Estadística

Distribuciones Unidimensionales
Población y Muestra
Variables estadísticas
Tablas estadísticas
Gráficos estadísticos
Parámetros estadísticos
Calculadora científica Casio
Problemas resueltos

Introducción

La palabra **Estadística** proviene de que fueron los Estados los impulsores de las primeras elaboraciones de datos, sobre todo, en el siglo XVIII. La Estadística es *la Ciencia que utiliza conjuntos de datos numéricos para obtener a partir de ellos inferencias basadas en el cálculo de probabilidades.*

Así pues, la estadística estudia fenómenos que, de alguna manera, se pueden cuantificar; que generan conjuntos de datos. Trabajando con estos datos el estadístico, utilizando las técnicas apropiadas, tratará de simplificar al máximo la información disponible, a fin de que pueda ser clara y útil. Además, si el fenómeno lo requiere, tratará de deducir, de inferir, las leyes que expliquen el comportamiento de ese fenómeno.

La idea de **inferencia** es la de deducción arriesgada y, por tanto, con posibilidad de error (error que debe ser conocido), pues la estadística se mueve en el campo del azar, de lo probable, y es ahí donde pretende buscar las leyes que determinan su comportamiento a fin de poder tomar las decisiones oportunas, conociendo con antelación la significación de esos resultados.

Veamos tres situaciones:

I. En un regimiento de 900 soldados, deseamos conocer la estatura de todos ellos.

II. Se desea conocer la talla de los varones españoles que tengan una edad comprendida entre 18 y 30 años.

III. Un fabricante de bombillas se encuentra con una producción de 10.000 unidades que, por algún fallo en el proceso de fabricación, comprobado posteriormente, teme que sean defectuosas. Desea, pues, conocer sus características antes de lanzarlas al mercado, porque, en caso de ser defectuosas, considera preferible destruirlas.

En el primer caso es muy fácil tallar a los 900 soldados. Los resultados obtenidos se tabularán y, si se quiere, se representarán gráficamente.

En el segundo caso es prácticamente imposible tallar a los millones de individuos que cumplen esas características. No quedará más remedio que tomar una *muestra* (un pequeño conjunto de ellos) y efectuar la medida en éstos. Las conclusiones se harán extensivas a la totalidad.

En el tercer caso, las características que el fabricante de bombillas desea comprobar son la luminosidad y la duración de cada una de ellas. La primera, la luminosidad, la puede comprobar exahustivamente (en cada una de las 10.000 bombillas), pero sería un proceso demasiado largo. La duración, sin embargo, no puede medirla en todas, pues la única forma de averiguar la vida de una lámpara es dejarla encendida hasta que se funda y cronometrar su duración (es un proceso destructivo). No le quedará más remedio que proceder al estudio de una muestra y extender las conclusiones a la totalidad.

La estadística se puede dividir en dos partes:

Estadística descriptiva La estadística descriptiva o deductiva, trata del recuento, ordenación y clasificación de los datos obtenidos por las observaciones. Se construyen tablas y se representan gráficos que permiten simplificar, en gran medida, la complejidad de todos los datos que intervienen en la distribución. Asimismo se calculan parámetros estadísticos que caracterizan la distribución. ***En esta parte de la estadística no se hace uso del cálculo de probabilidades, y únicamente se limita a realizar deducciones directamente a partir de los datos y parámetros obtenidos*** **(caso I).**

Estadística inferencial La estadística inferencial o inductiva plantea y resuelve el problema de establecer previsiones y conclusiones válidas generales sobre una población a partir de los resultados obtenidos de una muestra. ***Utiliza resultados obtenidos mediante la estadística descriptiva y se apoya fuertemente en el cálculo de probabilidades*** **(casos II y III).**

Población y Muestra

Se denomina **población** al conjunto de todos los elementos que cumplen una determinada característica, que deseamos medir o estudiar. Los elementos de la población se llaman **individuos** (debido a su origen demográfico).

Se llama **muestra** a cualquier subconjunto de la población (precisamente aquel subconjunto sobre el que se va a realizar el estudio). El número de elementos de la muestra se llama **tamaño** de la misma. Cuando la muestra coincide con la población se llama **censo**.

Ejemplo: En los problemas planteados anteriormente, las *poblaciones* estaban formadas por:

- 900 soldados de un regimiento.
- Todos los varones españoles de edades entre 18 y 30 años.
- 10.000 bombillas.

Los *individuos* son:

- cada uno de los soldados.
- cada uno de los varones españoles de 18 a 30 años.
- cada una de las 10.000 bombillas.

Un mismo conjunto puede ser población o muestra según los casos. Por ejemplo, el conjunto de seres vivos encontrados en una parcela de una Hectárea, será población si nuestro interés se limita a esa parcela; será muestra si, pretendiendo conocer la ecología de una región, acotamos esa parcela para sacar de su estudio conclusiones extensivas a toda la región.

Variables estadísticas

Se llama ***variable estadística*** al conjunto de valores que toma un carácter estadístico. Dependiendo del carácter, una variable estadística puede ser *cuantitativa* o *cualitativa*.

Variable estadística discreta

Una variable estadística se llama discreta cuando puede tomar un número finito de valores o infinito numerable. Variables estadísticas discretas son por ejemplo:

- Número de empleados de una fábrica.
- Número de hijos de 20 familias.
- Número de goles marcados por la Selección Nacional de Fútbol en cada una de las últimas 30 temporadas.

Variable estadística continua

Una variable estadística es continua cuando puede tomar, al menos teóricamente, todos los valores posibles dentro de un cierto intervalo de la recta real.

Variables estadísticas continuas son por ejemplo:

- La altura de un individuo.
- El peso de un individuo.
- El tamaño de los objetos.
- El tiempo que tarda en caer un objeto.

En la práctica muchas medidas de carácter continuo se hacen discretas, como, por ejemplo, la talla de los individuos: la estatura suele redondearse en cm.. Otras veces, fundamentalmente para obtener resultados teóricos, una variable discreta puede tratarse como continua.

Los valores de las variables estadísticas se acostumbran a representar por $\mathbf{x_1, x_2, x_3, \cdots\cdots\cdots, x_n}$

Frecuencias absolutas y relativas

Frecuencia absoluta — Se llama frecuencia absoluta del valor x_i, y lo representamos por $\mathbf{f_i}$, ***al número de veces que se repite dicho valor.***

Frecuencia absoluta acumulada — Se llama frecuencia absoluta acumulada del valor x_i, y lo representamos por $\mathbf{F_i}$, ***a la suma de las frecuencias absolutas de todos los valores anteriores a*** $\mathbf{x_i}$ ***más la frecuencia absoluta de*** $\mathbf{x_i}$***:***

$$\mathbf{F_i = f_1 + f_2 + f_3 + \cdots\cdots\cdots + f_i}$$

La frecuencia absoluta no es suficiente para reflejar la intensidad con que se repite un determinado valor de la variable estadística. Por ejemplo, no es lo mismo obtener tres veces un cinco en diez lanzamientos de un dado que obtenerlo en 1000 lanzamientos del mismo.

Frecuencia relativa

Se llama frecuencia relativa de un valor x_i y lo representaremos por $\mathbf{f_r}$, ***al cociente entre la frecuencia absoluta de*** $\mathbf{x_i}$ ***y el número total de datos que intervienen en la distribución***:

$$\mathbf{f_r} = \frac{\mathbf{f_i}}{\sum \mathbf{f_i}}$$

siendo $\sum \mathbf{f_i}$ el número total de datos.

Si cada frecuencia relativa se multiplica por 100 se obtiene el tanto por ciento correspondiente a cada valor.

Frecuencia relativa acumulada

Se llama frecuencia relativa acumulada del valor x_i, y la representaremos por $\mathbf{F_r}$, **a la suma de las frecuencias relativas de todos los valores anteriores a** $\mathbf{x_i}$ **más la frecuencia relativa de** $\mathbf{x_i}$**.**

En todos los casos, la suma de las frecuencias absolutas debe ser igual al total de elementos, la suma de las frecuencias relativas será igual a la unidad, y la suma de los porcentajes deberá ser igual a 100.

Tratamiento de la información. Tablas estadísticas

A continuación vamos a estudiar cómo debemos proceder ordenadamente para analizar la muestra:

1. **Recogida de datos**: Consiste en la toma de datos numéricos procedentes de la muestra.

2. **Ordenación de los datos**: Una vez recogidos los datos los colocaremos en orden creciente o decreciente.

3. **Recuento de frecuencias**: Efectuaremos el recuento de los datos obtenidos.

4. **Agrupación de los datos**: En caso de que la variable sea continua o bien discreta pero con un número de datos muy grande es muy aconsejable agrupar los datos.

   Se llama **intervalos de clase** ***a cada uno de los intervalos en que pueden agruparse los datos de una variable estadística***. Ahora bien, ¿Cuál es el número idóneo de intervalos que debemos escoger a la hora de agrupar? No existe una contestación tajante a esta pregunta; existen incluso varios criterios para dar respuesta a esta cuestión.

   Con carácter muy general podemos enunciar como uno de los criterios muy sencillos el de **Norcliffe** que establece que ***el número de intervalos debe ser aproximadamente igual a la raíz***

cuadrada positiva del número de datos. Para obtener la amplitud de dichos intervalos se procede del siguiente modo:

- *Se localizan los valores menor y mayor de la distribución.*
- *Se restan. Si la diferencia es divisible entre el número de intervalos, el cociente nos da su amplitud. En caso contrario se busca el primer número entero por exceso de la diferencia que sea divisible por el número de intervalos. El cociente de esta división nos da la amplitud de los mismos.*
- *Se forman los intervalos de modo que contengan todos los datos.*
- Con el fin de que la clasificación esté bien hecha se deben construir de tal manera que ***el límite superior de un intervalo coincida con el límite inferior del siguiente***. *Se debe adoptar el criterio de que los intervalos sean cerrados por la izquierda y abiertos por la derecha.*
- ***El punto medio entre los extremos de cada intervalo se llama*** **marca de clase** que será la variable que utilizaremos para el cálculo de los parámetros estadísticos.

5. **Construcción de la tabla estadística**: En la tabla deberán figurar los valores de la variable (y en caso de que se encuentre agrupada en intervalos, los límites superior e inferior, así como las marcas de clase), frecuencias absolutas y frecuencias relativas. A veces es conveniente incluir las frecuencias absolutas acumuladas y las frecuencias relativas acumuladas.

   En muchas ocasiones es interesante trabajar con ***porcentajes****; éstos se obtienen multiplicando las frecuencias relativas por 100.*

Gráficos estadísticos

Los gráficos tienen un mayor poder de comunicación de los resultados al utilizar formas visuales de fácil comprensión (*la información entra por los ojos)*.

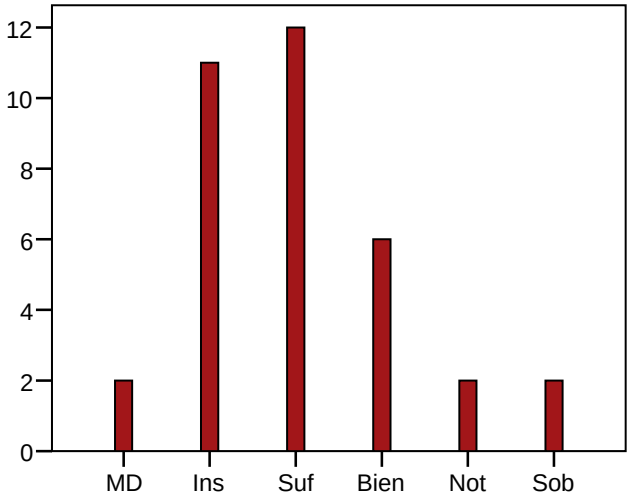
La claridad que requiere un gráfico hace que no sea indiferente el modelo utilizado. Según como sea la naturaleza del carácter estudiado, utilizaremos uno u otro tipo de representación gráfica.

Diagrama de barras

Recibe el nombre de **diagrama de barras** el gráfico que ***asocia a cada valor de la variable una barra estrecha, generalmente vertical, de longitud proporcional a la frecuencia (o a la cantidad) con que se presenta.*** Suele usarse cuando la variable es discreta.

Ejemplo: **Las notas de matemáticas de los 35 alumnos de una clase vienen dadas por la siguiente tabla:**

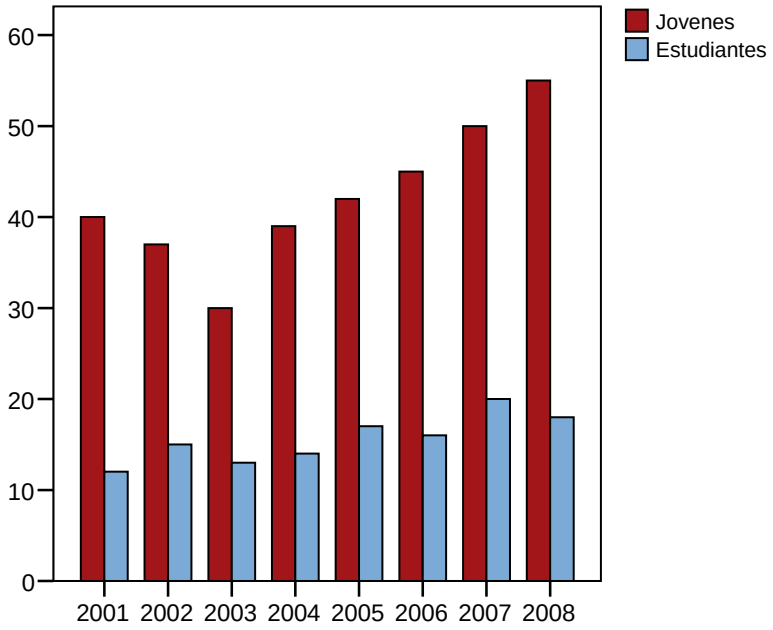
Calificaciones	Nºdealumnos
MD	2
Ins	11
Suf	12
Bien	6
Not	2
Sob	2



Ejemplo **En un barrio de cierta ciudad se ha dado la siguiente distribución**.

Año	Jóvenes mayores de 15 años	Estudiantes de Bachillerato
2001	40	12
2002	37	15
2003	30	13
2004	39	14
2005	42	17
2006	45	16
2007	50	20
2008	55	18

En este caso existe una conexión entre los dos caracteres estudiados, pues los alumnos de Bachillerato son todos mayores de 15 años.



Histogramas

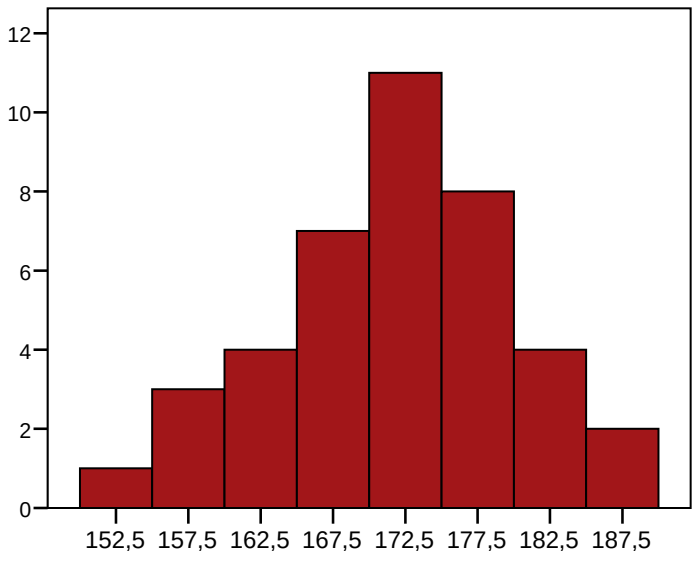
Los histogramas son diagramas de barras que se utilizan específicamente para distribuciones de variable estadística continua, o bien para distribuciones de variable estadística discreta con gran número de datos y que se han agrupado por intervalos (clases). Generalmente se acostumbra a agrupar los datos obtenidos en intervalos de igual amplitud.

Estos gráficos asocian a cada intervalo un rectángulo de superficie proporcional a la frecuencia correspondiente a dicho intervalo. Para construir el histograma se representa sobre el eje de abscisas los límites de los intervalos. Sobre dicho eje se construyen unos ***rectángulos que tienen por base la amplitud del intervalo y por alturas los cocientes entre las frecuencias absolutas y las longitudes de los intervalos correspondientes* (densidad de frecuencia).** De esta manera, el área del rectángulo coincide con la frecuencia del intervalo.

<u>Intervalos de la misma amplitud</u> ***(alturas directamente proporcionales a la frecuencia)***

En un grupo de 40 alumnos se ha estudiado su talla, obteniéndose los resultados de la tabla adjunta.

Intervalos en cm.	x_i	f_i
$[150,155)$	152'5	1
$[155,160)$	157'5	3
$[160,165)$	162'5	4
$[165,170)$	167'5	7
$[170,175)$	172'5	11
$[175,180)$	177'5	8
$[180,185)$	182'5	4
$[185,190)$	187'5	2
		40

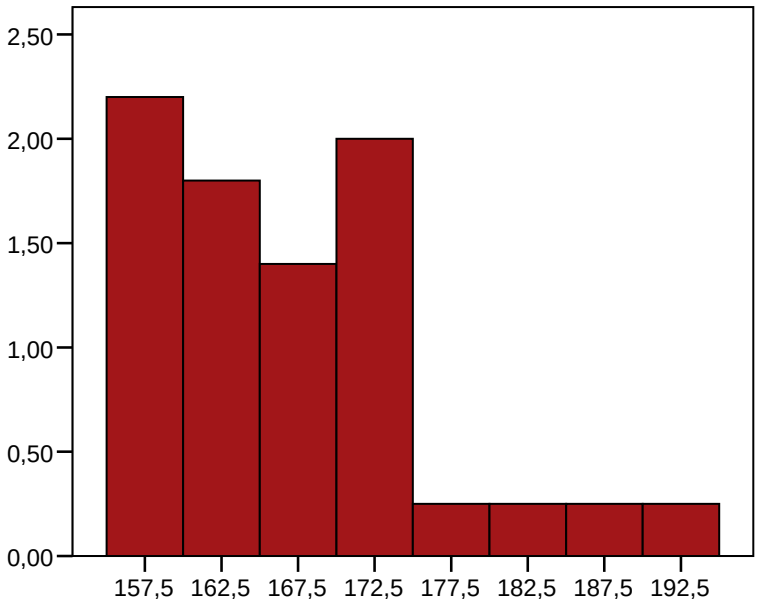


Intervalos de distinta amplitud

Cuando los intervalos sean de distinta amplitud, las alturas de los rectángulos deben ser siempre igual a la densidad de frecuencia. En el ejemplo siguiente se observa que no todos los intervalos tienen la misma amplitud. El último intervalo tiene una amplitud de 20 y por tanto la altura del rectángulo es de 0'25 que es el resultado de dividir la frecuencia del intervalo, en este caso 5, entre la anchura del mismo, 20. Observa que el producto de la amplitud por la altura del rectángulo da la frecuencia indicada en la figura.

Ejemplo: **En un grupo de 42 alumnos se ha estudiado su talla, obteniéndose los resultados de la tabla adjunta.**

Intervalos	f_i	d_i
[155,160)	11	2′2
[160,165)	9	1′8
[165,170)	7	1′4
[170,175)	10	2
[175,195)	5	0′25
	42	

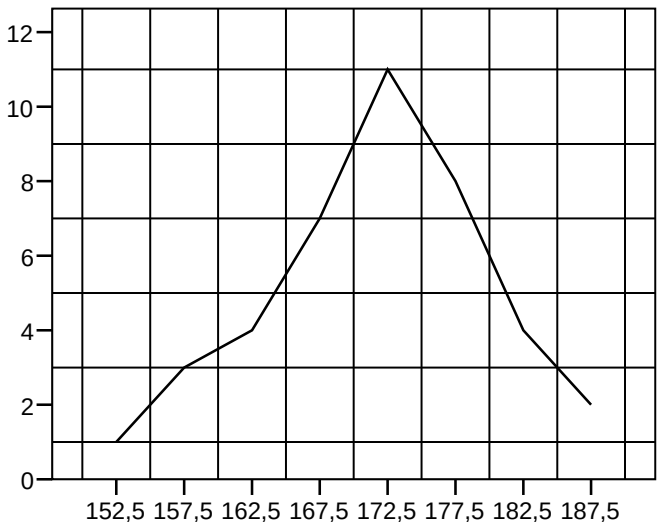


Polígono de frecuencias

Se considera polígono de frecuencias a la línea poligonal (línea quebrada) que une los puntos correspondientes a las frecuencias de cada valor de la variable estadística, o de los extremos superiores de las barras. Si los datos vienen dados en intervalos, unirá los puntos correspondientes a las marcas de clase.

Ejemplo: **En un grupo de 40 alumnos se ha estudiado su talla, obteniéndose los resultados de la tabla adjunta.**

Intervalos	x_i	f_i
[150,155)	152,5	1
[155,160)	157,5	3
[160,165)	162,5	4
[165,170)	167,5	7
[170,175)	172,5	11
[175,180)	177,5	8
[180,185)	182,5	4
[185,190)	187,5	2
		40



Diagramas de sectores

En los gráficos de diagramas de sectores ***cada suceso viene representado por un sector circular de amplitud proporcional a su frecuencia absoluta***. La amplitud de cada sector se obtiene mediante una sencilla regla de tres. Los diagramas de sectores son especialmente adecuados para representar las distintas partes de un todo y para representar varias situaciones similares y poder establecer comparaciones. El diagrama de sectores correspondiente al ejemplo anterior es:

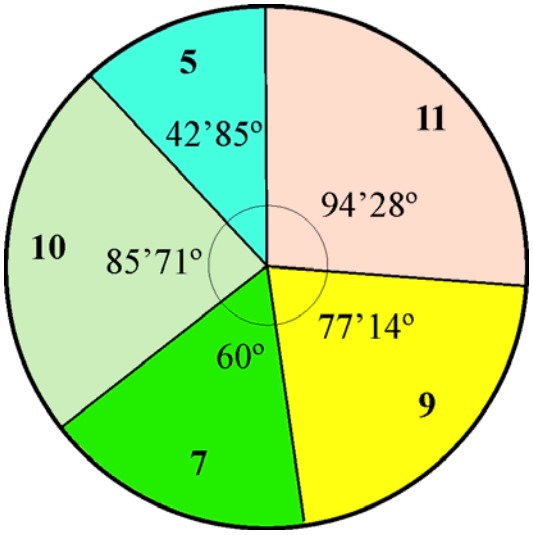
$$\left.\begin{array}{l} 42 \rightarrow 360^o \\ 11 \rightarrow x^o \end{array}\right\} \Rightarrow x = 94'28^o$$

$$\left.\begin{array}{l} 42 \rightarrow 360^o \\ 9 \rightarrow x^o \end{array}\right\} \Rightarrow x = 77'14^o$$

$$\left.\begin{array}{l} 42 \rightarrow 360^o \\ 7 \rightarrow x^o \end{array}\right\} \Rightarrow x = 60^o$$

$$\left.\begin{array}{l} 42 \rightarrow 360^o \\ 10 \rightarrow x^o \end{array}\right\} \Rightarrow x = 85'71^o$$

$$\left.\begin{array}{l} 42 \rightarrow 360^o \\ 5 \rightarrow x^o \end{array}\right\} \Rightarrow x = 42'85^o$$



Pirámides de población

Son un caso particularmente importante de histogramas. En realidad ***se trata de dos histogramas de distribución de edades, uno para hombres y otro para mujeres. Se utilizan para estudiar conjuntamente la variable edad y el atributo sexo.*** En el eje de ordenadas se representa los intervalos de edades cuya anchura puede ser anual, quinquenal o decenal, dependiendo del detalle necesario. En el eje de abscisas se representa el sexo. Para la modalidad mujer se toma el semieje positivo, y para la modalidad hombre el semieje negativo.

Pictogramas

Los **pictogramas** ***son gráficos en los que utilizan dibujos alusivos a la distribución que se pretende estudiar, y que mediante su forma, tamaño, etc., ofrecen una descripción lo más expresiva posible de la distribución estadística***

Cartogramas

Se llama **cartograma** ***a los gráficos que se realizan sobre un mapa, señalando sobre determinadas zonas con distintos colores o rayados lo que se trata de poner de manifiesto***.

Por ejemplo, se suelen utilizar este tipo de diagramas para representar la densidad demográfica de una nación, la renta per cápita, las horas de sol anuales sobre una determinada parte de la tierra, los índices de lluvias de una nación, etc.

Parámetros estadísticos

Los parámetros estadísticos resumen en un número un aspecto relevante de la distribución estadística que pueda dar una idea de la misma o compararla en ese aspecto, con otras. Evidentemente, todo proceso de síntesis conlleva una pérdida de información; pero se gana en el hecho de que es más fácil trabajar con unos pocos parámetros con significado muy preciso que con la totalidad de los datos. Los parámetros estadísticos suelen clasificarse, según el papel que juegan, de la siguiente manera:

Medidas de Centralización

O de tendencia central, indican valores con respecto a los que los datos parecen agruparse. Las más importantes son: **Moda** (*el valor que se presenta con mayor frecuencia*), **Media Aritmética** (*suma de todos los valores de una variable estadística dividido por el número de valores*), **Media Geométrica**, **Media Armónica** y **Mediana** (*el valor del individuo que ocuparía el lugar central si se colocaran ordenados de menor a mayor*).

Medidas de Dispersión

Sirven para medir el grado de alejamiento (dispersión) de los datos. Son la **Desviación Media, Varianza, Desviación Típica, Rango, etc.**

Medidas de Posición

Dividen un conjunto ordenado de datos en grupos con la misma cantidad de individuos. Son los **Cuantiles** (*Cuartiles*, *Deciles*, *Percentiles*, etc.).

Media aritmética

Se llama media aritmética de una variable estadística a la suma de todos los valores de dicha variable dividida por el número de valores.

Sea x una variable estadística que toma los valores $x_1, x_2, x_3, ..., x_n$ con frecuencias absolutas $f_1, f_2, f_3, ..., f_n$ respectivamente. La media aritmética de la variable x se representa por $\overline{x}$, y viene dada por la siguiente expresión:

$$\overline{x} = \frac{x_1 f_1 + x_2 f_2 + \cdots\cdots\cdots + x_n f_n}{f_1 + f_2 + ... + f_n} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i}$$

Si la variable x es continua, o aún siendo discreta, y por tratarse de muchos datos se encuentran agrupados en clases, se toman como valores $x_1, x_2, x_3, ..., x_n$ las marcas de clase.

Ejemplo: Las calificaciones en la asignatura Historia del Arte de los 40 alumnos de una clase viene dada por la tabla adjunta. Hallar la calificación media.

Calificaciones	1	2	3	4	5	6	7	8	9
Número de alumnos	2	2	4	5	8	9	3	4	3

En la práctica, los cálculos se disponen de la siguiente forma:

x_i	f_i	$x_i f_i$
1	2	2
2	2	4
3	4	12
4	5	20
5	8	40
6	9	54
7	3	21
8	4	32
9	3	27
	40	212

La media aritmética será:

$$\overline{x} = \frac{1\cdot 2 + 2\cdot 2 + 3\cdot 4 + 4\cdot 5 + 5\cdot 8 + 6\cdot 9 + 7\cdot 3 + 8\cdot 4 + 9\cdot 3}{40} = \frac{212}{40} = 5'3$$

Luego la calificación media en Historia del Arte es 5'3

Media ponderada

Hasta aquí hemos considerado todos los datos con la misma importancia, es decir, como si todos ellos tuvieran la misma fiabilidad. No obstante, puede suceder que en algún caso, debido al criterio adoptado o a las circunstancias en que se obtuvieron los datos, sea necesario dar más importancia a unos datos que a otros. En este caso, la media se llama **media ponderada** y su expresión es:

$$\overline{x} = \frac{x_1 f_1 a_1 + x_2 f_2 a_2 + \cdots\cdots\cdots + x_n f_n a_n}{f_1 a_1 + f_2 a_2 + \ldots\ldots\ldots + f_n a_n} = \frac{\sum_{i=1}^{n} x_i f_i a_i}{\sum_{i=1}^{n} f_i a_i}$$

donde a_i son las distintas ponderaciones o pesos que se adjudican a los datos.

Ejemplo: **En un curso de ESO los alumnos durante un período evaluativo han realizado las siguientes pruebas: un examen, dos controles y tres intervenciones en clase. Las pruebas, según acuerdo del Seminario, se valoran de la siguiente forma: 50% el examen, 30% los controles y 20% las intervenciones. Si un determinado alumno ha obtenido: 7 en el examen, 6 y 8 en los controles y 10, 5 y 2 en las intervenciones, obtener su nota media de la evaluación.**

$$\overline{x} = \frac{7\cdot 0'5 + (6+8)\cdot \frac{0'30}{2} + (10+5+2)\cdot \frac{0'2}{3}}{1\cdot 0'5 + 2\cdot \frac{0'30}{2} + 3\cdot \frac{0'20}{3}} = \mathbf{6'73}$$

Si en este mismo ejemplo hubiésemos considerado todas las notas con la misma importancia, la nota media obtenida sería:

$$\overline{x} = \frac{7+6+8+10+5+2}{6} = 6'33$$

Moda

Se llama moda de una variable estadística al valor de dicha variable que presenta mayor frecuencia absoluta. Se representa por M_0.

La moda no tiene por qué ser única, puede haber varios valores de la variable con la mayor frecuencia. En este caso se dirá que la distribución es bimodal, trimodal, etc., según que sean 2, 3, etc. los valores de la variable que presentan mayor frecuencia. También se aplica este nombre a distribuciones en las que destacan varios valores con frecuencias muy altas, prácticamente iguales, aunque no todas sean máximas. Si los datos están agrupados hablaremos del intervalo que más se repite o intervalo modal. *Si los intervalos son de amplitud variable, tenemos que sustituir la frecuencia de cada intervalo por su correspondiente* ***densidad de frecuencia***, que como sabemos es el cociente entre la frecuencia y la amplitud del intervalo. Aquí, la clase modal será la clase que tenga mayor densidad de frecuencia.

Ejemplo: Consideremos la siguiente distribución:

Intervalos	**0 – 4**	**4 – 10**	**10 – 20**	**20 – 40**	**40 – 70**
f_i	**20**	**100**	**180**	**260**	**240**

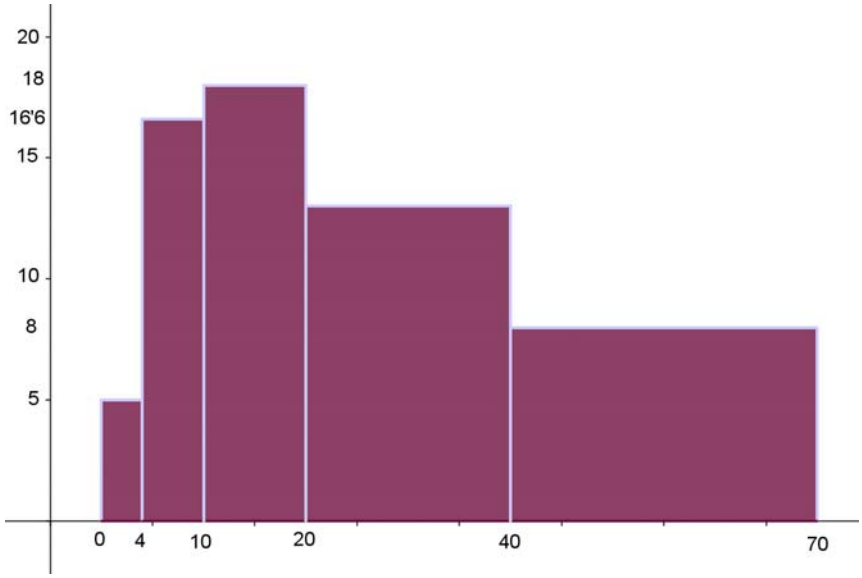
Construimos la tabla con las densidades de frecuencia, ya que la amplitud de los intervalos no es constante. La moda está en el intervalo $10-20$ por ser el que presenta, mayor densidad de frecuencia. Al intervalo que contiene la moda se le llama clase modal o intervalo modal.

Intervalos	x_i	f_i	d_i
$0-4$	2	20	5
$4-10$	7	100	16′6
$10-20$	15	180	18
$20-40$	30	260	13
$40-70$	55	240	8

La Moda es, aproximadamente, la marca de clase correspondiente a dicho intervalo, es decir:

$$M_o = 15$$

$n = 800 \quad \bar{x} = 30'55 \quad \sigma_n = 17'98$



Mediana

Si los datos de la muestra estudiada se ordenan siguiendo un criterio de crecimiento o decrecimiento, se denomina mediana al valor del dato que ocupa el lugar central, o dicho de otro modo, *la*

mediana es el valor que divide a la serie de datos en dos partes exactamente iguales. Se representa con el símbolo $\mathbf{M_e}$.

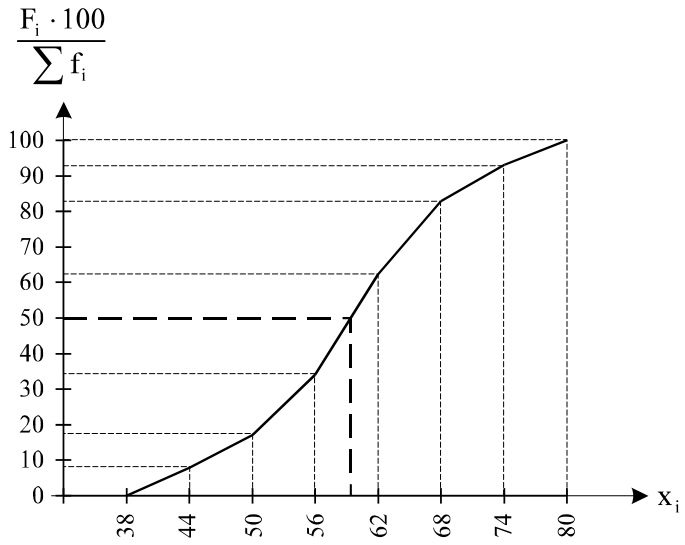
La Mediana es el primer parámetro de centralización que depende del orden de los datos y no de sus valores. Como consecuencia de la definición, se tiene que el 50% de los datos son menores o iguales a la Mediana y el 50% restante son mayores o iguales.

Cálculo de la Mediana a través del polígono de porcentajes acumulados

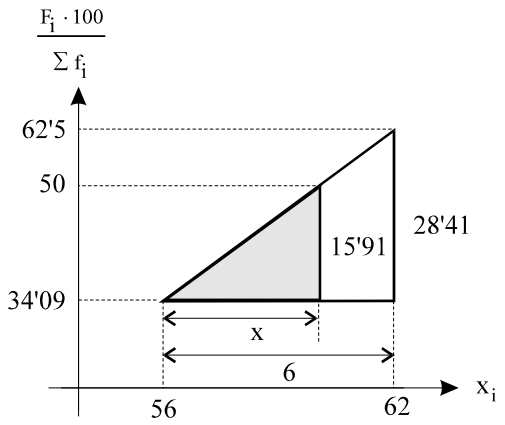
Los polígonos de porcentajes acumulados son útiles para calcular parámetros estadísticos como la Mediana y los Percentiles. Los porcentajes acumulados se obtienen al multiplicar la frecuencia absoluta acumulada por 100 y dividir el resultado por la suma de todas las frecuencias absolutas. $\mathbf{F_i \cdot \frac{100}{\sum f_i}}$

Representamos el polígono de porcentajes acumulados, situando en el eje "x" los valores de la variable (si es discreta) o de los intervalos (si es continua) y en el eje "y" la frecuencia absoluta acumulada en porcentaje. ***Para obtener la Mediana se traza desde el punto $(0,50)$, correspondiente al 50 %, una paralela al eje "x". Ésta corta al polígono de frecuencias absolutas acumuladas en un punto; por éste se traza una paralela al eje "y", que corta al eje "x" en el punto buscado***.

Ejemplo: Calcular la Mediana en el ejemplo sobre el test de satisfacción en el trabajo que se ha aplicado a 88 empleados de una fábrica.



Clases	f_i	F_i	$\frac{F_i \cdot 100}{\sum f_i}$
$[38-44)$	7	7	7'95
$[44-50)$	8	15	17'04
$[50-56)$	15	30	34'09
$[56-62)$	25	55	62'50
$[62-68)$	18	73	82'95
$[68-74)$	9	82	93'18
$[74-80)$	6	88	100
	88		



$$\frac{28'41}{6} = \frac{15'91}{x} \Rightarrow x = 3'36$$

Por lo tanto la mediana será:

$$\mathbf{M_e} = 56 + 3'36 = \mathbf{59'36}$$

Rango o recorrido

Se llama rango o recorrido de una distribución a la diferencia entre el mayor y el menor valor de la variable estadística. Bajo el supuesto de que los valores de la variable estén ordenados en sentido creciente, su expresión matemática sería:

$$\mathbf{R = x_n - x_1}$$

Varianza y Desviación típica de una población

Se llama ***varianza*** *de una variable a la media aritmética de los cuadrados de las desviaciones respecto a la media* y se representa con el símbolo $\boldsymbol{\sigma^2}$. La fórmula que nos da la varianza de una población es:

$$\boldsymbol{\sigma^2 = \frac{\sum_{i=1}^{n} f_i x_i^2}{\sum_{i=1}^{n} f_i} - \bar{x}^2}$$

La media aritmética es el centro de gravedad de la distribución estadística. Si nos imaginamos el diagrama de barras o el histograma de frecuencias apoyado en un punto del eje horizontal de forma que quedase en equilibrio, el valor de este punto en dicho eje sería el valor de la media.

No es suficiente con un parámetro de centralización, es necesario un parámetro de dispersión que nos indique si los datos estudiados están más concentrados o más dispersos. *El parámetro que mide la mayor o menor dispersión de los datos respecto de la media se llama la* ***Desviación típica*** y corresponde a la raíz cuadrada de la varianza. Lógicamente si los datos están más concentrados en torno a la media aritmética la desviación típica será menor, y si los datos están más dispersos la desviación típica será mayor.

$$\boldsymbol{\sigma = \sqrt{\frac{\sum_{i=1}^{n} f_i x_i^2}{\sum_{i=1}^{n} f_i} - \bar{x}^2}}$$

En la calculadora la desviación típica viene representada por el símbolo $\boldsymbol{\sigma_x}$ ó $\boldsymbol{\sigma_n}$ y representa ***la desviación típica de la población*** que estamos estudiando.

Ejemplo: El número de horas que dedica al estudio un alumno de 1º de Bachillerato durante la semana es el siguiente: 3'5, 5, 4, 6, 5'5, 3. Calcular el rango, la varianza y la desviación típica.

x_i	f_i	$x_i f_i$	$x_i^2 f_i$
3	1	3	9
3'5	1	3'5	12'25
4	1	4	16
5	1	5	25
5'5	1	5'5	30'25
6	1	6	36
	6	27	128'5

Rango: $6 - 3 = 3$

Varianza: $\sigma^2 = \frac{x_i^2 f_i}{f_i} - \bar{x}^2 = \frac{128'5}{6} - 4'5^2 = 1'16$

Media: $\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{27}{6} = 4'5$

Desviación típica: $\sigma = \sqrt{1'16} = 1'08$

Ejemplo: **Se ha aplicado un test sobre satisfacción en el trabajo a 88 empleados de una fábrica, obteniéndose los resultados de la tabla adjunta. Calcular el rango, la varianza y la desviación típica.**

Rango: $80-38=42$

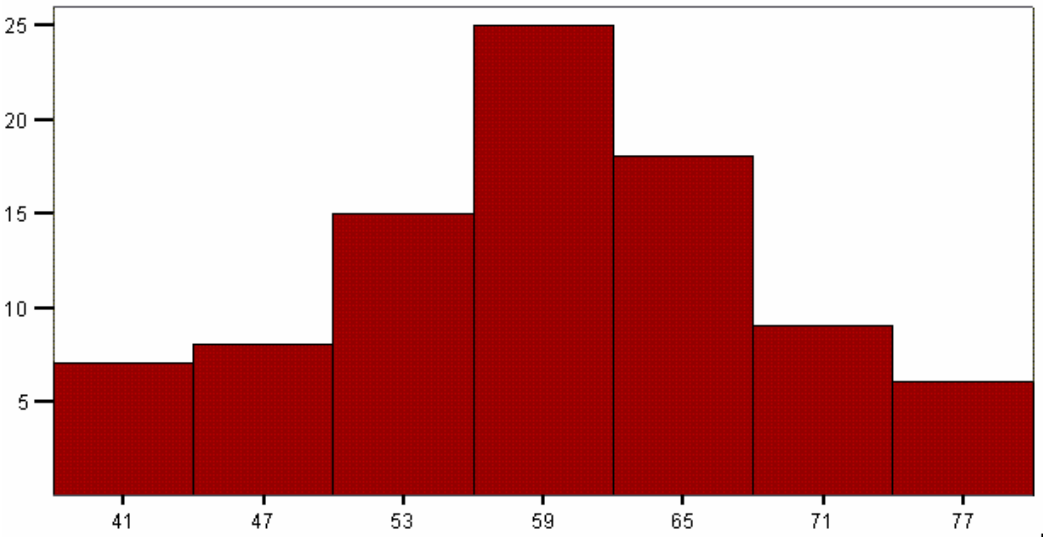
Media: $\overline{x}=\dfrac{5204}{88}=59'14$

Varianza: $\sigma^2=88'73$

Desviación típica:

$\sigma=\sqrt{88'73}=9'4$

Clases	Marcas de clase x_i	f_i	$x_i f_i$	$x_i^2 f_i$
$[38-44)$	41	7	287	11.767
$[44-50)$	47	8	376	17.672
$[50-56)$	53	15	795	42.135
$[56-62)$	59	25	1.475	87.025
$[62-68)$	65	18	1.170	76.050
$[68-74)$	71	9	639	45.369
$[74-80)$	77	6	462	35.574
		88	5.204	315.592



Muestra. Desviación típica de una muestra

Para conocer la intención de voto de los ciudadanos ante unas elecciones no se puede preguntar a todos los electores porque la población es muy grande. Si queremos comprobar si funcionan todas las bombillas que fabrica una determinada empresa es absurdo probarlas todas para realizar el experimento. En situaciones en las que no se pueden observar todos los individuos de una población, se hace necesario elegir una ***muestra***, para conocer las características de la población.

Una muestra es una parte de la población que sirve para representarla.

Los métodos más empleados para elegir muestras son los métodos aleatorios que dependen del azar. Para determinar el tamaño de una muestra existen procedimientos estadísticos basados en la dispersión de los datos: *cuanto mayor nos parezca la desviación típica de una población, más grande debe ser la muestra que se elija*.

En la calculadora la desviación típica de una ***muestra*** *viene representada por el símbolo* $\mathbf{s_x}$ ó $\boldsymbol{\sigma_{n-1}}$.

El coeficiente de variación

La dispersión de los datos en una población no puede determinarse exclusivamente a partir de la desviación típica, pues no siempre una desviación típica mayor indica mayor dispersión. Por ejemplo, la desviación típica del peso de un grupo de 5 caballos suele ser mayor que la desviación típica del peso de un grupo de 5 conejos. Parece evidente que habrá que tener en cuenta las medias de los pesos de ambos grupos de animales, para establecer algún tipo de comparación relativa o proporcional.

Una medida de la dispersión relativa de dos conjuntos de datos es el ***coeficiente de variación***, que se define como:

$$\mathbf{CV = \frac{\sigma}{\bar{x}}}$$

- *El coeficiente de variación se usa para comparar la dispersión de distribuciones que tienen diferentes medias y distintas desviaciones.*
- Dados dos conjuntos, aquel que tenga un coeficiente de variación mayor es el más disperso, el más heterogéneo. Su valor no depende de la unidad de medida utilizada.
- El coeficiente de variación suele darse en porcentajes: $CV = \frac{\sigma}{\bar{x}} \cdot 100$.
- Un porcentaje del 30% indica que la media es poco representativa como medida del promedio, debiéndose optar por la Mediana o la Moda.

Ejemplo: **Se han medido los pesos y las alturas de 6 personas, obteniéndose los datos de la tabla adjunta.**
¿Qué están más dispersos, los pesos o las alturas?

Pesos (kg)	Alturas (m)
65	**1′7**
60	**1′5**
63	**1′7**
63	**1′7**
68	**1′75**
68	**1′8**

Calculamos la media y la desviación típica de cada variable:

$\bar{x}_P = 64'5$ kg $\quad \sigma_P = 2'87$ kg $\quad \bar{x}_A = 1'69$ m $\quad \sigma_A = 0'093$ m

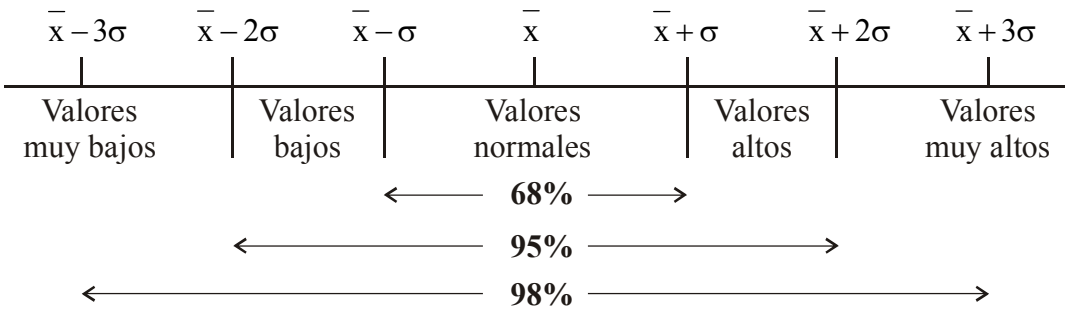
Para comparar la dispersión, hallamos el coeficiente de variación:

$$CV_P = \frac{\sigma_P}{\bar{x}_P} = \frac{2'87}{64'5} = 0'044 \rightarrow 4'4\ \% \qquad CV_A = \frac{\sigma_A}{\bar{x}_A} = \frac{0'093}{1'69} = 0'055 \rightarrow 5'5\ \%$$

Se observa que las alturas están más dispersas que los pesos, y sin embargo, la desviación típica de los pesos es mayor que la de las alturas.

Agrupación de datos en torna a la media aritmética

Si una distribución tiene una sola moda y es simétrica, es decir, cuando el histograma tiene forma de campana, llamada Campana de Gauss, se dice que la variable aleatoria se distribuye normalmente (altura de los individuos de una población, peso, coeficiente intelectual, etc.). En una distribución normal se puede hacer una clasificación de la población utilizando estos intervalos:



Medidas de posición. Cuantiles

Al estudiar la Mediana hemos visto que, una vez ordenados de menor a mayor los datos de una distribución, la mediana divide a éstos en partes iguales. Análogamente, tiene interés estudiar otros parámetros que dividan a los datos de la distribución en función de otras cuantías.

Reciben genéricamente la denominación de **cuantiles** ***aquellos valores que dividen la distribución en intervalos, de forma que cada uno de ellos tenga la misma frecuencia.*** Los cuantiles toman denominaciones específicas según sea el número de intervalos en que se divide la distribución. así:

Cuartiles *Se llama cuartiles a tres valores que dividen a la serie de datos en cuatro partes iguales, conteniendo cada una de ellas el 25% de la población***.** Se representan por $\mathbf{Q_1}$, $\mathbf{Q_2}$ **y** $\mathbf{Q_3}$ y se designan cuartil primero, segundo y tercero respectivamente. Los cuartiles Q_1, Q_2, y Q_3 son los valores que superan, exactamente, al 25%, 50% y 75% de los valores de la distribución respectivamente. El cuartil Q_2 coincide con la mediana de la distribución. Hay dos valores, uno que separa a la población en un 25% por debajo y un 75% por encima, y el otro que deja por debajo al 75% y por encima al 25% de la población. Se llaman **cuartil inferior (CI) y cuartil superior (CS)**, y corresponden a Q_1 y Q_3 respectivamente.

Si el problema que estudiamos son las notas en una determinada asignatura "Estar por encima del cuartil superior" significa estar entre el 25% de los mejores.

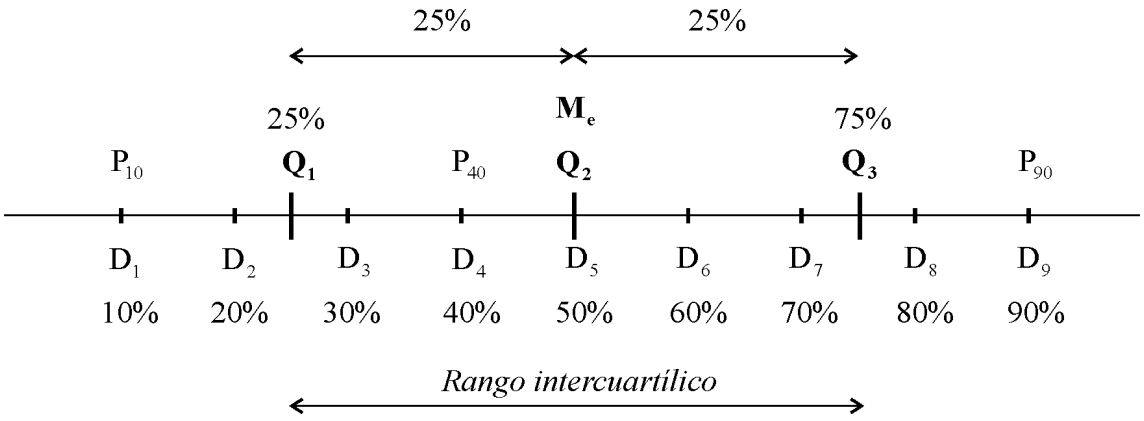
Deciles *Se llama deciles a nueve valores que dividen a la serie de datos en diez partes iguales, conteniendo cada una de ellas la décima parte de la población*. Se representan por $\mathbf{D_1, D_2, \cdots\cdots, D_9}$ y se designan decil primero, segundo, tercero,cuarto, ..., y noveno respectivamente. Hablar del decil 4 significa dejar por debajo del valor que representa al 40% de la población.

Percentiles *Se llaman centiles o percentiles a 99 valores que dividen a la serie de datos en cien partes iguales. Se representan* por $\mathbf{P_1, P_2, \cdots\cdots, P_{99}}$ y se designan percentil primero, segundo, tercero, cuarto,....., y nonagésimo noveno respectivamente. Hablar del centil 38 significa dejar por debajo del valor que representa al 38% de la población. Cuando se dice "Según su inteligencia abstracta este chico está en el centil 85" significa que su inteligencia abstracta, es superior a la del 85% de la población e inferior al 15% restante. Los centiles son muy utilizados por los psicólogos para dar los resultados de los tests

En una tabla de frecuencias con datos agrupados en intervalos, suponemos que los datos de cada intervalo se reparten uniformemente en él. Así, si en un intervalo de amplitud *L* hay *n* elementos, supondremos que la distancia entre cada uno es *L/n*. *Según esto, los valores de las frecuencias acu-*

muladas deben asignarse a los extremos superiores de los intervalos, pues es al final de cada intervalo cuando se han contabilizado todos los individuos.

A continuación se representa un gráfico donde se muestran las relaciones entre los distintos cuantiles.



Obsérvese que la Mediana coincide con el cuartil segundo (Q_2), el decil quinto (D_5) y el percentil de orden cincuenta (P_{50}), es decir: $M_e = Q_2 = D_5 = P_{50}$

Ejemplo **Las calificaciones en Matemáticas de los 40 alumnos de una clase vienen dadas por la tabla adjunta.**

x	1	2	3	4	5	6	7	8	9
f	2	2	4	5	8	9	3	4	3

a) Calcular los cuartiles 1º y 3º y los percentiles de orden 30 y 70.
b) ¿Cuál es el valor de la Mediana?

a) Calculamos las columnas correspondientes a la frecuencias absolutas acumuladas y al porcentaje de frecuencias absolutas acumuladas.

x	f	F	$F \cdot \frac{100}{40}$
1	2	2	5
2	2	4	10
3	4	8	20
4	5	13	32′5
5	8	21	52′5
6	9	30	75
7	3	33	82′5
8	4	37	92′5
9	3	40	100
	40		

Cuartil 1º
El cuartil primero corresponde al percentil 25 y es $Q_1 = P_{25} = 4$ por ser éste el primer valor de la variable cuyo porcentaje de frecuencia absoluta acumulada excede al 25%.

Cuartil 3º
El cuartil tercero corresponde al percentil 75 que coincide con un valor del porcentaje de frecuencias absoluta acumuladas para $x = 6$. El cuartil es la semisuma del valor de la variable correspondiente a dicha frecuencia y el siguiente.

Percentil 30
El percentil 30 es: $P_{30} = 4$ por ser éste el primer valor de la variable cuyo porcentaje de frecuencia absoluta acumulada excede al 30%.

Percentil 70

El percentil 70 es: $P_{30} = 6$ por ser éste el primer valor de la variable cuyo porcentaje de frecuencia absoluta acumulada excede al 70%.

b) La Mediana corresponde al percentil 50, es decir, $M_e = P_{50}$. El primer valor de la variable cuyo porcentaje de frecuencia absoluta acumulada excede al 50% es el 5 por tanto $M_e = 5$.

Cálculo gráfico de los Percentiles

Debido a que los Percentiles son parámetros del tipo de la Mediana, su cálculo se realiza de forma análoga. Representamos el polígono de porcentajes acumulados, situando en el eje "x" los valores de la variable (si es discreta) o de los intervalos (si es continua) y en el eje "y" la frecuencia absoluta acumulada en porcentaje. *Para calcular un percentil p_k, se señala el porcentaje correspondiente k, en el eje vertical, que está graduado de 0 a 100. Por este punto, se traza una paralela al eje X que corta al polígono de frecuencias absolutas acumuladas en un punto; por este punto se traza una paralela al eje "Y", que corta al eje "X" en el punto buscado p_k. Para calcular numéricamente, de forma exacta, el valor de p_k, no hay más que razonar adecuadamente con los valores del intervalo correspondiente y aplicar una semejanza de triángulos.*

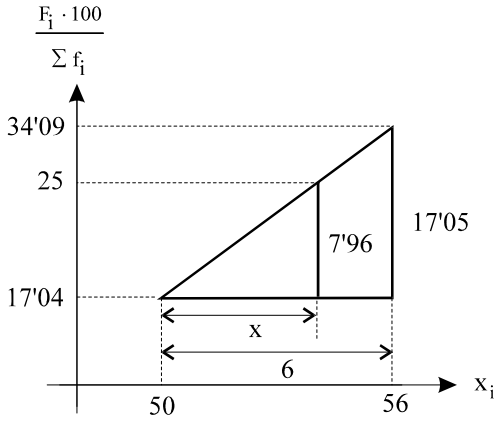
Ejemplo: Se ha aplicado un test sobre satisfacción en el trabajo a 88 empleados de una fábrica, obteniéndose los siguientes resultados:

Puntuaciones	Nº de trabajadores
$[38-44)$	7
$[44-50)$	8
$[50-56)$	15
$[56-62)$	25
$[62-68)$	18
$[68-74)$	9
$[74-80)$	6

Clases	f_i	F_i	$\frac{F_i \cdot 100}{88}$
$[38-44)$	7	7	7'95
$[44-50)$	8	15	17'04
$[50-56)$	15	30	34'09
$[56-62)$	25	55	62'5
$[62-68)$	18	73	82'95
$[68-74)$	9	82	93'18
$[74-80)$	6	88	100
	88		

a) Calcular el cuartil inferior, el decil 7 y el percentil de orden 90.
b) ¿Qué centil corresponde a 45 puntos?

Cuartil inferior

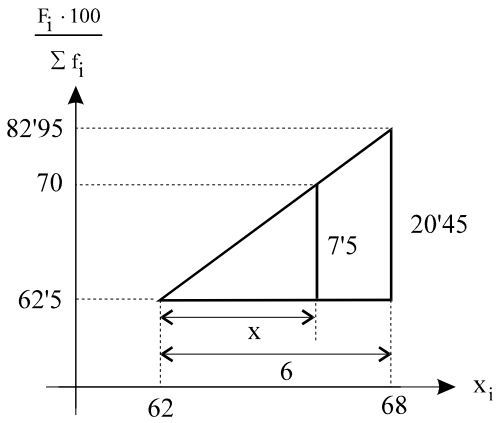


El cuartil inferior corresponde al percentil 25.

$$\frac{17'05}{6} = \frac{7'96}{x} \Rightarrow x = 2'8011$$

$$Q_1 = 50 + 2'8011 = 52'8011$$

Decil 7

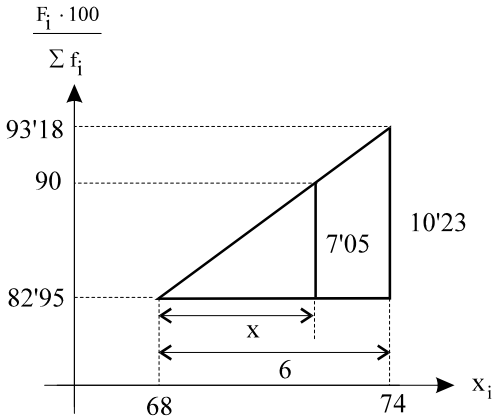


El decil 7 corresponde al percentil 70.

$$\frac{20'45}{6} = \frac{7'5}{x} \Rightarrow x = 2'20$$

$$D_7 = 62 + 2'2 = 64'2$$

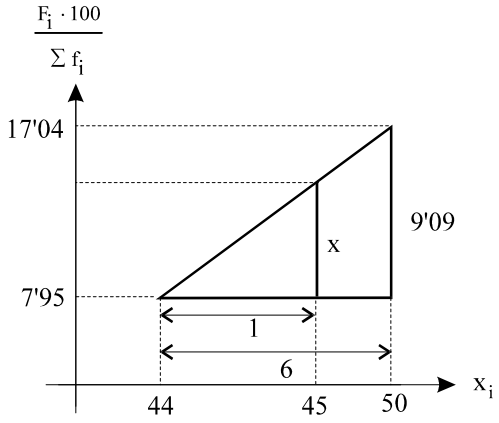
Percentil 90



$$\frac{10'23}{6} = \frac{7'05}{x} \Rightarrow x = 4'13$$

$$P_{90} = 68 + 4'13 = 72'13$$

¿Qué centil corresponde a 45 puntos?



$$\frac{9'09}{6} = \frac{x}{1} \Rightarrow x = 1'51$$

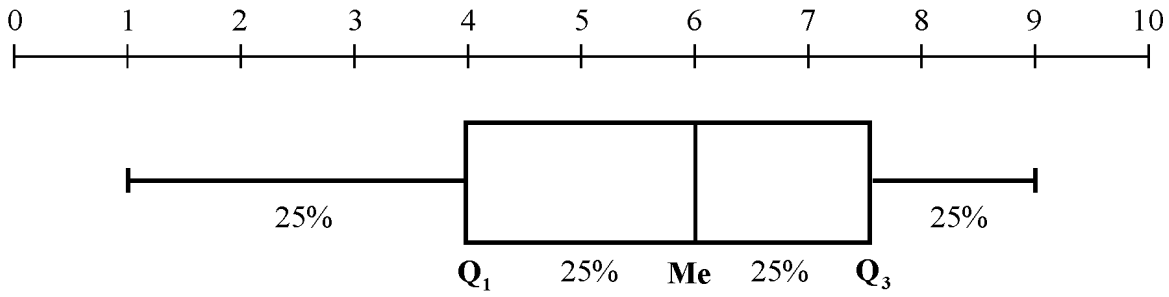
$$7'95 + 1'51 = 9'46$$

Corresponde aproximadamente al centil 9.

Diagramas de caja y bigotes

Un gráfico de caja y bigote es una representación gráfica que permite estudiar la simetría o asimetría de una distribución así como la dispersión de los datos. Por su facilidad de construcción e interpretación, permite comparar a la vez varios grupos de datos, sin saturarse de ellos.

La siguiente gráfica corresponde a la distribución de notas de un examen de matemáticas. En la parte superior figura la escala sobre la que se mueve la variable y en la parte inferior está el ***diagrama de caja y bigotes***.



- *La población total se parte en 4 trozos, cada uno de ellos con el 25% de los individuos, previamente ordenados de menor a mayor*.
- El 50% de los valores centrales se destacan mediante un rectángulo (***caja***), donde el lado más largo del rectángulo muestra el recorrido intercuartílico (desde Q_1 a Q_3). Este rectángulo está dividido por un segmento vertical que indica la posición de la *Mediana* y por lo tanto su relación con los cuartiles primero y tercero. Si la Mediana está relativamente en el centro de la caja se dice que la distribución es *Simétrica*. Si la mediana está considerablemente más cerca del primer cuartil indica que los datos son positivamente asimétricos (la media suele ser mayor que la mediana en estos casos). Si la mediana está considerablemente más cerca del tercer cuartil indica que los datos son negativamente asimétricos (la media suele ser menor que la mediana en estos casos).
- Los valores extremos (el 25% de los menores y el 25% de los mayores) se representan mediante sendos segmentos llamado ***bigotes***. Los bigotes se trazan hasta abarcar la totalidad de los individuos, con la condición de que *cada lado no se alargue más de una vez y media la longitud de la caja*. *Tukey* (1997) sugiere una regla sencilla para determinar los límites de los bigotes. Existen límites interiores y límites exteriores. *Los límites interiores son barreras hasta las cuales se "permiten" datos de la muestra por estar muy cerca del resto y son los que definen los extremos de los bigotes*. Los límites se construyen de la siguiente manera:

  Límite interior inferior = Límite del bigote inferior = $Q_1 - 1'5 \cdot (Q_3 - Q_1)$

  Límite interior superior = Límite del bigote superior = $Q_3 + 1'5 \cdot (Q_3 - Q_1)$

  Límite exterior inferior = $Q_1 - 3 \cdot (Q_3 - Q_1)$

  Límite exterior superior = $Q_3 + 3 \cdot (Q_3 - Q_1)$
- Si existen valores de la variable comprendidos entre los límites interiores y exteriores se consideran valores *"atípicos"* y se indican con un asterisco (*). Si existen valores fuera de los límites exteriores se consideran valores todavía más atípicos y se indican con un punto ($\circ$).
- La caja y los bigotes se colocan paralelos a un eje rotulado a una escala determinada.

Ejemplo: **La variable (x) de la tabla adjunta representa el número de hijos de una familia y la frecuencia (f) el número de familias que tienen esa cantidad de hijos. Representar la distribución mediante un diagrama de caja.**

x	0	1	2	3	4	5	6
f	10	20	41	29	14	5	1

Completamos la tabla con las frecuencias absolutas acumuladas y con el porcentaje de frecuencias absolutas acumuladas.

x	f	F	$F \cdot \frac{100}{120}$
0	10	10	8′33
1	20	30	25
2	41	71	59′16
3	29	100	83′33
4	14	114	95
5	5	119	99′16
6	1	120	100

Calculamos los valores mínimo, máximo y los cuartiles correspondientes a las estaturas.

$$x_{min} = 0 \quad Q_1 = 1'5 \quad M = Q_2 = 2 \quad Q_3 = 3 \quad x_{máx} = 6$$

La longitud de la caja es:

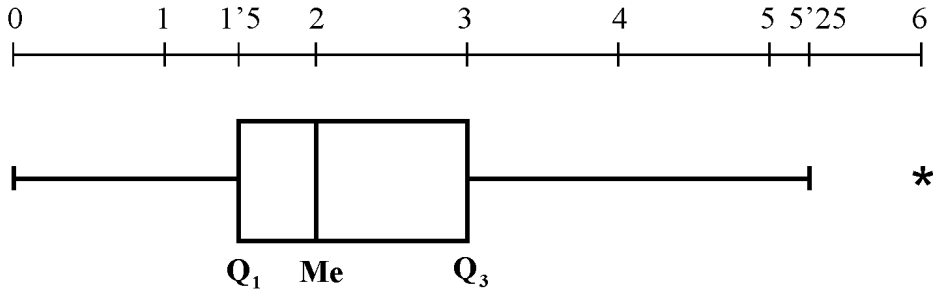
$$Q_3 - Q_1 = 3 - 1'5 = 1'5 \qquad 1'5 \cdot (Q_3 - Q_1) = 1'5 \cdot 1'5 = 2'25$$

Para calcular los extremos de los bigotes tenemos:

$$1'5 - 2'25 = -0'75 < 0 \qquad \text{y} \qquad 3 + 2'25 = 5'25 < 6$$

El valor mínimo de la distribución es 0, que es mayor que el límite inferior $-0'75$, por lo tanto 0 se tomará como valor extremo del bigote de la izquierda. El valor máximo de la distribución es 6, que es mayor que el límite superior $5'25$, por lo tanto $5'25$ se tomará como extremo del bigote de la derecha. Por tanto, los extremos de los bigotes son: 0 y $5'25$

La familia que tiene 6 hijos se encuentra fuera del extremo superior del bigote y la representamos mediante un asterisco (*). El asterisco señala la situación del individuo "atípico" de la población correspondiente al número de hijos.



De la observación del diagrama se puede concluir que:

- El bigote de la derecha es más largo que el de la izquierda, lo que indica que hay más familias que tienen menos hijos.
- Las pulsaciones que se hallan entre el 50% y el 75% de los datos (parte derecha de la caja) están más dispersas que entre el 25% y el 50% (parte izquierda de la caja).
- Al estar la mediana más cerca del primer cuartil indica que los datos son positivamente asimétricos.

Ejemplo: **Representa la siguiente distribución de las estaturas de los 40 alumnos de una clase mediante un diagrama de caja:**

149 150 154 156 157 158 159 160 160 160 161 162 162 163 163 163 163 164 165 166 166 166 167 167 167 168 168 168 169 169 170 170 170 171 172 173 174 175 175 189

Completamos la tabla con las frecuencias absolutas acumuladas y con el porcentaje de frecuencias absolutas acumuladas.

x	f	F	%
149	1	1	2′5
150	1	2	5
154	1	3	7′5
156	1	4	10
157	1	5	12′5
158	1	6	15
159	1	7	17′5
160	3	10	25
161	1	11	27′5
162	2	13	32′5
163	4	17	42′5
164	1	18	45
165	1	19	47′5
166	3	22	55
167	3	25	62′5
168	3	28	70
169	2	30	75
170	3	33	82′5
171	1	34	85
172	1	35	87′5
173	1	36	90
174	1	37	92′5
175	2	39	97′5
189	1	40	100

Calculamos los valores mínimo, máximo y los cuartiles correspondientes a las estaturas.

$$x_{min} = 149 \quad Q_1 = 160'5 \quad M = Q_2 = 166$$

$$Q_3 = 169'5 \quad x_{máx} = 189$$

La longitud de la caja es:

$$Q_3 - Q_1 = 169'5 - 160'5 = 9$$

$$1'5 \cdot (Q_3 - Q_1) = 1'5 \cdot 9 = 13'5$$

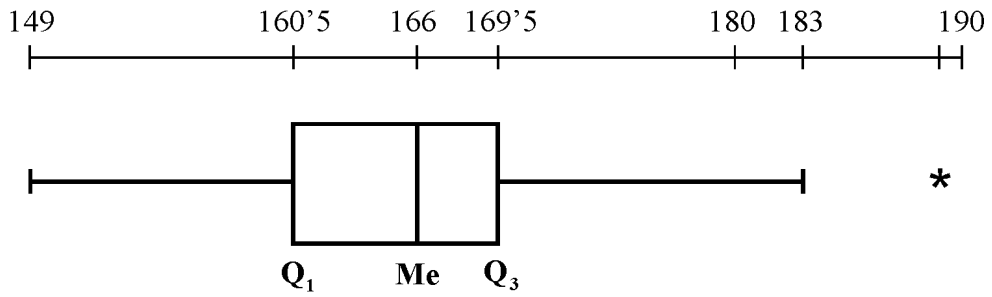
Para calcular los extremos de los bigotes tenemos:

$$160'5 - 13'5 = 147 < 149$$

$$169'5 + 13'5 = 183 < 189$$

El valor mínimo de la distribución es 149, que es mayor que el límite inferior 147, por lo tanto 149 se tomará como valor extremo del bigote de la izquierda. El valor máximo de la distribución es 189, que es mayor que el límite superior 183, por lo tanto 183 se tomará como extremo del bigote de la derecha. Por tanto, los extremos de los bigotes son: 149 y 183.

El estudiante que mide 189 cm queda fuera del extremo superior del bigote. Este valor se representa en el diagrama de caja mediante un asterisco (*), que señala la situación de los individuos “atípicos” de la población de estudiantes.



Los diagramas de caja son especialmente útiles para efectuar comparaciones entre dos o más conjuntos de datos.

Ejemplo: Comparar mediante diagramas de cajas la edad de dos colectivos de 20 personas.

Distribución 1	***Distribución 2***
35 38 32 28 30 29 27 19 48 40	**36 25 37 24 39 20 36 45 31 31**
39 24 24 34 26 41 29 48 28 22	**39 24 29 23 41 40 33 24 34 40**

Distribución 1 Haciendo la tabla correspondiente obtenemos como resultados:

$$x_{min} = 19 \quad Q_i = 26'5 \quad M = Q_2 = 29'5 \quad Q_3 = 38'5 \quad x_{máx} = 48$$

La longitud de la caja es:

$$Q_3 - Q_1 = 38'5 - 26'5 = 12'5 \qquad 1'5 \cdot (Q_3 - Q_1) = 1'5 \cdot 12'5 = 18'75$$

Para calcular los extremos de los bigotes tenemos:

$$26'5 - 18'75 = 7'75 < 19 \qquad 38'5 + 18'75 = 57'25 > 48$$

El valor mínimo de la distribución es 19, que es mayor que el límite inferior, por lo tanto 19 se tomará como valor extremo del bigote inferior. El valor máximo de la distribución es 48, que es menor que el límite superior, por lo tanto 48 se tomará como extremo del bigote superior.

Distribución 2 Haciendo la tabla correspondiente obtenemos como resultados:

$$x_{min} = 20 \quad Q_i = 24'5 \quad M = Q_2 = 33'5 \quad Q_3 = 39 \quad x_{máx} = 45$$

La longitud de la caja es:

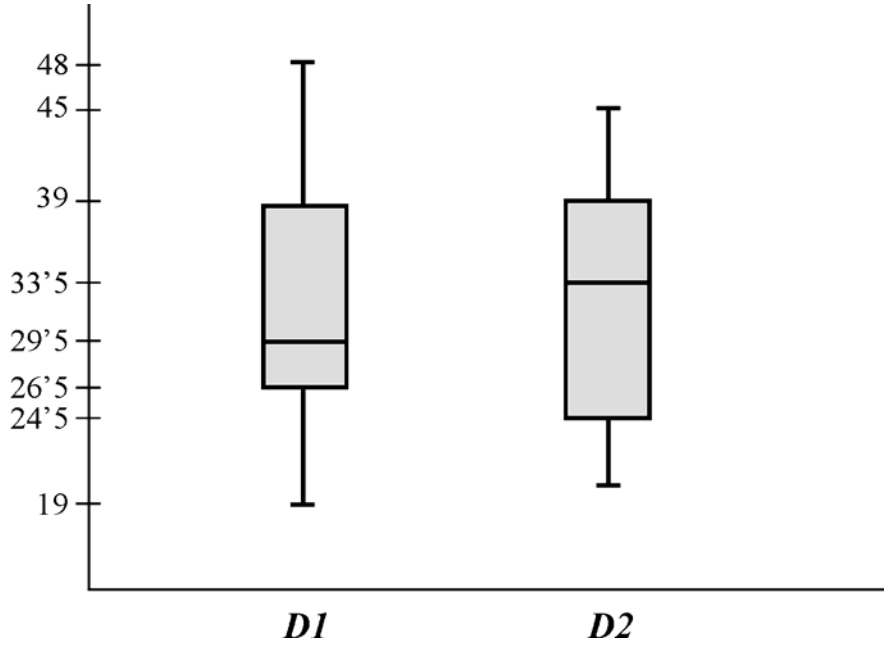
$$Q_3 - Q_1 = 39 - 24'5 = 4'5 \qquad 1'5 \cdot (Q_3 - Q_1) = 1'5 \cdot 4'5 = 6'75$$

Para calcular los extremos de los bigotes tenemos:

$$24'5 - 6'75 = 17'75 < 20 \qquad 39 + 6'75 = 45'75 > 45$$

El valor mínimo de la distribución es 20, que es mayor que el límite inferior, por lo tanto 20 se tomará como extremo del bigote inferior. El valor máximo de la distribución es

45, que es menor que el límite superior, por lo tanto 45 se tomará como extremo del bigote superior.



- Si observamos la caja de la D1, la parte inferior de la caja es menor que la superior, lo que quiere decir que las edades comprendidas entre el 25% y el 50% de la población están más agrupadas que entre el 50% y el 75%. Si observamos la caja de la D2, la parte inferior de la caja es mayor que la superior, lo que quiere decir que las edades comprendidas entre el 25% y el 50% de la población están más dispersas que entre el 50% y el 75%.
- Si observamos la caja de la D1, el bigote inferior es más corto que el superior, por ello el 25% de los más jóvenes están más concentrados que el 25% de los mayores. Si observamos la caja de la D2, el bigote inferior es más corto que el superior, por ello el 25% de los más jóvenes están más concentrados que el 25% de los mayores.
- Si observamos la caja de la D1, el rango intercuartílico es $Q_3 - Q_1 = 12$ lo que significa que el 50% de la población está comprendida en 12 años. Si observamos la caja de la D2, el rango intercuartílico es $Q_3 - Q_1 = 14'5$ lo que significa que el 50% de la población está comprendida en $14'5$ años.

¡OJO!

Es necesario destacar que, a veces, las medidas de centralización no siempre son una descripción adecuada de todos los datos de una distribución. Para ello veamos a continuación, con alguna variante, un anecdótico ejemplo expuesto por J. C. Stanley en "Measurement in Today's Schools".

En cierta ocasión se sentaron cinco hombres en un banco de un parque. Dos de ellos eran vagabundos y todos sus bienes ascendían a 10 € cada uno. El tercero era un obrero que no tenía más propiedades que una cuenta bancaria con 1000 €. El cuarto era un administrativo que entre su vivienda y su cuenta bancaria tenía unos bienes valorados en 50000 €. El quinto era un agraciado de la Primitiva que tenía un capital igual a 2400000 €. Calcular las medidas de centralización.

La serie estadística es la siguiente: 10; 10; 1000; 50000; 2400000

Si calculamos las medidas de centralización de esta serie estadística se obtiene:

$$\overline{x} = \frac{10+10+1000+50000+2400000}{5} = 490204\ €$$

$$M_0 = 10\,€ \qquad M_e = 1000\,€$$

Vemos que la media no da una idea de cómo es la distribución; tampoco la moda permite asegurar nada, pues si bien 10 € son los bienes del 40% de la distribución (los dos vagabundos), este valor se encuentra muy lejos y es prácticamente insignificante para el multimillonario de la Primitiva. Por último, con el conocimiento de la mediana, que describe muy bien el capital del obrero, nada permite afirmar de los capitales de los otros cuatro señores.

Con el fin de evitar contradicciones como la presente en la anécdota, se deben evitar grandes diferencias numéricas entre los datos de una distribución. Por otra parte, los parámetros estadísticos de una distribución informan mejor de ésta cuanto mayor es el número de datos.

Calculadora científica

Para visualizar o bajarte de Internet los manuales de los distintos modelos de las calculadoras CASIO ir a la página web: http://world.casio.com/calc/download/es/manual/

Modelo MS (fx 82, fx 83, fx 85, fx 270, fx 300 y fx 350)

a) Se pone la calculadora en modo SD. Para ello se pulsa la tecla [MODE] y la que indique la calculadora para el modo SD.

b) Se borran los datos pulsando [SHIFT] [CLR] (Scl) [1] [=]

c) Se escribe cada dato y se pulsa [DT] o [DATA]. Si la frecuencia es mayor que uno, se pulsa, después del dato la tecla [;] y se escribe la frecuencia.

d) Si se introduce un dato erróneo, se puede buscar subiendo con la tecla redonda central e introduciéndolo de nuevo.

e) Media aritmética: [SHIFT] [S-VAR] ($\bar{x}$) [1] [=]

f) Desviación típica: [SHIFT] [S-VAR] ($x\sigma_n$) [2] [=]

g) Varianza: se eleva al cuadrado la desviación típica: [x^2]

h) Cociente de variación: [$x\sigma_n$] [÷] [$\bar{x}$]

Ejemplo Calcular σ_{n-1} , σ_n , $\bar{x}$, n , $\sum x$ y $\sum x^2$ para los datos siguientes:

55 , 54 , 51 , 55 , 53 , 53 , 54 , 52

Solución Ponemos la calculadora en modo estadístico**: SD**

Borramos todos los datos anteriores: [SHIFT] [CLR] [1] (Scl) [=] (Stat clear)

Introducimos los datos como se indica en el apartado c).

55 [DT] 54 [DT] 51 [DT] 55 [DT] 53 [DT] 53 [DT] 54 [DT] 52 [DT]

Para realizar los cálculos que nos pide el problema pulsamos en la calculadora las teclas tal como se indica en los apartados e), f) y g).

Desviación estándar de muestra ($\sigma n-1$) = ***1,407885953*** SHIFT S-VAR 3 =

Desviación estándar de población (σn) = ***1,316956719*** SHIFT S-VAR 2 =

Media aritmética ($\bar{x}$) = ***53,375*** SHIFT S-VAR 1 =

Número de datos (n) = ***8*** SHIFT S-SUM 3 =

Suma de valores (Σx) = ***427*** SHIFT S-SUM 2 =

Suma de los cuadrados de los valores (Σx^2) = ***22805*** SHIFT S-SUM 1 =

Ejemplo

Se han obtenido los datos de la tabla siguiente sobre el número de personas que viven en el hogar familiar. Calcula la media aritmética, la varianza, la desviación típica, el coeficiente de variación e interpreta los resultados.

Nº de personas	x_i	3	4	5	6
Frecuencia	n_i	6	15	12	7

Scl 3 ; 6 DT 4 ; 15 DT 5 ; 12 DT 6 ; 7 DT

Media

$\bar{x}$ **4,5**

Desviación típica

$x\sigma_n$ **0,95**

Coeficiente de variación

$x\sigma_n$ ÷ $\bar{x}$ **0,21**

Los datos se distribuyen alrededor de 4,5, con una desviación típica no muy grande:

$$0,21 = 21\% < 30\%$$

Modelo ES (fx 82, fx 83, fx 85, fx 300 y fx 350)

Para activar o desactivar la columna de frecuencia (FREQ) de la pantalla del editor STAT del modo STAT, utilice el procedimiento siguiente.

Para especificar esto:	Realice esta operación de tecla:
Mostrar la columna FREQ	SHIFT MODE ▼ 3 (STAT) 1 (ON)
Ocultar la columna FREQ	SHIFT MODE ▼ 3 (STAT) 2 (OFF)

Ejemplo de tabla estadística para una variable estadística discreta

Un profesor tiene anotadas en su cuaderno las notas de los 30 alumnos de una clase.

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

a) Completa la tabla siguiente.
b) Haz un diagrama de barras y dibuja un diagrama de sectores.
c) Calcula la media, la mediana, la moda, la desviación típica y el coeficiente de variación.
d) Calcula los cuartiles, el decil 4, el percentil 80 y el percentil 60.

x_i	f_i	f_r	%	F_i	F_r	Grados $f_r \cdot 360^o$	$F_i \cdot \frac{100}{30}$
				30	**1**		**100**
	30	**1**	**100**			**360º**	

Soluciones

c) $\sum x_i f_i = 175$ $\quad \sum x_i^2 f_i = 1277$ $\quad \overline{x} = 5'8333$ $\quad \sigma = 2'9221$ $\quad M_o = 8$ $\quad M_e = 7$

$$CV = \frac{\sigma_n}{\overline{x}} = \frac{2'9221}{5'8333} = 0'5009$$

d) $Q_1 = P_{25} = 4$ $\quad Q_2 = M_e = 7$ $\quad Q_3 = P_{75} = 8$ $\quad D_4 = P_{40} = 6$ $\quad P_{80} = 8$ $\quad P_{60} = \frac{7+8}{2} = 7'5$

Ejemplo de tabla estadística para una variable estadística continua

Las edades de las personas que acuden a un logopeda a lo largo de un mes son las de la tabla adjunta.

3	**2**	**11**	**13**	**4**	**3**	**2**	**4**	**5**	**6**	**7**	**3**	
4	**5**	**3**	**2**	**5**	**6**	**27**	**15**	**4**	**21**	**12**	**4**	
3	**6**	**29**	**13**	**6**	**17**	**6**	**13**	**6**	**5**	**12**	**26**	

a) Completa la tabla siguiente.
b) Representa los datos mediante un histograma y dibuja un diagrama de sectores.
c) Calcula la media, la mediana, la moda, la desviación típica y el coeficiente de variación.
d) Calcula el cuartil superior, los deciles 3 y 7 y los percentiles 20, 40 y 90.

- Hay 36 datos, por tanto el número de intervalos que debemos formar es $\sqrt{36} = 6$.
- El valor menor de la distribución es 2 y el mayor es 29 la diferencia es 27.
- El número entero más próximo a 27 por exceso que es divisible entre 6 es 30 por tanto los intervalos deben tener de amplitud $\frac{30}{6} = 5$.
- Si comenzamos el primer intervalo en el número 2 que es el menor de la distribución, el último intervalo acaba en 32 (3 más que el mayor de la distribución que es 29). Si comenzamos el primer intervalo en el número 1, el último intervalo acaba en 31 (2 más que el mayor de la distribución). Cualquiera de las dos opciones son igualmente válidas

Intervalos	**Marcas x_i**	**f_i**	**f_r**	**%**	**F_i**	**F_r**	**Grados $f_r \cdot 360^o$**	**$F_i \cdot \frac{100}{36}$**
1 - 6								
					36	**1**		**100**
		36	**1**	**100**			**360º**	

Soluciones

c) $\sum x_i f_i = 341$ $\quad \sum x_i^2 f_i = 5321$ $\quad \overline{x} = 9'4722$ $\quad \sigma = 7'6211$ $\quad M_o = 3'5$ $\quad M_e = 6'71$

$$CV = \frac{\sigma_n}{\overline{x}} = \frac{7'6211}{9'4722} = 0'8045$$

d) $Q_3 = P_{75} = 13'143$ $\quad D_3 = P_{30} = 4'176$ $\quad D_7 = P_{70} = 11'85$ $\quad P_{20} = 3'11$ $\quad P_{40} = 5'23$ $\quad P_{90} = 23$

Estadística Inferencial. Muestra

Salvo en el caso de poblaciones pequeñas, pocas veces en una investigación se cuenta con el tiempo, los recursos y los medios para estudiar una población completa. A veces ni siquiera podemos delimitar exactamente una población, otras veces la población total "aún no existe" como sucede en los estudios sobre predicción. Estos motivos de tiempo, coste, accesibilidad a los individuos y complejidad de las operaciones de recogida, clasificación y análisis de los datos hacen que la gran mayoría de los proyectos de investigación no estudien más que una parte representativa de la población, denominada ***muestra***. Esto se puede hacer así porque, si se selecciona correctamente la muestra, ésta puede aportarnos información representativa y exacta de toda la población.

Una muestra es una parte de la población que sirve para representarla.

La estadística inferencial o inductiva plantea y resuelve el problema de establecer previsiones y conclusiones válidas generales sobre una población a partir de los resultados obtenidos de una ***muestra***. *Utiliza resultados obtenidos mediante la estadística descriptiva y se apoya fuertemente en el cálculo de probabilidades*.

Los métodos más empleados para elegir muestras son los métodos aleatorios que dependen del azar. Para determinar el tamaño de una muestra existen procedimientos estadísticos basados en la dispersión de los datos: *cuanto mayor nos parezca la desviación típica de una población, más grande debe ser la muestra que se elija*.

Ejemplo: *En un centro escolar se desea conocer el nº de horas que estudian por término medio los 1000 alumnos y alumnas del centro. Como resulta prácticamente imposible encuestarlos a todos se extrae una muestra de 100 alumnos y alumnas. Nos plantean tres métodos para extraer la muestra y debemos encontrar el más objetivo:*

a) El director elige la muestra procurando que haya alumnos de todo tipo.

b) Se eligen los 100 primeros que llegan por la mañana al centro un día cualquiera.

c) Se numeran del 1 al 1000 y se eligen, al azar, 100 de ellos.

El método a) tiene el inconveniente de que depende de la subjetividad del director, lo que nunca debe ocurrir al elegir una muestra, ya que las ideas de quien la elija influirán en las conclusiones que se extraigan. El método b) tampoco es bueno ya que es posible que los 100 primeros alumnos que lleguen a clase sean los más responsables y esto influya sobre sus métodos de estudio. El c) es el único método válido

Se dice que un muestreo es aleatorio cuando los individuos de la muestra se eligen al azar, de modo que todos los individuos de la población tienen la misma probabilidad de ser elegidos. El muestreo aleatorio es el único que garantiza la fiabilidad de las conclusiones que se obtengan.

*En la calculadora <u>la desviación típica de una **muestra**</u> viene representada por el símbolo* $\mathbf{s_x}$ ó $\boldsymbol{\sigma_{n-1}}$.

Problemas resueltos de Estadística Descriptiva

Problema 1

Las calificaciones de un grupo de 25 alumnos son:

10, 8, 3, 2, 2, 2, 4, 4, 6, 2, 7, 1, 0, 9, 5, 4, 5, 3, 7, 8, 5, 4, 2, 3, 8

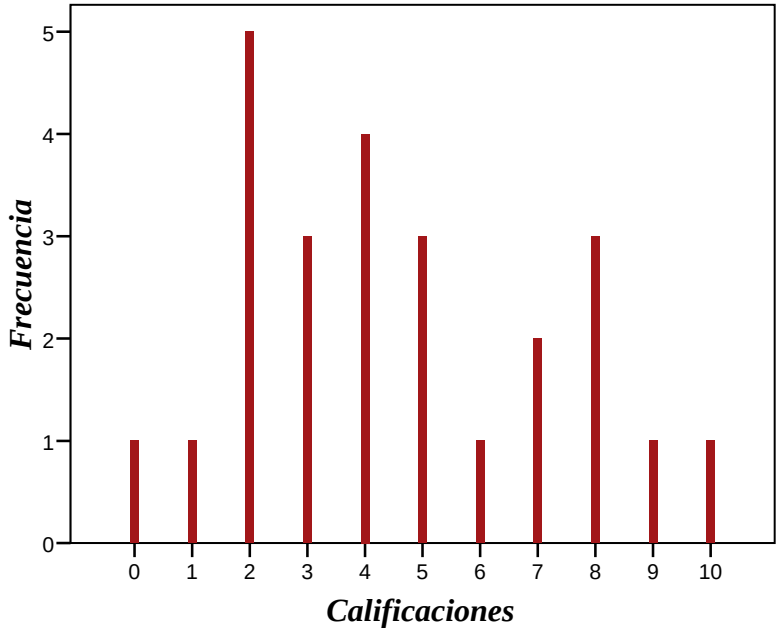
1) Calcula la media, la mediana, la moda, la desviación típica y el coeficiente de variación.
2) Representa un diagrama de frecuencias y el diagrama de sectores.
3) Calcula los cuartiles, el decil 7º y los percentiles 30 y 80.

Solución

x_i	f_i	f_r	F_i	F_r	$F_i \cdot \frac{100}{25}$
0	1	0,04	1	0,04	4
1	1	0,04	2	0,08	8
2	5	0,20	3	0,28	12
3	3	0,12	10	0,40	40
4	4	0,16	14	0,56	56
5	3	0,12	17	0,68	68
6	1	0,04	18	0,72	72
7	2	0,08	20	0,80	80
8	3	0,12	23	0,92	92
9	1	0,04	24	0,96	96
10	1	0,04	**25**	**1**	**100**
	25	**1**			

N		25
Media		4,56
Mediana		4,00
Moda		2
Desviación típica		2,6393
Varianza		6,9664
Rango		10
Mínimo		0
Máximo		10
Percentiles	25	2,00
	30	3,00
	50	4,00
	70	6,00
	75	7,00
	80	7,50

$$CV = \frac{\sigma_n}{\overline{x}} = \frac{2'6393}{4'56} = 0'5787$$



Problema 2

Realizada una encuesta en una ciudad, se han agrupado los establecimientos hoteleros por el número de plazas que permiten, obteniéndose la siguiente tabla:

Plazas	nº de hoteles
< 100	25
100 – 200	37
200 – 300	12
300 – 500	22
500 – 600	21
600 – 700	13
700 – 800	5
800 – 1000	3

1) **Calcula la media aritmética, la mediana, la moda, la desviación típica y el coeficiente de variación.**
2) **Representa un histograma de frecuencias y el diagrama de sectores correspondiente.**
3) **Calcula los cuartiles, los deciles 2 y 8 y los percentiles 41 y 72.**
4) **¿Qué centil corresponde a 325 plazas?**

Solución

Intervalos	x_i	f_i	d_i	f_r	F_i	F_r	$F_i \cdot \frac{100}{138}$
0 - 100	50	25	0,25	0,181	25	0,181	18,11
100 - 200	150	37	0,37	0,268	62	0,449	44,92
200 - 300	250	12	0,12	0,087	74	0,536	53,62
300 - 500	400	22	0,11	0,159	96	0,695	69,56
500 - 600	550	21	0,21	0,152	117	0,847	84,78
600 - 700	650	13	0,13	0,094	130	0,941	94,20
700 - 800	750	5	0,05	0,036	135	0,977	97,82
800 - 1000	900	3	0,015	0,022	**138**	**1**	**100**
		138		**1**			

N		138
Media		326,44
Mediana		258,39
Moda		150
Desviación típica		233,84
Varianza		54680,8
Rango		850
Mínimo		50
Máximo		900
Percentiles	20	104
	25	126
	41	185
	50	258
	72	516
	75	536
	80	569

Los valores de los percentiles deben ser números enteros, por ser la variable discreta.

A 325 plazas le corresponde el percentil 56.

$$CV = \frac{\sigma_n}{\overline{x}} = \frac{233'84}{326'44} = 0'7163$$

Problema 3

Las tallas de los 40 alumnos de una clase se encuentran en la tabla adjunta.

149	154	156	158	158	159	160	160
165	165	165	166	167	168	168	168
168	168	168	158	169	170	170	171
172	173	175	175	178	160	161	162
162	163	163	163	163	164	165	165

1) **Calcula la media aritmética, la mediana, la moda, la desviación típica y el coeficiente de variación, sin datos agrupados y con ellos.**
2) **Haz la representación gráfica del diagrama de frecuencias, histograma y diagrama de sectores**
3) **Calcula los cuartiles, el decil 9º y los centiles 30 y 70.**
4) **¿Qué centil corresponde a 162'6 cm?**

Solución con datos agrupados

Intervalos	x_i	f_i	f_r	F_i	F_r	$F_i \cdot \frac{100}{138}$
146 – 151	148,5	1	0,025	1	0,025	2,50
151 – 156	153,5	1	0,025	2	0,050	5,00
156 – 161	158,5	8	0,200	10	0,250	25,00
161 – 166	163,5	13	0,325	23	0,575	57,50
166 – 171	168,5	11	0,275	34	0,850	85,00
171 – 176	173,5	5	0,125	39	0,975	97,50
176 - 181	178,5	1	0,025	**40**	**1**	**100**
		40	**1**			

N		40
Media		164,875
Mediana		164,846
Moda		163,5
Desv. típ.		6,0194
Varianza		36,2343
Rango		30,0
Mínimo		148,5
Máximo		178,5
Percentiles	25	158,50
	30	161,76
	50	164,84
	70	168,27
	75	169,18
	90	173,00

A 162′6 cm. le corresponde el percentil 36.

$$CV = \frac{\sigma_n}{\bar{x}} = \frac{6'0194}{164'875} = 0'036$$

Solución sin datos agrupados

x_i	f_i	f_r	F_i	F_r	$F_i \cdot \frac{100}{40}$
149	1	0,025	1	0,025	2,5
154	1	0,025	2	0,050	5,0
156	1	0,025	3	0,075	7,5
158	3	0,075	6	0,150	15,0
159	1	0,025	7	0,175	17,5
160	3	0,075	10	0,250	25,0
161	1	0,025	11	0,275	27,5
162	2	0,050	13	0,325	32,5
163	4	0,100	17	0,425	42,5
164	1	0,025	18	0,450	45,0
165	5	0,125	23	0,575	57,5
166	1	0,025	24	0,600	60,0
167	1	0,025	25	0,625	62,5
168	6	0,150	31	0,775	77,5
169	1	0,025	32	0,800	80,0
170	2	0,050	34	0,850	85,0
171	1	0,025	35	0,875	87,5
172	1	0,025	36	0,900	90,0
173	1	0,025	37	0,925	92,5
175	2	0,050	39	0,975	97,5
178	1	0,025	**40**	**1**	**100**
	40	**1**			

N		40
Media		164,80
Mediana		165,00
Moda		168
Desv. típ.		5,9338
Varianza		35,21
Rango		29
Mínimo		149
Máximo		178
Percentiles	25	160,50
	30	162,00
	50	165,00
	70	168,00
	75	168,00
	90	172,50

A 162′6 cm. le corresponde el percentil 38.

$$CV = \frac{\sigma_n}{\bar{x}} = \frac{5'9338}{164'80} = 0'036$$

Problemas sobre Estadística Descriptiva

1) **Construir un diagrama de sectores y de barras para la distribución de enfermos mentales ingresados en un hospital psiquiátrico durante un año.**

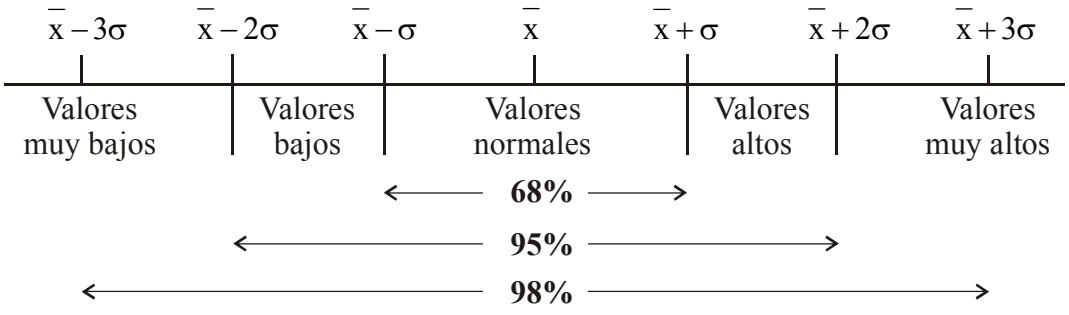
Enfermedades	Casos
Esquizofrenia	115
Psicosis maniacodepresiva	20
Neurosis	60
Demencia	33
Alcohólicos	70
Otros	210

2) **En los países europeos se está cuestionando la posibilidad de reducir la jornada laboral a 35 horas. En una empresa automovilística se ha confeccionado la siguiente tabla de la distribución estadística del salario/hora en euros de los obreros.**

Salario hora	6 – 7'5	7'5 – 9	9 – 10'5	10'5 – 12	12 – 13'5
% obreros	10	20	38	25	7

a) Calcula la media, la mediana, la moda y la desviación típica.
b) Dibuja el histograma correspondiente y el diagrama de sectores.
c) Calcula los cuartiles y el percentil 60.

3) *Cuando el histograma tiene forma de campana, llamada Campana de Gauss, se dice que la variable aleatoria se distribuye normalmente (altura de los individuos de una población, peso, coeficiente intelectual, etc.). En una distribución normal se puede hacer una clasificación de la población utilizando estos intervalos*:



En un grupo de Matemáticas se han obtenido las siguientes puntuaciones en un test de habilidad mental:

50, 23, 45, 36, 56, 34, 56, 67, 45, 34, 23, 45, 23, 67, 54, 21, 34, 43, 12, 78, 36, 49, 53, 27, 66, 31, 45, 22, 33, 44, 48, 53, 57, 77, 31, 23, 47, 52, 33, 37, 64, 21.

a) Comprobar si en el intervalo $\bar{x}-\sigma$, $\bar{x}+\sigma$ se encuentra aproximadamente el 68% de los datos.

b) Comprobar si en el intervalo $\bar{x}-2\sigma$, $\bar{x}+2\sigma$ se encuentra aproximadamente el 95% de los datos.

c) Comprobar si en el intervalo $\bar{x}-3\sigma$, $\bar{x}+3\sigma$ se encuentra aproximadamente el 98% de los datos.

4) **En un estudio estadístico sobre la altura de los españoles y los alemanes se han obtenido los siguientes resultados:**

	Españoles	Alemanes
Media	172'2 cm.	175'4 cm.
Desviación típica	4'4 cm.	6'5 cm.

¿Quién es más alto, un español que mide 177cm o un alemán que mide 181cm?

5) **La distribución de estaturas de una muestra de 40 individuos nacidos en el mismo año y en el mismo mes viene dada por la tabla siguiente:**

Estatura (cm)	151	156	161	166	171	176
Nº de individuos	2	4	11	14	5	4

Clasifica los 40 individuos de una forma razonada en altos, normales y bajos.

6) **Se realiza una estadística en dos centros de enseñanza, uno público y otro privado, referente a la nota global del bachillerato de cada uno de los alumnos que van a acudir a los exámenes de selectividad. Las distribuciones de frecuencias son las siguientes:**

Centro privado

Nota global de cada alumno.	Frecuencias
5,5	10
6.5	15
7.5	20
8.5	30
9.5	15

Centro público

Nota global de cada alumno.	Frecuencias
[5 , 6]	250
(6 , 7]	150
(7 , 9]	100
(9, 10]	20

a) ¿Cuál es el motivo de que los datos se presenten en dos tablas de diferente tipo?
b) Calcula la media y la desviación típica de las dos distribuciones.
c) Representa gráficamente los diagramas que mejor se ajusten a cada distribución.
d) Un alumno del centro privado tiene una nota global de un 8'5 y otro del centro público una nota de un 7. ¿Cuál de los dos es mejor alumno comparativamente?

7) **A la finalización del curso "Informática e Internet" se realizó un examen tipo test a los 300 alumnos obteniéndose la siguiente tabla relativa al número de preguntas acertadas:**

Nº preguntas acertadas	Nº de alumnos
0-10	10
10-15	20
15-20	60
20-23	100
23-25	70
25-30	30
30-40	10

a) Representa gráficamente la distribución de frecuencias anterior.
b) Hallar la media, la mediana, la moda y la desviación típica.
c) Calcula los cuartiles, los deciles 3 y 8 y los percentiles 30, 50 y 80.

Para la realización de la segunda parte del curso se convocan sesenta plazas.

Una vez finalizado este segundo curso, se realiza un examen a los alumnos obteniéndose las siguientes notas:

Notas	Nº Alumnos
4	8
5	12
5.5	15
6	14
6.5	6
8	5

a) ¿Por qué no se agrupan los datos en intervalos, como anteriormente?
b) Calcular la media, la mediana, moda y la desviación típica.
c) ¿Qué resulta más meritorio, obtener 28 preguntas acertadas en el primer examen u obtener un 6.5 en el segundo?

8) **El volumen de ventas de una empresa de telefonía en el año 2002 se reparte de la siguiente manera:**
- **dentro de la telefonía móvil fue de 7'51 millones de euros, mientras que la media en el sector fue de 6'61 millones de euros y la varianza de 86'5.**
- **en el caso de la empresa de telefonía fija, las ventas fueron de 8'41 millones de euros, siendo la media en su sector de 7'2 millones de euros y la varianza de 117'79.**

¿Cuál de estas dos empresas está mejor situada en cuanto a su volumen de ventas? Razone la respuesta.

9) **Una alumna de primer curso de Economía, tras los exámenes de Febrero, quiere saber en qué asignatura de las cursadas en el primer cuatrimestre ocupa una mejor posición relativa según la nota obtenida. Para satisfacer su curiosidad dispone de la siguiente información:**

Asignaturas	*Nota obtenida por la alumna*	*Nota media de la asignatura*	*Desviación típica de las Notas de la asignatura*
Estadística	7,0	6,0	1,2
Matemáticas	6,5	6,0	1,7
Tª Económica	6,0	5,0	2,0
Contabilidad	7,2	7,0	1,4
Derecho Civil	8,5	7,5	2,1
Historia Económica	9,0	8,0	1,3

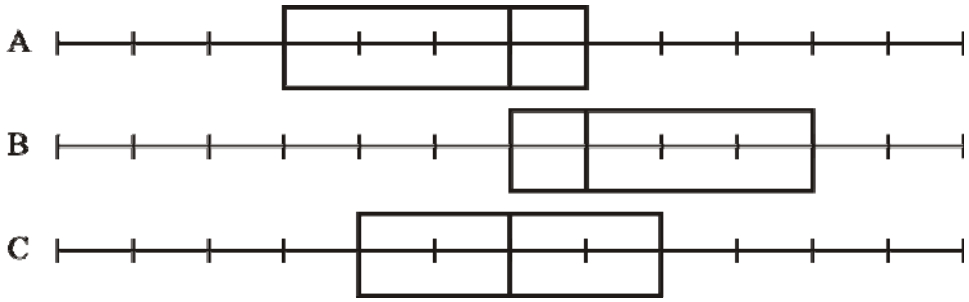
Determinar en qué asignatura está situada en una mejor posición relativa.

10) **El gasto de dos grupos de familias durante un periodo de tiempo ha sido el siguiente**:

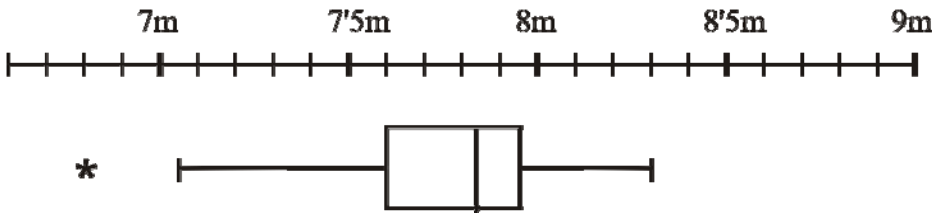
GRUPO A		***GRUPO B***	
Gasto	***Nª de familias***	***Gasto***	***Nº de familias***
10	**14**	**8**	**10**
12	**16**	**10**	**20**
14	**20**	**11**	**25**
16	**15**	**13**	**20**
18	**18**	**15**	**10**
20	**17**	**18**	**10**
		20	**5**

Determinar cuál de los dos grupos es más homogéneo respecto a su gasto, con explicación de los pasos aplicados y de los resultados obtenidos.

11) **Los siguientes diagramas de caja corresponden a las variables A, B y C. Interpreta en cada uno de ellos todo lo relativo a concentración de datos, dispersión y simetría**.



12) **Interpreta el siguiente diagrama de caja, relativo a marcas de saltadores de longitud**.



13) **En la tabla adjunta se dan los datos de los alumnos matriculados en los centros de la ciudad de Denia en Infantil, Primaria y Secundaria para el curso 2008/2009.**

a) Elabora un diagrama de barras y un diagrama de sectores donde se reflejen los alumnos matriculados en Educación Infantil por centros.

b) Elabora un diagrama de barras y un diagrama de sectores donde se reflejen los alumnos matriculados en Educación Primaria por centros.

c) Elabora un diagrama de barras y un diagrama de sectores donde se reflejen los alumnos matriculados en la ESO según los distintos cursos, y aparte elabora otro diagrama de barras y uno de sectores con el total de alumnos de la ESO por centros.

d) Calcula el total de alumnos de cada uno de los centros correspondientes a las 8 primeras filas y representa los datos en un diagrama de barras y un diagrama de sectores.

e) Con los datos de la tabla haz un análisis de la situación de la ecuación Infantil, Primaria y Secundaria en Denia.

	Educación Infantil	Educación Primaria	1º ESO	2º ESO	3º ESO	4º ESO
Cervantes	146	307				
Llebeig	197	414				
Montgó	173	318				
Pou de la Muntanya	148	291				
Les Vessanes	98	151				
Carmelitas	152	315	64	60	60	55
Maristas	152	320	60	60	52	
Paidos	153	324	60	66	60	60
La Xara	57	79				
Jesús Pobre	16	28				
Alfa y Omega	76	145	23	23	24	24
Historiador Chabás			150	159	182	98
María Ibars			121	120	105	110
3r Instituto			106	60	15	