugar = 31
[15 , 20)	**6**	37			
[20 , 25]	**13**	50			
	N = 50				

Determinemos sus valores concretos :

$$P_{20} = e_i + \frac{\frac{20.N}{100} - N_{i-1}}{n_i}.a_i = 5 + \frac{\frac{20.50}{100} - 5}{10}.5 = 7'5$$

$$P_{62} = e_i + \frac{\frac{62.N}{100} - N_{i-1}}{n_i}.a_i = 10 + \frac{\frac{62.50}{100} - 15}{16}.5 = 15$$

Utilizando los porcentajes acumulados, calculemos el **cuartil 1º** y la **mediana** :

Intervalos	n	r	p	P		
[0 , 5)	**5**	0'10	10	10		
[5, 10)	**10**	0'20	20	30	⇐	Cuartil 1º (percentil 25) en [5,10)
[10 , 15)	**16**	0'32	32	62	⇐	Mediana (percentil 50) en [10,15)
[15 , 20)	**6**	0'12	12	74		
[20 , 25]	**13**	0'26	26	100		
	N = 50	1'00	100			

Determinemos sus valores concretos :

$$P_{25} = e_i + \frac{\frac{25.N}{100} - N_{i-1}}{n_i}.a_i = 5 + \frac{\frac{25.50}{100} - 5}{10}.5 = 8'75$$

$$P_{50} = e_i + \frac{\frac{50.N}{100} - N_{i-1}}{n_i}.a_i = 10 + \frac{\frac{50.50}{100} - 15}{16}.5 = 13'125$$

3

x	n
2	**6**
3	**15**
4	**10**
5	**9**

De la presente distribución, calculemos :

Media, varianza y desviación típica.
Moda.
Mediana, Percentil 82, Cuartiles y amplitud semi-intercuartílica.

La variable establecida puede ser discreta o continua sin agrupar en intervalos. Realicemos los cálculos en ambos supuestos.

x	n
2	6
3	15
4	10
5	9
	40

N	P
6	15
21	52'5
31	77'5
40	100

n.x	n.x²
12	24
45	135
40	160
45	225
142	**544**

Media $\overline{x} = \frac{\sum n_i . x_i}{N} = \frac{142}{40} = 3'55$	**Varianza** $\sigma^2 = \frac{\sum n_i . x_i{}^2}{N} - \overline{x}^2 = \frac{544}{40} - 3'55^2 = 0'9$	**Desviación típica** $\sigma = \sqrt{0'9975} = 0'99875$
Moda **3**	**Mediana** (percentil 50) **3**	**Percentil 82** **5**
Cuartil 1º (percentil 25) **3**	**Cuartil 3º** (percentil 75) **4**	**Rango semi-intercuartílico** $\frac{Q_3 - Q_1}{2} = \frac{4-3}{2} = 0'5$

Los valores anteriores, relativos a percentiles, son válidos si la variable es DISCRETA. En el supuesto de tratarse de una variable CONTINUA (con datos no agrupados), deberíamos entender que el valor identifica el intervalo situado a la izquierda en la siguiente tabla :

Intervalo	x	n
[1'5,2'5)...	2	6
[2'5,3'5)...	3	15
[3'5,4'5)...	4	10
[4'5,5'5]...	5	9
		40

N	P
6	15
21	52'5
31	77'5
40	100

Los percentiles pedidos se obtendrían del modo siguiente :

Mediana	**en [2'5,3'5)**	$\mathbf{Me = P_{50} = 2'5 + \frac{\frac{50.40}{100} - 6}{15}.1 = 3'433}$
Percentil 82	**en [4'5,5'5]**	$\mathbf{P_{82} = 4'5 + \frac{\frac{82.40}{100} - 31}{9}.1 = 4'700}$
Cuartil 1º	**en [2'5,3'5)**	$\mathbf{Q_1 = P_{25} = 2'5 + \frac{\frac{25.40}{100} - 6}{15}.1 = 2'767}$
Cuartil 3º	**en [3'5,4'5)**	$\mathbf{Q_3 = P_{75} = 3'5 + \frac{\frac{75.40}{100} - 21}{10}.1 = 3'400}$

4

Interv.	n
[10,12)	5
[12,14)	11
[14,16)	19
[16,18)	21
[18,20]	4

De la distribución de la izquierda, calcular :

Media, varianza y desviación típica.
Moda
Mediana, Percentil 59 y Decil 3º.
Desviación media.
Coeficientes de asimetría y curtosis.

Interv.	n	a
[10,12)	5	11
[12,14)	11	13
[14,16)	19	15
[16,18)	21	17
[18,20]	4	19
	60	

N	P
5	8'333
16	26'667
35	58'333
56	93'333
60	100'000

n.a	n.a²
55	605
143	1859
285	4275
357	6069
76	1444
916	**14252**

Media	Varianza	Desviación típica
$\bar{x} = \frac{\sum n_i . a_i}{N} = \frac{916}{60} = 15'2667$	$\sigma^2 = \frac{\sum n_i . a_i^{\;2}}{N} - \bar{x}^2 = \frac{14252}{60} - 15'2667^2 = 4'$	$\sigma = \sqrt{4'4622} = 2'1124$

Moda	en [16,18)	$\mathbf{Mo = 16 + \frac{4}{4+19} . 2 = 16'3478}$
Mediana (percentil 50)	en [14,16)	$\mathbf{Me = P_{50} = 14 + \frac{\frac{50.60}{100} - 16}{19} . 2 = 15'4737}$
Percentil 59	en [16,18)	$\mathbf{P_{59} = 16 + \frac{\frac{59.60}{100} - 35}{21} . 2 = 16'0381}$
Decil 3º (percentil 30)	en [14,16)	$\mathbf{D_3 = P_{30} = 14 + \frac{\frac{30.60}{100} - 16}{19} . 2 = 14'2105}$

Desviación media

$\lvert x - \bar{x} \rvert$	$n.\lvert x - \bar{x} \rvert$
4'2667	21'3333
2'2667	24'9333
0'2668	5'0668
1'7333	36'4000
3'7333	14'9333
	102'6667

Asimetría y Curtosis

$x - \bar{x}$	$n.(x - \bar{x})^3$	$n.(x - \bar{x})^4$
-4'2667	-388'3615	1657'0090
-2'2667	-128'1019	290'3644
-0'2668	-0'3603	0'0961
1'7333	109'3618	189'5604
3'7333	208'1375	777'0466
	-199'3244	**2914'0765**

Desviación media	$D = \frac{\sum n_i . \lvert x_i - \bar{x} \rvert}{N} = \frac{102'6667}{60} = 1'7111$
Asimetría (-0'3524 < 0) Algo asimétrica hacia la izquierda	$As_1 = \frac{\frac{\sum n_i . (x_i - \bar{x})^3}{N}}{\sigma^3} = \frac{\frac{-199'3244}{60}}{2'1124^3} = -0'3524$
Curtosis (-0'5608 < 0) Ligeramente aplanada (Platicúrtica)	$K = \frac{\frac{\sum n_i . (x_i - \bar{x})^4}{N}}{\sigma^4} - 3 = \frac{\frac{2914'0765}{60}}{2'1124^4} - 3 = -0'5608$

5

La distribución de las estaturas en centímetros de los alumnos de un centro, expresados en porcentajes, es la siguiente:

Estaturas	Porcentajes
Menos de 150	**0'3**
De 150 a 154	**1'6**
De 155 a 159	**9'4**
De 160 a 164	**20'5**
De 165 a 169	**31'5**
De 170 a 174	**22'5**
De 175 a 179	**10'7**
De 180 y más	**3'5**

a) Siendo abiertos los intervalos primero y el último, ¿ qué valores sería razonable considerar para los límites extremos de esos intervalos ?
b) Si suponemos que en el Centro hay 1200 alumnos, ¿ cuáles serían las frecuencias absolutas?
c) Calcular la estatura media y la desviación típica.
d) ¿ Entre qué estaturas se encuentra la quinta parte de las estaturas centrales ?.

a)
Al referirse a intervalos de 5 cm. de amplitud en los restantes casos, debemos considerar que el primer intervalo es de *145 a menos de 150* y, el último, *de 180 a 185*.

b)

Estaturas	p	n = p . 1200 / 100	n
[145,150)	0'3	3'6	**4**
[150,155)	1'6	19'2	**19**
[155,160)	9'4	112'8	**113**
[160,165)	20'5	246	**246**
[165,170)	31'5	378	**378**
[170,175)	22'5	270	**270**
[175,180)	10'7	128'4	**128**
[180,185)	3'5	42	**42**
			N=1200

P	N
0'3	4
1'9	23
11'3	136
31'8	382
63'3	760
85'8	1030
96'5	1158
100'0	1200

c)

Estaturas	n	x	n.x	n.x²
[145,150)	4	147'5	590'0	87025'00
[150,155)	19	152'5	2897'5	441868'75
[155,160)	113	157'5	17797'5	2803106'25
[160,165)	246	162'5	39975'0	6495937'50
[165,170)	378	167'5	63315'0	10605262'50
[170,175)	270	172'5	46575'0	8034187'50
[175,180)	128	177'5	22720'0	4032800'00
[180,185)	42	182'5	7665'0	1398862'50
	1200		201535'0	33899050'00

De aquí resulta :

$$\overline{\mathbf{x}} = \frac{\mathbf{201535}}{\mathbf{1200}} = \mathbf{167'95}$$

$$\mathbf{s_x^2} = \frac{\mathbf{33899050}}{\mathbf{1200}} - \mathbf{167'95^2} = \mathbf{42'006} \qquad \mathbf{s_x} = \sqrt{\mathbf{42'006}} = \mathbf{6'481}$$

d)
La quinta parte representa el 20%. Con relación al centro (50%), cubrirán desde el 40% al 60%.
Se nos pide que calculemos los percentiles 40 y 60 de la distribución de estaturas.

La tabla de porcentajes acumulados del apartado **b)** nos permite deducir que :
Los percentiles 40 y 60 se encuentran en el intervalo [165,170) .

Sus valores concretos son :

$$P_{40} = e_i + \frac{\frac{40.N}{100} - N_{i-1}}{n_i}.a_i = 165 + \frac{\frac{40.1200}{100} - 382}{378}.5 = 166'963$$

$$P_{60} = e_i + \frac{\frac{60.N}{100} - N_{i-1}}{n_i}.a_i = 165 + \frac{\frac{60.1200}{100} - 382}{378}.5 = 169'471$$

6

Partiendo de la siguiente distribución de frecuencias acumuladas, determinar la media, mediana y moda de la siguiente distribución de edades. Analice la relación entre ellas.

Edad	N
[10,12)	**4**
[12,14)	**11**
[14,16)	**24**
[16,18)	**34**
[18,20]	**40**

Calculemos los parámetros pedidos, con el fin de observar en qué medida se verifica la relación $\bar{x} - Mo = 3.(\bar{x} - Me)$

Para obtener las frecuencias absolutas, a partir de las acumuladas, aplicamos el concepto que define a estas últimas. En la práctica, las frecuencias absolutas se obtienen restando la correspondiente acumulada de la anterior.

Edad	N	n	x	n.x	n.x^2
[10,12)	4	4	11	44	484
[12,14)	11	7	13	91	1183
[14,16)	24	13	15	195	2925
[16,18)	34	10	17	170	2890
[18,20]	40	6	19	114	2166
		40		614	9648

$\bar{x} = \frac{614}{40} = 15'35$	Lugar que ocupa la mediana : L = 50 . 40 / 100 = 20 La mediana está en [14,16) : $Me = 14 + \frac{20-11}{13}.2 = 15'3846$	La moda se encuentra en [14 , 16). Su valor concreto es : $Mo = 14 + \frac{10}{10+7}.2 = 15'1765$

Comprobemos la relación existente entre ellas :

$$\bar{x} - Mo = 15'35 - 15'1765 = 0'1735$$
$$3.(\bar{x} - Me) = 3.(15'35 - 15'3845) = -0'1035$$

No se verifica la relación esperada, si bien la diferencia no es muy grande.

Esta relación teórica sólo se verifica en situaciones ideales y excepcionales (por ejemplo en distribuciones simétricas, donde $\bar{x} = Mo = Me$).

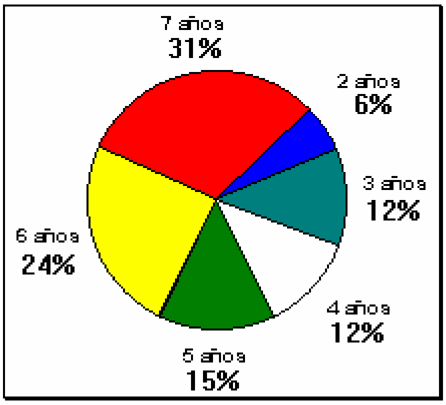
7

Completar la tabla de frecuencias siguiente :

Nº de suspensos	n	N
0	3	
1		10
2	12	
3		30
4		
N =	50	

Nº de suspensos	n	N	
0	3	**3**	coincide con el valor de n
1	**7**	10	para que al acumular resulte N=10
2	12	**22**	acumulando 12
3	**8**	30	para que al acumular resulte N=30
4	**20**	**50**	Última acumulada =N=50 y n=20 por diferencia con la anterior

8



Calcular la amplitud semi-intercuartílica de la distribución de las edades de 400 niños, representada a la izquierda.

Conocidos los porcentajes y el total de observaciones (N=400), podemos construir la distribución de frecuencias absolutas :

n = p . N / 100

x	p	n	P		
2	6	24	6		
3	12	48	18		
4	12	48	30	⇐	Primer cuartil (percentil 25)
5	15	60	45		
6	24	96	69		
7	31	124	100	⇐	Tercer cuartil (percentil 75)
		400			

La amplitud o rango semi-intercuartílico será pues : $\dfrac{\mathbf{Q_3 - Q_1}}{\mathbf{2}} = \dfrac{\mathbf{7-4}}{\mathbf{2}} = \mathbf{1'5}$

9

Una variable X tiene por media 12 y desviación típica 3. Si elevamos todos los valores al cuadrado construimos la nueva variable Y = X^2 . ¿ Cuál es el valor de su media aritmética ?.

Observemos la expresión de la varianza : $s_x^2 = \dfrac{\sum_{i=1}^{n} n_i . x_i^2}{N} - \overline{x}^2$

La primera parte de la expresión contiene los cuadrados de los valores de la variable X; es decir, los valores definidos como la nueva variable Y.

Con esto : $s_x^2 = \dfrac{\sum_{i=1}^{n} n_i . y_i}{N} - \overline{x}^2 \Rightarrow s_x^2 = \overline{y} - \overline{x}^2 \Rightarrow \overline{y} = s_x^2 + \overline{x}^2 = 3^2 + 12^2 = 153$

10

Una variable X tiene como media 8 y varianza 4. ¿ Qué transformación lineal hemos de realizar con ella, para obtener una nueva variable Y que tenga por media 42 y desviación típica 10 ?.

Se entiende por transformación lineal a una relación del tipo : **Y = a + b.X**
Hemos de calcular los parámetros **a** y **b** desconocidos.

Haciendo uso de las propiedades de la media y la desviación típica, resulta :
Sobre la media $\mathbf{\overline{Y} = a + b.\overline{X} \Rightarrow 42 = a + b.8}$
En relación con la desviación típica $\mathbf{s_Y = b.s_X \Rightarrow 10 = b.2 \Rightarrow b = 5 \Rightarrow a = 42 - 5.8 = 2}$

La transformación realizada fue : **Y = 2 + 5.X**

11

Las calificaciones de un alumno en dos test de conocimientos fueron 5'4 y 41. El primer test dio como media 5 con varianza 2 y, el segundo, media 38 con varianza 12.
¿ En qué test obtuvo mejor calificación con relación al grupo total de alumnos ?.

Nos encontramos con dos distribuciones de calificaciones medidas en distintas escalas. Para poder comparar tendremos que referir ambas series de valores a otras equivalentes entre sí (igual media y desviación típica).

El proceso de tipificación nos proporciona lo que deseamos (siempre obtendremos una distribución con media 0 y desviación típica 1).

Tipificando ambas calificaciones se obtiene :

Nota del test 1º : $\mathbf{5'4} \quad \rightarrow \quad \mathbf{z_1 = \dfrac{5'4 - 5}{\sqrt{2}} = 0'283}$

Nota del test 2º : $\mathbf{41} \quad \rightarrow \quad \mathbf{z_1 = \dfrac{41 - 38}{\sqrt{12}} = 0'866}$

La nota obtenida en el segundo test es superior a la del primero en términos comparativos.

12

Estatura en cm.	Alumnos
[140,145)	**12**
[145,150)	**35**
[150,155)	**51**
[155,160)	**?**
[160,165)	**7**

a) Determinar la frecuencia desconocida, sabiendo que la estatura media es de 151’5 cm.
b) Calcule la amplitud semi-intercuartílica.
c) Moda de la distribución y coeficiente de asimetría que la utiliza.
d) Percentil correspondiente a una estatura de 153 cm.. Explique su significado.
e) ¿ Entre qué estaturas se encuentran las 25 centrales ?.
f) Porcentaje de alumnos que miden más de 157 cm.

a)

	x	n	n.x
[140,145)	142’5	12	1710
[145,150)	147’5	35	5162’5
[150,155)	152’5	51	7777’5
[155,160)	157’5	**f**	157'5.f
[160,165)	162’5	7	1137’5
		105+f	15787’5+157'5.f

La tabla de cálculos de la media conduce a :

$$151'5 = \frac{15787'5 + 157'5.f}{105 + f}$$

Resolviendo deducimos que : f = 20

b)

	n	N
[140,145)	12	12
[145,150)	35	47
[150,155)	51	98
[155,160)	20	118
[160,165)	7	125
	N=125	

Lugar Q_1 = 125 . 25 / 100 = 31’25
Q_1 se encuentra en [145,150)

$$Q_1 = 145 + \frac{31'25 - 12}{35}.5 = 147'75$$

Lugar Q_3 = 125 . 75 / 100 = 93’75
Q_3 se encuentra en [150,155)

$$Q_3 = 150 + \frac{93'75 - 47}{51}.5 = 154'5833$$

Luego : $Q = \dfrac{Q_3 - Q_1}{2} = \dfrac{154'5833 - 147'75}{2} = 3'4167$

c) 1º) Moda en [150,155) : $Mo = 150 + \dfrac{20}{35+20}.5 = 151'8182$

x	n	n.x	n.x^2
142’5	12	1710	243675
147’5	35	5162’5	761468’75
152’5	51	7777’5	1186068’75
157’5	20	3150	496125
162’5	7	1137’5	184843’75
	125	18937’5	2872181’25

$$s = \sqrt{\frac{2872181'25}{125} - 151'5^2}$$

$$s = 5'02$$

$$As = \frac{\bar{x} - Mo}{s} = -0'0634$$

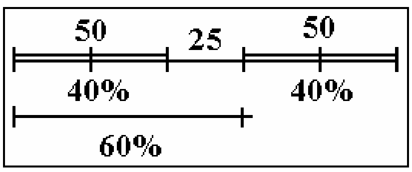
d) 153 se encuentra en [150,155)

	n	N
[140,145)	12	12
[145,150)	35	47
[150,155)	51	98
[155,160)	20	118
[160,165)	7	125
	N=125	

$$P_k = 150 + \frac{k.\frac{125}{100} - 47}{51}.5 = 153$$

Resolviendo : k = 62’08 ≈ 62

e)



Lugar = 125 . 40 / 100 = 50 ; en [150,155) :

$$P_{40} = 150 + \frac{50 - 47}{51}.5 = 150'29$$

Lugar = 125 . 60 / 100 = 75 ; en [150,155) :

$$P_{60} = 150 + \frac{75 - 47}{51}.5 = 152'75$$

Entre 150’29 y 152’75

f) 157 se encuentra en [155,160)

$$P_k = 155 + \frac{k.\frac{125}{100} - 98}{20}.5 = 157$$

Resolviendo : k = 84’8% (porcentaje inferiores a 157)

Luego, miden más de 157 cm. : 100% - 84’8% = 15’2%

13

Edad	Hombres	Mujeres
22 a 25	**7**	**3**
19 a 22	**9**	**5**
16 a 19	**5**	**6**
13 a 16	**11**	**9**
10 a 13	**8**	**2**

a) Determine el número de hombres con edades comprendidas entre los 11 y 15 años.
b) ¿ Cuál de los dos grupos de edades está más disperso ?.
c) Con relación al grupo integrado por los del mismo sexo, ¿quién resulta más joven, un hombre o una mujer de 20 años ?.

	Hombre					Mujer		
	x	n	N	n.x	n.x²	n	n.y	n.y²
[10,13)	11'5	8	8	92	1058	2	23	264'5
[13,16)	14'5	11	19	159'5	2312'75	9	130'5	1892'25
[16,19)	17'5	5	24	87'5	1531'25	6	105	1837'5
[19,22)	20'5	9	33	184'5	3782'25	5	102'5	2101'25
[22,25)	23'5	7	40	164'5	3865'75	3	70'5	1656'75
		40		688	12550	25	431'5	7752'25

a) 11 pertenece al intervalo [10,13) : $P_k = 10 + \dfrac{k.\frac{40}{100} - 0}{8}.3 = 11 \quad \Rightarrow \quad k = 6'67\%$

15 pertenece al intervalo [13,16) : $P_k = 13 + \dfrac{k.\frac{40}{100} - 8}{11}.3 = 15 \quad \Rightarrow \quad k = 38'33\%$

Entre 11 y 15 el 38'33-6'67 = 31'66%. Luego hay : 40 . 31'66 / 100 = 12'664 ≈ 13 hombres

b) Calculamos las varianzas de ambos grupos :

$$\bar{x} = \frac{688}{40} = 17'2 \quad ; \quad s_x^2 = \frac{12550}{40} - 17'2^2 = 17'91 \quad ; \quad s_x = \sqrt{17'91} = 4'232$$

$$\bar{y} = \frac{431'5}{25} = 17'26 \quad ; \quad s_y^2 = \frac{7752'25}{25} - 17'26^2 = 12'1824 \quad ; \quad s_y = \sqrt{12'1824} = 3'49$$

Siendo 17'91 > 12'1824 ⇒ Grupo hombres más disperso *de forma aboluta*

Pese a ser las medias prácticamente iguales, debemos emplear el coeficiente de variación para estudiar la *variabilidad relativa* de ambos grupos :

$$CV_x = \frac{4'232}{17'2}.100 = 24'605\% \quad ; \quad CV_y = \frac{3'49}{17'26}.100 = 20'220\%$$ ⇒ hombres más disperso

c) Tipificamos 20 en ambos grupos :

$$Z_{\text{hombre}} = \frac{20 - 17'2}{\sqrt{17'91}} = 0'662 \quad ; \quad Z_{\text{mujer}} = \frac{20 - 17'26}{\sqrt{12'1824}} = 0'785$$

Como 0'662 < 0'785 ⇒ Hombre más joven

14

La tabla siguiente nos muestra las calificaciones de 10 alumnos, en un test de cálculo matemático, al inicio del curso y al finalizar el mismo.

Alumno	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Inicio	**4**	**5**	**1**	**5**	**2**	**3**	**2**	**1**	**1**	**3**
Final	**6**	**8**	**5**	**9**	**3**	**6**	**7**	**6**	**4**	**9**

a) Determine la media, desviación típica, mediana y moda de las calificaciones al inicio y al final del curso.
b) Calcule la media y desviación típica del incremento o mejora de la calificación obtenida.

a)

Inicio	x	4	5	1	5	2	3	2	1	1	3	27
	x^2	16	25	1	25	4	9	4	1	1	9	95

$$\overline{x} = \frac{27}{10} = 2'7 \quad ; \quad s_x = \sqrt{\frac{95}{10} - 2'7^2} = 1'487$$

Ordenando valores :

1	1	1	2	2	3	3	4	5	5

Mediana = 2'5 Moda = 1

Final	y	6	8	5	9	3	6	7	6	4	9	63
	y^2	36	64	25	81	9	36	49	36	16	81	433

$$\overline{y} = \frac{63}{10} = 6'3 \quad ; \quad s_y = \sqrt{\frac{433}{10} - 6'3^2} = 1'9$$

Ordenando valores :

3	4	5	6	6	6	7	8	9	9

Mediana = 6 Moda = 6

b)

Mejora	d	2	3	4	4	1	3	5	5	3	6	36
	d^2	4	9	16	16	1	9	25	25	9	36	150

$$\overline{d} = \frac{36}{10} = 3'6 \quad ; \quad s_d = \sqrt{\frac{150}{10} - 3'6^2} = 1'428$$

Media de la diferencia : $\overline{d} = \overline{y} - \overline{x} = 6'3 - 2'7 = 3'6$ (No es válido para dispersiones)

15

Nº Suspensos	Alumnos
0	16
1	20
2	14
3	15
4	10
5	5

a) Determine la media, desviación típica, coeficiente de variación, mediana y moda del número de suspensos.
b) Coeficiente de asimetría de Fisher.
c) Puntuación diferencial y tipificada correspondiente a 2 suspensos.

a) De la siguiente tabla de cálculos obtenemos :

$$\bar{x} = \frac{158}{80} = 1'975 \qquad s = \sqrt{\frac{496}{80} - 1'975^2} = 1'5164 \qquad CV = \frac{1'5164}{1'975}.100 = 76'78\%$$

<u>Mediana :</u> N/2 = 40 $\Rightarrow$ Me = 2   <u>Moda</u> = 1

x	n	N	n.x	n.x^2	$x - \bar{x}$	$n.(x-\bar{x})^3$
0	16	16	0	0	-1’975	-123’2598
1	20	36	20	20	-0’975	-18’5372
2	14	50	28	56	0’025	0’0002
3	15	65	45	135	1’025	16’1534
4	10	75	40	160	2’025	83’0377
5	5	80	25	125	3’025	138’4032
	80		158	496		95’7975

b)

$$As = \frac{\frac{\sum n.(x-\bar{x})^3}{N}}{s^3} = \frac{\frac{95'7975}{80}}{1'5164^3} = 0'3434$$

Ligeramente asimétrica a la derecha (o positiva)

c)

$$x = 2$$

$$d = x - \bar{x} = 2 - 1'975 = 0'025$$

$$z = \frac{x - \bar{x}}{s} = \frac{0'025}{1'5164} = 0'016$$

16

Estatura	Niños
155-159	**4**
150-154	**13**
145-149	**12**
140-144	**A**
135-139	**2**
130-134	**1**

La altura en cm. de los niños de 12 años, examinados durante la última semana en la unidad de crecimiento del centro hospitalario “Crecebien”, viene representada en la tabla de la izquierda. Sabiendo que la altura media de los mismos es 147’75 cm., calcular :

a) La frecuencia A del tercer intervalo.

b) La simetría de la distribución a partir de la comparación de media, mediana y moda.

c) El percentil correspondiente a un niño que mide 1’43 m..

x	n	n.x
132	1	132
137	2	274
142	A	142.A
147	12	1764
152	13	1976
157	4	628
TOTAL	**32+A**	**4774+142.A**

a)

$$\mathbf{\bar{x} = 147'75 = \frac{4774 + 142.A}{32 + A}}$$

Resolviendo la ecuación anterior obtenemos el valor de A :

147’75.(32+A)=4774+142.A $\rightarrow$

$\rightarrow$ 4728+147’75.A=4774+142.A $\rightarrow$

$\rightarrow$ 5’75.A = 46 $\rightarrow$ **A = 8**

b) Calculemos la mediana y la moda de la distribución :

Intervalos	n	N
[129’5 , 134’5)	1	1
[134’5 , 139’5)	2	3
[139’5 , 144’5)	8	11
[144’5 , 149’5)	12	**23**
[149’5 , 154’5)	**13**	36
[154’5 , 159’5)	4	40

Moda en [149’5 , 154’5) :

$$\mathbf{Mo = 149'5 + \frac{4}{4+12}.5 = 150'75}$$

Lugar que ocupa la mediana = 40/2 = 20

Mediana en [144’5 , 149’5) : $\mathbf{Me = 144'5 + \frac{20-11}{12}.5 = 148'25}$

Utilizando los coeficientes de asimetría : $\mathbf{As_2 = \frac{\bar{x} - Mo}{s}}$ $\quad$ $\mathbf{As_3 = \frac{3.(\bar{x} - Me)}{s}}$

y siendo siempre positiva la desviación típica ,concluiremos que la simetría resultará del análisis del signo del numerador.

$$\mathbf{\bar{x} - Mo = 147'75 - 150'75 = -3 \quad < 0}$$

$$\mathbf{3.(\bar{x} - Me) = 3.(147'75 - 148'25) = -1'5 \quad < 0}$$

Luego es asimétrica izquierda (o negativa).

c) La altura 1’43 m. (= 143 cm.) se encuentra en el intervalo [139’5 , 144’5) :

$$\mathbf{P_k = 143 = 139'5 + \frac{\frac{k.40}{100} - 3}{8}.5 \quad \Rightarrow \quad 3'5 = \frac{0'4.k - 3}{8}.5 \quad \Rightarrow \quad \frac{3'5.8}{5} + 3 = 0'4.k \quad \Rightarrow \quad k = \frac{8'6}{0'4} = 21'5}$$

Luego corresponde al percentil 21’5.

17

X	n
10-12	**10**
7-9	**100**
4-6	**60**
1-3	**30**

Dada la siguiente distribución de frecuencias., calcular :

a) Media y desviación típica.

b) Número de observaciones comprendidas entre las puntuaciones directas 3’5 y 9’5.

c) Puntuaciones típicas de los percentiles 20 y 80.

Ordenamos los intervalos de menor a mayor, expresándolos mediante sus extremos reales.

Intervalos	n	x	n.x	n.x^2	N
[0’5 , 3’5)	30	2	60	120	30
[3’5 , 6’5)	60	5	300	1500	90
[6’5 , 9,5)	100	8	800	6400	190
[9’5 , 12’5]	10	11	110	1210	200
Totales	**200**		**1270**	**9230**	

a) $\overline{x} = \frac{1270}{200} = 6'35 \qquad s^2 = \frac{9230}{200} - 6'35^2 = 5'8275 \qquad s = \sqrt{5'8275} = 2'414$

b) De la observación directa de la tabla se concluye que es 160 (60+100).

c) Percentil 20 : Lugar = 20 x 200 / 100 = 40 (Observando N) se encuentra en [3’5 , 6’5)

$$P_{20} = 3'5 + \frac{40-30}{60}.3 = 4 \quad \rightarrow \quad z = \frac{4-6'35}{2'414} = -0'9735$$

Percentil 80 : Lugar = 80 x 200 / 100 = 160 (Observando N) se encuentra en [6’5 , 9,5)

$$P_{80} = 6'5 + \frac{160-90}{100}.3 = 8'6 \quad \rightarrow \quad z = \frac{8'6-6'35}{2'414} = 0'9321$$

18

x	n
0	**6**
1	**12**
2	**21**
3	**11**

Haciendo uso de coeficientes basados en medidas de posición, estudie la asimetría y el apuntamiento de la distribución.

Tales coeficientes son el de asimetría de Yule y el de curtosis de Kelley.

Obtengamos los percentiles que intervienen en su cálculo a través de la columna de porcentajes acumulados (P) :

x	n	r	p	P
0	6	0'12	12	12
1	12	0'24	24	36
2	21	0'42	42	78
3	11	0'22	22	100
	50			

Cuartil 1º : (25%) 1
Cuartil 3º : (75%) 2
Mediana : (50%) 2

Percentil 10 : (10%) 0
Percentil 90 : (90%) 3

Con ellos :

$$Y = \frac{Q_3 - 2.Me + Q_1}{Q_3 - Q_1} = \frac{2 - 2.2 + 1}{2 - 1} = -1$$

(asimétrica a la izquierda o negativa)

$$K = \frac{Q}{P_{90} - P_{10}} - 0'263 = \frac{\frac{Q_3 - Q_1}{2}}{P_{90} - P_{10}} - 0'263 = \frac{\frac{2-1}{2}}{3 - 0} - 0'263 = -0'0963$$

(ligeramente platicúrtica o aplastada)

19

Determine las medias aritmética, geométrica y armónica de la variable X que toma los valores siguientes :

5 , 1 , 5 , 4 , 8.

Media aritmética : $\bar{x} = \frac{\sum x_i}{N} = \frac{5+1+5+4+8}{5} = \frac{23}{5} = 4'6$

Media geométrica : $\bar{x}_G = \sqrt[N]{x_1 . x_2 . \ ... \ . x_N} = \sqrt[5]{5.1.5.4.8} = \sqrt[5]{800} = 800^{1/5} = 800^{0'2} = 3'807$

Media armónica : $\bar{x}_A = \frac{N}{\sum\left(\frac{1}{x_i}\right)} = \frac{5}{\frac{1}{5}+\frac{1}{1}+\frac{1}{5}+\frac{1}{4}+\frac{1}{8}} = \frac{5}{1'775} = 2'817$

20

x	n
1	3
2	10
3	7
	20

Determine las medias aritmética, geométrica y armónica de la distribución.

Generalizamos las expresiones correspondientes al figurar frecuencias :

Media aritmética :

$$\bar{x} = \frac{\sum n_i . x_i}{N} = \frac{3.1 + 10.2 + 7.3}{20} = \frac{44}{20} = 2'2$$

Media geométrica :

$$\bar{x}_G = \sqrt[N]{x_1^{n_1} . x_2^{n_2} .....x_n^{n_n}} = \sqrt[20]{1^3 . 2^{10} . 3^7} =$$

$$= \sqrt[20]{2239488} = 2239488^{1/20} = 2239488^{0'05} = 2'077$$

Media armónica :

$$\bar{x}_A = \frac{N}{\sum \left(\frac{n_i}{x_i} \right)} = \frac{20}{\frac{3}{1} + \frac{10}{2} + \frac{7}{3}} = \frac{20}{10'333} = 1'935$$

21

Con el fin de estudiar la edad media y la dispersión de edades en un centro educativo, el director solicita estos datos a los responsables de los distintos niveles, resultando :

- **200 alumnos de Primaria con media 11 años y varianza 2’5.**
- **140 alumnos de Secundaria con media 14’6 años y varianza 2.**
- **165 alumnos de Bachillerato con media 17’1 años y varianza 0’9.**

¿ Cuál es la edad media y la varianza del colectivo total de alumnos del centro ?.

Media conjunta de los 3 grupos

$$\overline{X} = \frac{\sum n_i . \overline{x}_i}{\sum n_i} = \frac{200.11 + 140.14'6 + 165.17'1}{200 + 140 + 165} = \frac{7065'5}{505} = 13'99$$

Varianza conjunta de los 3 grupos

$$S^2 = \frac{\sum n_i . S_i^2}{\sum n_i} + \frac{\sum n_i . \left(\overline{x}_i - \overline{X}\right)^2}{\sum n_i} =$$

$$= \frac{200.2'5 + 140.2 + 165.0'9}{505} + \frac{200.(11 - 13'99)^2 + 140.(14'6 - 13'99)^2 + 165.(17'1 - 13'99)^2}{505} =$$

$$= \frac{928'5}{505} + \frac{3436'0105}{505} = 1'839 + 6'804 = 8'643$$

22

De las 10 observaciones de dos variables X , Y, conocemos :

$\Sigma X = 114$; $\Sigma X^2 = 1410$; $\Sigma Y = 34$; $\Sigma Y^2 = 154$; $\Sigma XY = 398$.

Determine la media y varianza de la variable V = X - Y.

Calculemos la media y varianza de X, la media y varianza de Y, así como la covarianza.

$$\overline{X} = \frac{114}{10} = 11'4 \qquad \overline{Y} = \frac{34}{10} = 3'4 \qquad S_X^2 = \frac{1410}{10} - 11'4^2 = 11'04 \qquad S_Y^2 = \frac{154}{10} - 3'4^2 = 3'84$$

$$S_{XY} = \frac{\sum X_i . Y_i}{N} - \overline{X}.\overline{Y} = \frac{398}{10} - 11'4.3'4 = 1'04$$

Con ello :

$$\overline{V} = \overline{X} - \overline{Y} = 11'4 - 3'4 = 8$$

$$S_V^2 = S_X^2 + S_Y^2 - 2.S_{XY} = 11'04 + 3'84 - 2.1'04 = 12'8$$

23

El estudio de las faltas de asistencia a clase de alumnos de un grupo de 3º de Secundaria produjo los resultados siguientes :

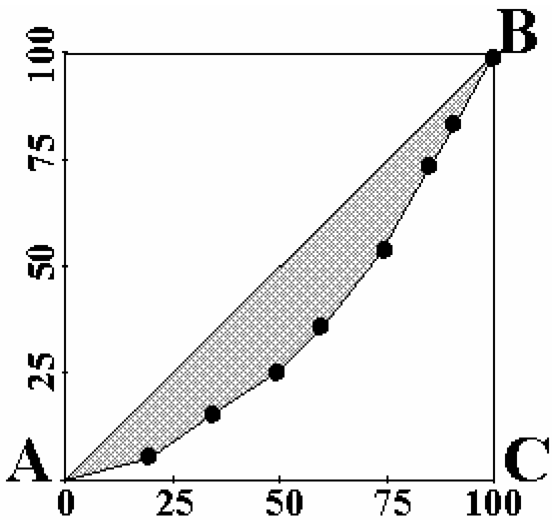
Faltas	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Alumnos	**4**	**3**	**3**	**2**	**3**	**2**	**1**	**2**

Determine la mediala y estudie analítica y gráficamente el grado de concentración de la distribución.

Los cálculos de la mediala, índice de Gini y curva de Lorenz, se obtienen a partir de la siguiente tabla auxiliar:

x_i	n_i	$N_i = \Sigma n_i.$	$P_i = (N_i.. /N).100$	$t_i = n_i. x_i$	$T_i = \Sigma t_i.$	$Q_i = (T_i.. /T).100$	$P_i - Q_i$
1	4	4	20	4	4	5'195	14'805
2	3	7	35	6	10	12'987	22'013
3	3	10	50	9	19	24'675	25'325
4	2	12	60	8	27	35'065	24'935
5	3	15	75	15	42	54'545	20'455
6	2	17	85	12	54	70'130	14'870
7	1	18	90	7	61	79'221	10'779
8	2	20	100	16	77	100	0
	N = 20		**TP = 515**	**T = 77**			**TD =133'182**

Uniendo el origen del rectángulo (0 , 0) con los sucesivos puntos (P_i , Q_i) obtenemos la **curva de Lorenz** de la derecha.



Las sumas TD y TP permiten obtener el **índice de Gini** :

$$G = \frac{TD}{TP - 100} = \frac{133'182}{515 - 100} = 0'3209$$

Concluimos la presencia de una cierta concentración (lo cuál también se advierte con la gráfica).

Mediala = 5

ya que el primer valor que iguala o supera a 50 en la columna Q_i es 54'545, el cuál corresponde a x = 5.

24

Un análisis del pago de impuesto en el sector de hostelería ofreció los resultados siguientes (importes mensuales por 10.000 pesetas) :

Importe	**[0,2)**	**[2,4)**	**[4,6)**	**[6,8)**	**[8,10)**	**[10,12]**
Empresas	**2**	**6**	**26**	**40**	**21**	**5**

Determine la mediala y estudie analítica y gráficamente el grado de concentración de la distribución.

Los cálculos de la mediala, índice de Gini y curva de Lorenz, se obtienen a partir de la siguiente tabla auxiliar:

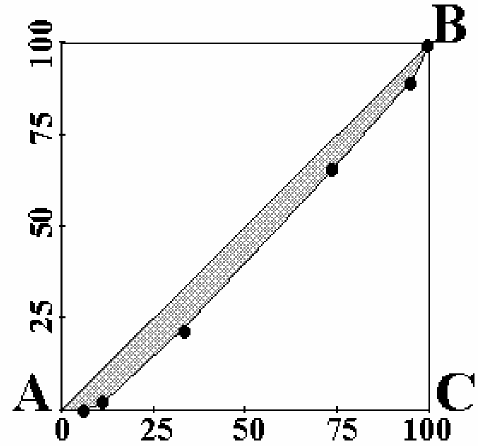
	x_i	n_i	$N_i = \Sigma\, n_i.$	$P_i = (N_i.. /N).100$	$t_i = n_i.\, x_i$	$T_i = \Sigma\, t_i.$	$Q_i = (T_i.. /T).100$	P_i - Q_i
[0,2)	1	2	2	2	2	2	0'297	1'703
[2,4)	3	6	8	8	18	20	2'967	5'033
[4,6)	5	26	34	34	130	150	22'255	11'745
[6,8)	7	40	74	74	280	430	63'798	10'202
[8,10)	9	21	95	95	189	619	91'840	3'160
[10,12]	11	5	100	100	55	674	100	0
		N =100		**TP = 313**	**T = 674**			**TD =31'843**

Con TD y TP obtenemos el **índice de Gini** :

$$G = \frac{TD}{TP - 100} = \frac{31'843}{313 - 100} = 0'1495$$

Concluimos que existe una concentración muy baja (lo cuál manifestará también la gráfica de Lorenz).

Uniendo el origen del rectángulo (0 , 0) con los sucesivos puntos (P_i , Q_i) obtenemos la **curva de Lorenz** de la derecha.



Mediala en el intervalo [6 , 8)

ya que el primer valor que iguala o supera a 50 en la columna Q_i es 63'798, el cuál corresponde al intervalo indicado.

De aquí :

$$Ml = e_i + \frac{50 - Q_{i-1}}{Q_i - Q_{i-1}}.a_i = 6 + \frac{50 - 22'255}{63'798 - 22'255}.2 = 7'3357$$

25

x	f
0	2
1	8
2	10
3	3
4	1

Haciendo uso del cálculo de momentos ordinarios de órdenes 1º al 4º, determine el valor de
la media, varianza, asimetría y curtosis de la distribución de la izquierda.

Tabla de cálculo de momentos ordinarios :

		a_1	a_2	a_3	a_4
x	**n**	**n.x**	**n.x²**	**n.x³**	**n.x⁴**
0	2	0	0	0	0
1	8	8	8	8	8
2	10	20	40	80	160
3	3	9	27	81	243
4	1	4	16	64	256
Totales :	**24**	**41**	**91**	**233**	**667**

Orden	$a_k = \sum \frac{n}{N}.x^k = \frac{\sum n.x^k}{N}$	m_k
1	$a_1 = \frac{41}{24} = 1'7083$	$m_1 = 0$
2	$a_2 = \frac{91}{24} = 3'7917$	$m_2 = a_2 - a_1^2 = 3'7917 - 1'7083^2 = 0'8734$
3	$a_3 = \frac{233}{24} = 9'7083$	$m_3 = a_3 - 3.a_2.a_1 + 2.a_1^3 = \ \ldots \ = 0'2468$
4	$a_4 = \frac{667}{24} = 27'7917$	$m_4 = a_4 - 4.a_3.a_1 + 6.a_2.a_1^2 - 3.a_1^4 = \ \ldots \ = 2'2954$

Con los momentos calculados :

Media $\mu = \overline{x} = a_1 = 1'7083$

Varianza $\sigma^2 = s_x^2 = m_2 = 0'8734$

Coeficiente de asimetría $As = \frac{m_3}{\left(\sqrt{m_2}\right)^3} = \frac{0'2468}{\left(\sqrt{0'8734}\right)^3} = 0'3024$

Coeficiente de curtosis $K = \frac{m_4}{m_2^{\ 2}} - 3 = \frac{2'2954}{0'8734^2} - 3 = 0'0091$

26

Haciendo uso del coeficiente de variación, compare la dispersión o variabilidad relativa de las dos variables descritas en cada uno de los apartados siguientes :

a) El peso medio de los toros de una ganadería es de 410 kg. con desviación típica de 1 kg. y, el peso medio de los perros de una granja es de 8 kg. con igual desviación típica.

b) Dos fábricas producen tornillos con igual longitud media (50 mm.), siendo la desviación típica de la primera de 2 mm. y de 12 mm. la de la segunda.

a) $CV_T = \frac{1}{410}.100 = 0'2439\%$ $\quad CV_P = \frac{1}{8}.100 = 12'5\%$ $\quad \Rightarrow$ El peso de los perros tiene mayor variabilidad

b) $CV_A = \frac{2}{50}.100 = 4\%$ $\quad CV_B = \frac{12}{50}.100 = 24\%$ $\quad \Rightarrow$ Los de la 2ª tienen mayor variabilidad

27

X	n_A	n_B
0-6	4	4
7-13	6	7
14-20	9	9
21-27	12	8
28-34	9	2

La tabla muestra la comprensión lectora (X) de dos grupos de sujetos educados en niveles socioculturales altos (A) y bajos (B). Si a partir de la puntuación X=19 se considera una comprensión lectora buena, calcular :

a) El porcentaje de personas en cada grupo con una buena comprensión lectora.

b) ¿ Cuál de los dos grupos presenta mayor variabilidad ? (Razone adecuadamente su respuesta).

Expresamos los intervalos con extremos reales, obteniendo la tabla de cálculos de percentiles, media y varianza de ambos grupos.

	x	n_A	N_A	$n_A.x$	$n_A.x^2$	n_B	N_B	$n_B.x$	$n_B.x^2$
[-0'5,6'5)	3	4	4	12	36	4	4	12	36
[6'5,13'5)	10	6	10	60	600	7	11	70	700
[13'5,20'5)	17	9	19	153	2601	9	20	153	2601
[20'5,27'5)	24	12	31	288	6912	8	28	192	4608
[27'5,34'5]	31	9	40	279	8649	2	30	62	1922
		40		792	18798	30		489	9867

a)

Calculemos el orden **k** del percentil que es igual a 19. Este nos da el porcentaje de los que tienen menos de 19 puntos, luego, como deseamos saber el porcentaje de los superiores a 19, la respuesta será su diferencia hasta 100.
El valor 19 se encuentra en el intervalo [13'5,20'5) :

En el grupo A :

$$P_k = 19 = 13'5 + \frac{\frac{k.40}{100} - 10}{9}.7 \quad \rightarrow \quad k = 42'68$$

Luego el 57'32% (100 - 42'68) tienen buena comprensión lectora en el grupo A.

En el grupo B :

$$P_k = 19 = 13'5 + \frac{\frac{k.30}{100} - 11}{9}.7 \quad \rightarrow \quad k = 60'24$$

Luego el 39'76% (100 - 60'24) tienen buena comprensión lectora en el grupo B.

b)

Mayor variabilidad la presentará aquel grupo que posea mayor dispersión entre sus valores. Con mayor rigor, si la media es representativa de las observaciones (no existen valores extremos exageradamente distanciados de la mayoría), es el coeficiente de variación el más adecuado para medir la variabilidad relativa entre dos series estadísticas (mayor coeficiente indica menor homogeneidad; un menor valor indicará menor dispersión o variabilidad).

Si comparamos mediante las varianzas :

$$\overline{X}_A = \frac{792}{40} = 19'8 \quad ; \quad S_A^2 = \frac{18798}{40} - 19'8^2 = 77'91 \quad ; \quad \overline{X}_B = \frac{489}{30} = 16'3 \quad ; \quad S_B^2 = \frac{9867}{30} - 16'3^2 = 63'21$$

el grupo A presenta una mayor variabilidad.

Si comparamos mediante los coeficientes de variación :

$$CV_A = \frac{S_A}{\overline{X}_A}.100 = \frac{\sqrt{77'91}}{19'8}.100 = 44'58\% \qquad CV_B = \frac{S_B}{\overline{X}_B}.100 = \frac{\sqrt{63'21}}{16'3}.100 = 48'78\%$$

luego, concluimos que el grupo B presenta una mayor variabilidad relativa (44'58 **<** 48'78), en contra de lo obtenido comparando varianzas.

28

X	n_A	n_B
0-6	4	4
7-13	6	7
14-20	9	9
21-27	12	8
28-34	9	2

La tabla muestra la comprensión lectora (X) de dos grupos de sujetos educados en niveles socioculturales altos (A) y bajos (B). Si a partir de la puntuación X=19 se considera una comprensión lectora buena, calcular :

a) El porcentaje de personas en cada grupo con una buena comprensión lectora.

b) ¿ Cuál de los dos grupos presenta mayor variabilidad ? (Razone adecuadamente su respuesta).

Expresamos los intervalos con extremos reales, obteniendo la tabla de cálculos de percentiles, media y varianza de ambos grupos.

	x	n_A	N_A	$n_A.x$	$n_A.x^2$	n_B	N_B	$n_B.x$	$n_B.x^2$
[-0'5,6'5)	3	4	4	12	36	4	4	12	36
[6'5,13'5)	10	6	10	60	600	7	11	70	700
[13'5,20'5)	17	9	19	153	2601	9	20	153	2601
[20'5,27'5)	24	12	31	288	6912	8	28	192	4608
[27'5,34'5]	31	9	40	279	8649	2	30	62	1922
		40		792	18798	30		489	9867

a)

Calculemos el orden **k** del percentil que es igual a 19. Este nos da el porcentaje de los que tienen menos de 19 puntos, luego, como deseamos saber el porcentaje de los superiores a 19, la respuesta será su diferencia hasta 100.

El valor 19 se encuentra en el intervalo [13'5,20'5) :

<u>En el grupo A :</u>

$$P_k = 19 = 13'5 + \frac{\frac{k.40}{100} - 10}{9}.7 \quad \rightarrow \quad k = 42'68$$

Luego el 57'32% (100 - 42'68) tienen buena comprensión lectora en el grupo A.

<u>En el grupo B :</u>

$$P_k = 19 = 13'5 + \frac{\frac{k.30}{100} - 11}{9}.7 \quad \rightarrow \quad k = 60'24$$

Luego el 39'76% (100 - 60'24) tienen buena comprensión lectora en el grupo B.

b)

Mayor variabilidad la presentará aquel grupo que posea mayor dispersión entre sus valores. Con mayor rigor, si la media es representativa de las observaciones (no existen valores extremos exageradamente distanciados de la mayoría), es el coeficiente de variación el más adecuado para medir la variabilidad relativa entre dos series estadísticas (mayor coeficiente indica menor homogeneidad; un menor valor indicará menor dispersión o variabilidad).

Si comparamos mediante las varianzas :

$$\overline{X}_A = \frac{792}{40} = 19'8 \quad ; \quad S_A^2 = \frac{18798}{40} - 19'8^2 = 77'91 \quad ; \quad \overline{X}_B = \frac{489}{30} = 16'3 \quad ; \quad S_B^2 = \frac{9867}{30} - 16'3^2 = 63'21$$

el grupo A presenta una mayor variabilidad.

Si comparamos mediante los coeficientes de variación :

$$CV_A = \frac{S_A}{\overline{X}_A}.100 = \frac{\sqrt{77'91}}{19'8}.100 = 44'58\% \qquad CV_B = \frac{S_B}{\overline{X}_B}.100 = \frac{\sqrt{63'21}}{16'3}.100 = 48'78\%$$

luego, concluimos que el grupo B presenta una mayor variabilidad relativa (44'58 **<** 48'78), en contra de lo obtenido comparando varianzas.

EJERCICIOS PROPUESTOS

1

Las edades de los alumnos que asisten a clase de repaso en una academia son las siguientes.

14	16	16	19	17	17	15	17	17	15
19	15	15	16	17	14	15	16	17	16
16	15	16	18	14	15	14	17	13	18
16	16	15	16	17	15	17	14	16	16
18	18	16	18	17	17	17	17	15	16

a) Construir la tabla completa de frecuencias.
b) Calcular la moda.
c) Determinar su media aritmética, varianza y desviación típica.
d) Obtener el valor de la mediana, del percentil 29 y de la amplitud semi-intercuartílica.

2

La tabla siguiente contiene los pesos en kg. de los alumnos de un curso.

40	43	58	48	47	41'5	40'5	43	47	52
51'5	57	43	44	56	44	50	50'5	46	42
44	40	45	50	50'5	49'5	41	55	58	51
50	45	43'5	45'5	53	59	39	40	38	39'5

a) Agrupar los valores en intervalos de 5 kg. de amplitud, comenzando por 35 kg., realizando un recuento de los mismos y confeccionando la tabla completa de frecuencias
b) Calcular la moda de dicha distribución de pesos.
c) Determinar su media aritmética, varianza y desviación típica.
d) Obtener el valor de la mediana, y del 8º decil.

3

Sea la siguiente distribución de frecuencias:

x	n
1	10
2	15
3	12
4	8

a) Calcular la media de esta distribución.
b) Si se suma a los valores de x_i la cantidad A, ¿qué relación guarda la media de la nueva distribución con la de la anterior ?. Generalizar este resultado y demostrar que si en una distribución de frecuencias de media m, se sustituyen los valores x_i por x_i + A, manteniendo las frecuencias, la media m' de la nueva distribución verifica : m'= A + m
c) Utilizando la igualdad obtenida, ¿cómo podría calcularse más fácilmente la media de la distribución siguiente ?

x	n
2752	36
2754	54
2756	24
2758	18

4

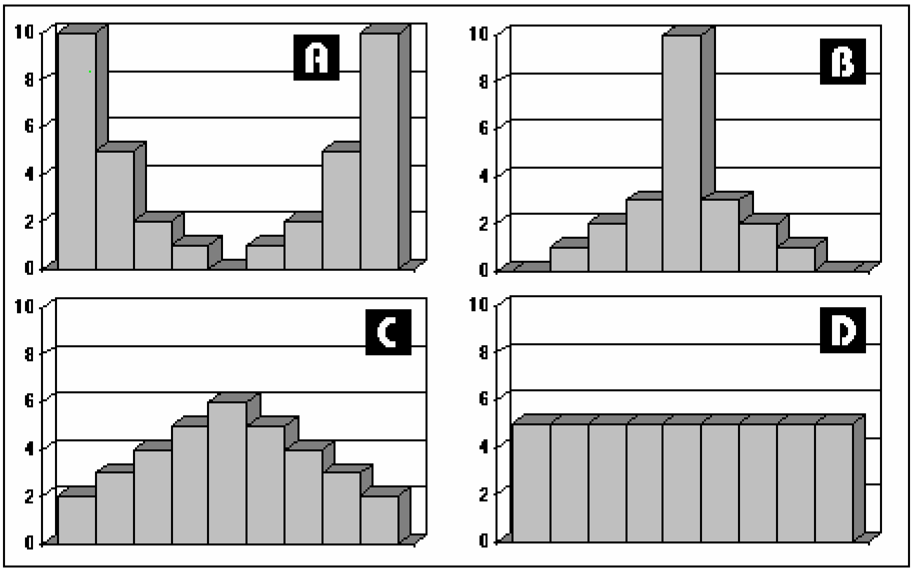
Una serie familias se han clasificado por su número de hijos, resultando :

Nº de hijos	0	1	2	3	4	5	6	7	8
Nº de familias	11	13	20	25	14	10	4	2	1

Se pide:
a) Calcular la tabla completa de frecuencias.
b) Representaciones gráficas.
c) Calcular la media, mediana y moda.
d) Hallar el recorrido, varianza y desviación típica.

5

Ordenar las cuatro distribuciones siguientes de mayor a menor dispersión.



6

Los precios de una chaqueta en once establecimientos fueron (en pts.):

5000 5200 5300 5600 6000 6400 6500 7200 7300 8400 9000

Calcular la desviación media respecto de la mediana y respecto de la media.

7

Si en una distribución de frecuencias duplicamos las amplitudes de los intervalos, ¿ qué sucederá, aproximadamente, con los valores de las frecuencias ?.

8

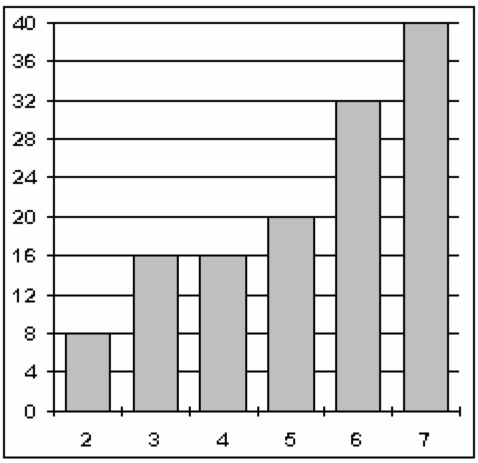
Represente el histograma correspondiente a la siguiente distribución de edades de los trabajadores de una fábrica.

Edades	Nº de trab.
de 20 a menos de 25	15
de 25 a menos de 35	20
de 35 a menos de 45	48
de 45 hasta 65	24

9

Ponga un ejemplo sencillo de una distribución de frecuencias simétrica. Calcule su moda, media y mediana, verificando que los tres parámetros coinciden.

10



A la izquierda se muestra el gráfico representativo de las frecuencias absolutas acumuladas de la distribución de edades de 40 individuos.

a) Obtenga su media, mediana y moda.

b) ¿ Cuántos tienen edades inferiores a cinco años y medio ?

11

Una variable X tiene como media 21 y varianza 9. Si se obtiene una nueva variable Y multiplicando los elementos de X por 4 y restándoles 8 unidades, ¿ cuál es el valor del coeficiente de variación de Y ?.

12

Una variable X toma los valores :

2 5 5 6 7

Realizada una transformación lineal con ella, se generó una nueva variable de la que conocemos que su media era 15 y que la puntuación X=2 se transformó en Y=13.
Calcule las cuatro puntuaciones Y desconocidas.

13

X	n
0	3
1	9
2	13
3	25
4	16
5	14

Estudie la simetría y el apuntamiento (curtosis) de la distribución de la izquierda.

NOTA :
Obtenga los distintos coeficientes conocidos. Compare los resultados.

14

Nota	**Alumnos**
9 - 10	2
7 - 8	0
5 - 6	4
3 - 4	14
1 - 2	12

La tabla de la izquierda nos muestra la distribución de calificaciones de los 32 alumnos de un curso.
a) Determine su media, mediana y moda.
b) ¿ Qué porcentaje de observaciones tienen nota inferior a 1’62 ?.
c) ¿ Entre qué valores se encuentra el 70% de las notas centrales ?
d) Obtenga el coeficiente de variación y la amplitud semi-intercuartílica.

15

Nota	n	N
[0 , 1)		1
[1 , 2)		1
[2 , 3)		5
[3 , 4)	3	
[4 , 5)		11
[5 , 6)	6	
[6 , 7)		19
[7 , 8]		

De la distribución de notas de 20 alumnos, calcular :
a) Frecuencias absolutas simples (f) y acumuladas (F) que faltan en la tabla.
b) Coeficiente de variación.
c) Porcentaje de alumnos con notas inferiores a 2'6.
d) ¿ Entre qué notas se encuentra el 10% de las calificaciones centrales ?.
e) Momentos ordinarios y centrales hasta el 4º orden.
f) Coeficientes de asimetría y curtosis, utilizando los momentos calculados en e).

16

Con el fin de estudiar la distribución de fallos en una pieza de tela, se realizó un recuento de los contenidos en cada metro. Los resultados fueron los siguientes :

Fallos	0	1	2	3	4	5	6	7	8	9
Nº de metros	25	8	4	1	1	1	2	1	3	4

a) Estudie el grado de concentración de la distribución de fallos a lo largo de la pieza de tela.
b) Calcule su media y su mediala.

17

La tabla siguiente muestra los fallos cometidos por alumnos en la realización de un test de 120 items.

Errores	[0 , 10)	[10 , 20)	[20 , 30)	[30 , 40)	[40 , 50)	[50 , 60)	[60 , 70)	[70 , 80)
Alumnos	25	20	22	16	29	24	38	26

a) Estudie el grado de concentración de la distribución de preguntas con respuesta errónea.
b) Calcule su mediala.

SOLUCIONES DE LOS EJERCICIOS PROPUESTOS

1

a)

x	n	r	p	N	R	P
13	1	0'02	2	1	0'02	2
14	5	0'10	10	6	0'12	12
15	10	0'20	20	16	0'32	32
16	14	0'28	28	30	0'60	60
17	13	0'26	26	43	0'86	86
18	5	0'10	10	48	0'96	96
19	2	0'04	4	50	1'00	100

b) Mo = 16

c) $\overline{\mathbf{x}}$ = 16'12 ; s^2 = 1'7856 ; s = 1'3363

d) Me = 16 ; P_{29} = 15 ; Q = 1

2

a)

Intervalo	n	r	p	N	R	P
[35,40)	3	0'075	7'5	3	0'075	7'5
[40,45)	14	0'350	35'0	17	0'425	42'5
[45,50)	8	0'200	20'0	25	0'625	62'5
[50,55)	9	0'225	22'5	34	0'850	85'0
[55,60]	6	0'150	15'0	40	1'000	100'0

b) Mo = 43'636

c) $\overline{\mathbf{x}}$ = 47'625 ; s^2 = 36'859 ; s = 6'071

d) Me = 46'875 ; D_8 = 53'889

3

a) $\overline{\mathbf{x}}$ = 2'4

b) 2'4 + A

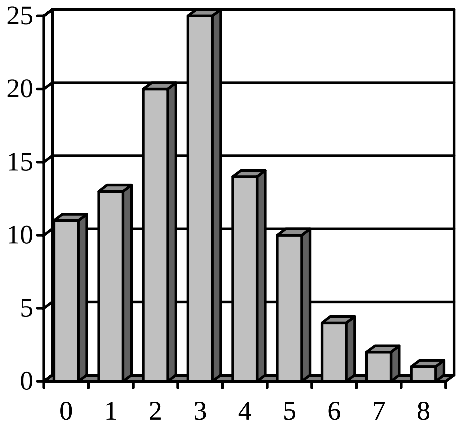
c) Realizando el cambio : $\mathbf{y} = \dfrac{\mathbf{x} - 2754}{2}$

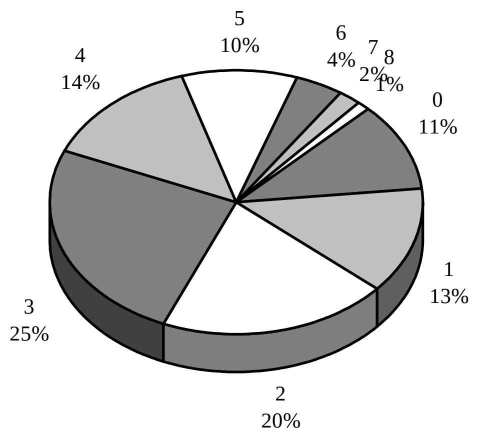
4

a)

x	n	r	p	N	R	P
0	11	0'11	11	11	0'11	11
1	13	0'13	13	24	0'24	24
2	20	0'20	20	44	0'44	44
3	25	0'25	25	69	0'69	69
4	14	0'14	14	83	0'83	83
5	10	0'10	10	93	0'93	93
6	4	0'04	4	97	0'97	97
7	2	0'02	2	99	0'99	99
8	1	0'01	1	100	1'00	100

b)





c) $\overline{\mathbf{x}}$ = 2'8 ; Me = 3 ; Mo = 3

d) R = 8 ; s^2 = 3'14 ; s = 1'772

5

A , D , C , B.

6

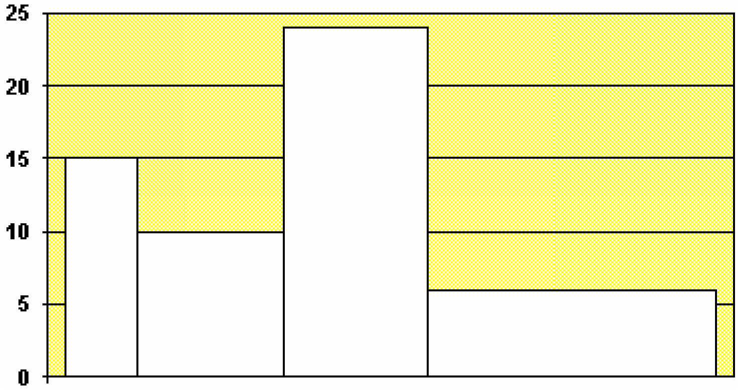
$\mathbf{D_{Me} = D_{\bar{x}} = 870}$

7

Se dividen por dos.

8

Las alturas deben ser proporcionales al área. Dividimos las frecuencias según sea la amplitud del intervalo.
Alturas : **15** **10** (20/2) **24** (48/2) **6** (24/4)



9

x	n
0	2
1	8
2	20
3	8
4	2
	40

$\bar{x}$ **= Me = Mo = 2**

10

a) $\bar{x}$ = 4'7 ; Me = 5 ; Mo = 6
b) 20

11

CV = 15'789

12

15 , 15 , 15'667 , 16'333

13

$$As = \frac{\dfrac{\sum n_i.(x_i - \bar{x})^3}{N}}{\sigma^3}$$ = - 0'299561 ligeramente asimétrica a la izquierda

$$As_1 = \frac{\bar{x} - Mo}{\sigma}$$ = 0'036786 ligeramente asimétrica a la derecha (prácticamente simétrica).

$$As_2 = \frac{3.(\bar{x} - Md)}{\sigma}$$ = - 0'110357 ligeramente asimétrica a la izquierda

Los coeficientes basados en la moda y la mediana hacen uso de una relación teórica entre los parámetros de centralización. Generalmente no conducen a la misma conclusión, salvo distribuciones claramente asimétricas.

$$K = \frac{\frac{\sum n_i.(x_i - \bar{x})^4}{N}}{\sigma^4} - 3$$ = - 0'620240 ligeramente aplastada (mesocúrtica)

14

a) 3'375 ; 3'0714 ; 3
b) 21%
c) 1'3 y 5'1
d) 60'9707% ; 1'1905

15

a) n = 1, 0, 4, 3, 3, 6, 2, 1
N = 1, 1, 5, 8, 11, 17, 19, 20
b) 38'6364
c) 17
d) 4'333 y 5
e) a_1 = 4'4 ; a_2 = 22'25 ; a_3 = 121'7 ; a_4 = 703'0625
m_1 = 0 ; m_2 = 0 ; 2'89 ; m_3 = -1'6320 ; m_4 = 21'2737
f) A = -0'3322 ; K = -0'4529 ⊗

16

Índice de Gini = 0'6567
Media = 2'14 ; Mediala = 8

17

Índice de Gini = 0'394
Mediala = 60'5263

⊗ Puede que sus resultados no coincidan exactamente con los ofrecidos. Todo depende del número de cifras decimales que emplee en sus cálculos.