11

Estadística

Objetivos

En esta quincena aprenderás a:

- Distinguir los conceptos de población y muestra.
- Diferenciar los tres tipos de variables estadísticas.
- Hacer recuentos y gráficos.
- Calcular e interpretar las medidas estadísticas de centralización más importantes.
- Calcular las principales medidas de dispersión.
- Entender la importancia de la elección de la muestra para que sea representativa.

1.Estadística descriptiva ..... pág. 184
Población y muestra
Variables estadísticas
Gráficos variables cualitativas
Gráficos variables cuantitativas discretas
Gráficos variables cuantitativas continuas

2.Medidas de centralización ..... pág. 187
Media, moda y mediana
Evolución de la media
Evolución de la mediana
Media y mediana comparadas
Medidas de posición

3.Medidas de dispersión ..... pág. 190
Desviación típica y recorrido
Cálculo de las medidas de dispersión
La media y la desviación típica

4.Representatividad de las muestras..pág. 192
Muestreo estratificado
Muestreo aleatorio. Sesgo

Ejercicios para practicar

Para saber más

Resumen

Autoevaluación

Actividades para enviar al tutor

Antes de empezar

Recuerda

El curso pasado ya estudiaste estadística, y en numerosas ocasiones has hecho estadística aunque no te hayas dado cuentas de ello. Veamos algunos ejemplos.

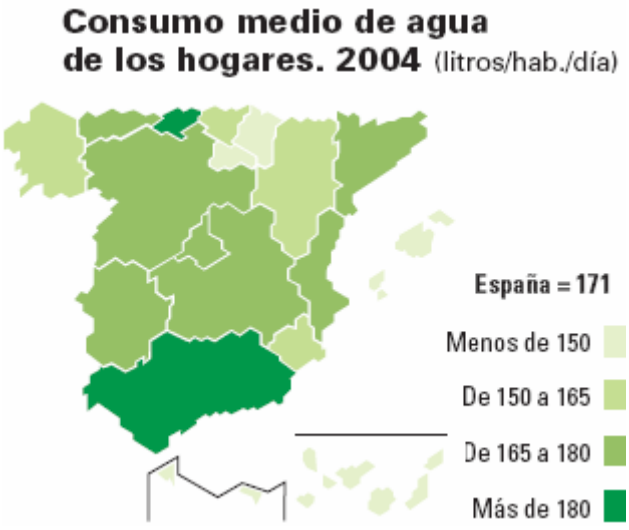
Nota media

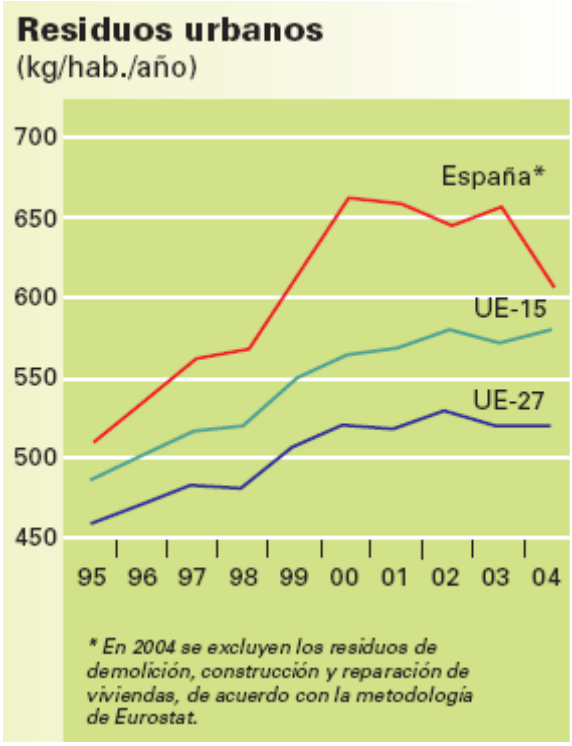
A lo largo de un curso escolar tendrás muchas ocasiones donde calcular este valor. Si una nota depende de dos exámenes y en uno tienes un 4, intentarás sacar al menos un 6 en la otra.

Al final del instituto, las medias del bachillerato y de la prueba selectividad. Comparaciones con la media local o nacional. Las medias de corte para determinadas carreras

Fútbol

El jugador que más goles ha marcado, el portero que menos ha encajado. La clasificación de la liga. La mejor mitad de liga. Los puestos de competiciones europeas, los de descenso, nº de veces internacional, nº de fases finales, minutos jugados, tiros a puerta, faltas.





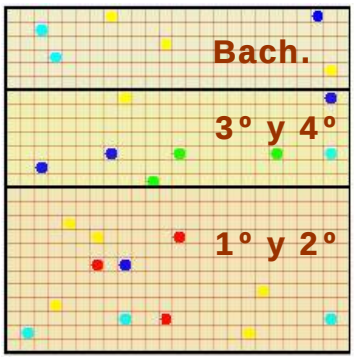
1. Estadística descriptiva

Población y muestra.

Población es el conjunto de individuos, con alguna característica común, sobre el que se hace un estudio estadístico.

La **muestra** es un subconjunto de la población, seleccionada de modo que ponga de manifiesto las características de la misma, de ahí que la propiedad más importante de las muestras es su representatividad.

El proceso seguido en la extracción de la muestra se llama **muestreo**



Si cada cuadrito representa a cada uno de los alumnos de un instituto ficticio y se les pregunta sobre su color favorito, el total de los cuadros es la población, 625 alumnos, y los 26 encuestados constituyen la muestra.

Variables estadísticas

La característica a estudiar en una población es la **variable estadística**.

Las variables estadísticas pueden ser esencialmente de dos tipos **cualitativas y cuantitativas.**

Las variables cualitativas son las que no aparecen en forma numérica sino como una categoría o atributo.

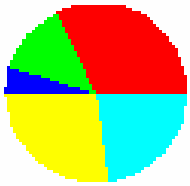
Las variables cuantitativas son las que pueden expresarse numéricamente, y a su vez pueden ser:

- ✓ Cuantitativas discretas, si sólo pueden tomar un número finito de valores.
- ✓ Cuantitativas continuas cuando pueden tomar cualquier valor de un intervalo.

- El color de los ojos, el queso preferido, el continente donde vives, son **variables estadísticas cualitativas**.
- El nº de ordenadores en casa, o de televisores y el nº de habitantes por vivienda, por ejemplo, son variables estadísticas **cuantitativas discretas.**
- El peso, la altura, la velocidad, la densidad, la presión, son **variables estadísticas cuantitativas continuas.**

Los datos:

xi	fi
	7
	3
	1
	6
	5
Total	22



Tienen este diagrama de sectores

Gráficos en variables cualitativas.

El diagrama de sectores es el mas indicado para este tipo de información. El porcentaje de datos de cada valor en una muestra se corresponde con el mismo porcentaje de sector de un círculo. Así por ejemplo, si los datos son A, A, A, A, A, B, B, B, C y C. Las frecuencias son (A,5), (B,3) y (C,2), los porcentajes serán (A,50%), (B,30%) y (C,20%) los que corresponde a un gráfico de sectores con (A, 180º), (B,108º) y (C, 72º).

$$\frac{\text{frecuencia}}{\text{nº total de datos}} = \frac{\text{grados del sector}}{360}$$

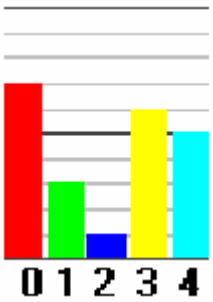
Gráficos en variables discretas.

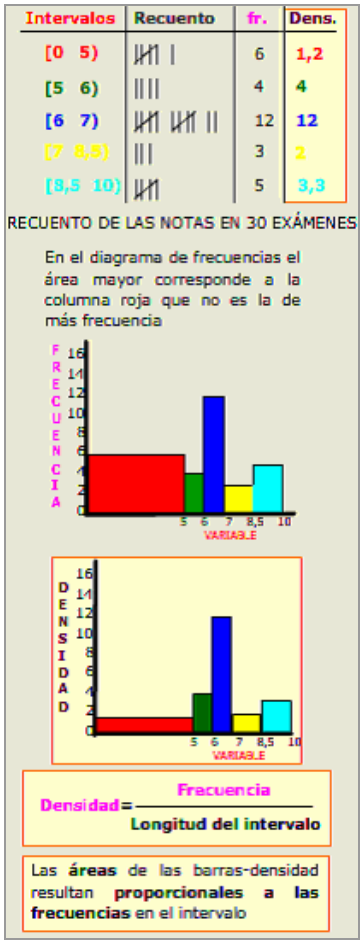
Diagrama de barras. Bastará que observes un ejemplo.

A los datos,

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

les corresponde el gráfico de la derecha.





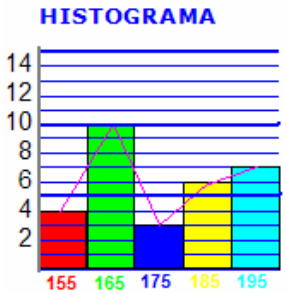
Gráficos en variables continuas.

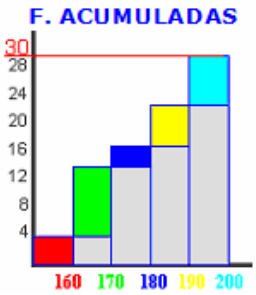
Histograma. Los datos se representan por rectángulos cuya base es la amplitud del intervalo representado y con la altura que nos indica la frecuencia absoluta, si todos los intervalos son de la misma amplitud. Si no es el caso, las alturas se calculan de manera que las áreas sean proporcionales a las frecuencias absolutas. A la izquierda tienes un ejemplo hecho.

Polígono de frecuencias. Uniremos los centros de la parte superior de todos los rectángulos para obtenerlo.

También se suele dibujar el histograma de las **frecuencias acumuladas**, en cada dato se acumula la frecuencia de los datos anteriores.

[150, 160]→4
[160, 170]→10
[170, 180]→3
[180, 190]→6
[190, 200]→7





EJERCICIOS resueltos

1. Clasifica las los siguientes ejemplos de variables estadísticas: Longitud de un camión, Carga máxima, nº de ruedas, nº de ejes, tipo de camión, marcas de neumáticos, tipo de tapicería, nº de puertas, altura máxima.

   Cualitativas: Tipo de camión, marcas de neumáticos, tipo tapicería
   C. discretas: Nº de ruedas, nº de ejes, nº de puertas
   C. continuas: Longitud de un camión, Carga máxima y altura máxima.

2. Calcula los grados que corresponden a cada valor en un gráfico de sectores hecho a partir de los datos: R, R , V , V , V , V , V , A, A y A

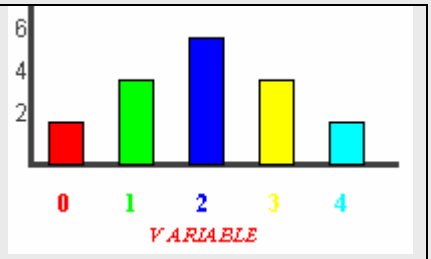
   Hacemos el recuento R → 2, V → 5 y A→3 Y calculamos

   $\frac{2}{10} = \frac{\text{Grados R}}{360}$, $\frac{5}{10} = \frac{\text{Grados V}}{360}$ y $\frac{3}{10} = \frac{\text{Grados A}}{360}$ y obtenemos

   Grados R = 72, Grados V = 180 y Grados A =108

3. Agrupa los datos siguientes y haz un diagrama de barras adecuado. Datos = { 0 1 0 2 3 4 1 2 2 1 2 2 3 4 3 2 1 3 }

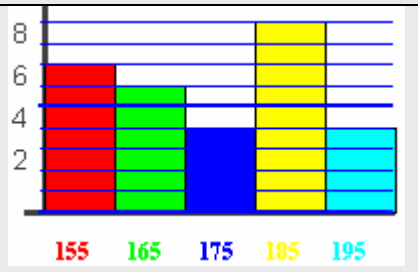
Marca	Frecuencia
0	2
1	4
2	6
3	4
4	2

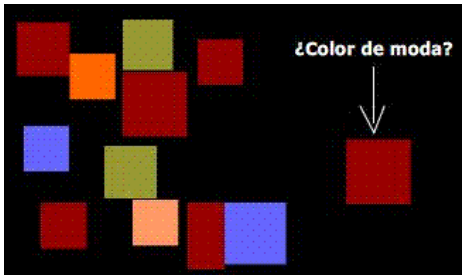


4. Clasifica los datos en intervalos y dibuja un histograma adecuado.

180	197	154	181	189	162	152	162	167	190
189	160	166	197	187	194	152	181	173	154
177	184	186	174	177	159	158	189	160	150

[150, 160] 155→ 7
[160, 170] 165→ 6
[170, 180] 175→ 4
[180, 190] 185→ 9
[190, 200] 195→ 4





Por ejemplo, si tenemos las observaciones 6,7,8,6,7,6,8,6,9 y agrupamos los datos vemos claramente que el valor 6 aparece mas que ningún otro. En este caso la **moda** es 6.

xi→fr
6→4
7→2
8→2
9→1

Si ordenamos los datos, y dado que el nº de datos es impar justo el 7 queda en el centro.

6 6 6 6 7 7 8 8 1

Si los datos fueran 6,7,8,6,7,6,8,6,5 una vez ordenados, y como hay una cantidad par de datos, dos de ellos ocuparían el centro:

5 6 6 6 6 7 7 8 8 1

y la mediana será (6+7)/2 = 6.5

2. Medidas de centralización

Media, mediana y moda.

Un conjunto N de observaciones, N números, puede que por si solo no nos diga nada. En cambio, si además nos dicen que están situados alrededor de uno o varios valores centrales ya tenemos una referencia que sintetiza la información.

Media. La suma de los N números dividida entre N. Por ejemplo, para **3**, **4** y **5**, (**3**+**4**+**5**)/3 = 12/3 = 4; para **1**, **1**, **4**, **8**, **8** y **8**, (**1**·2+**4**+**8**·3)/6=5**.**

$$\text{Media} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \ldots + x_n \cdot f_n}{f_1 + f_2 + \ldots + f_n}$$

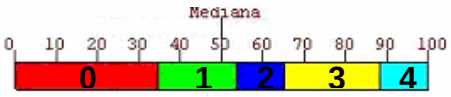
Moda. Si una observación se repite más que cualquier otra, será considerada la moda de esos datos. Por ejemplo, si tenemos las observaciones 6,7,8,6,7,6,8,6,9 y agrupamos los datos 6→4, 7→2, 8→2 y 9→1 vemos claramente que el valor 6 aparece mas que ningún otro. En este caso la moda es 6.

En el caso de variable continua, consideraremos por moda a la marca del intervalo de mayor frecuencia, cuando esto ocurra. También puede ocurrir que haya dos modas o que no haya ninguna que destaque.

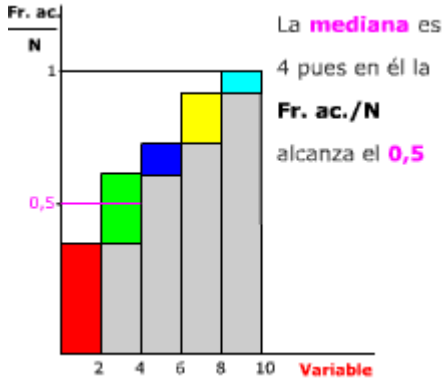
Mediana. El número tal que la mitad de las observaciones son mayores que él y la otra mitad menores.

En general, para pocos datos lo mejor es proceder según el ejemplo de la izquierda, según sea una cantidad para o impar.

Para cantidades mayores, habrá que agrupar los datos primero en una tabla. Y determinar segmentos de longitud proporcional a su frecuencia, disponerlos de forma lineal y marcar el centro como muestra el siguiente ejemplo.

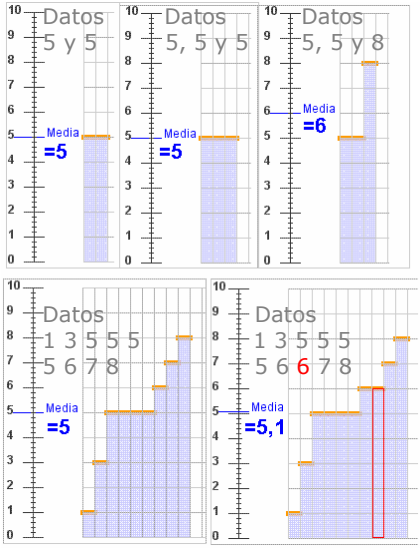


En este otro gráfico vemos indicada la mediana en un diagrama de Frecuencias relativas acumuladas:



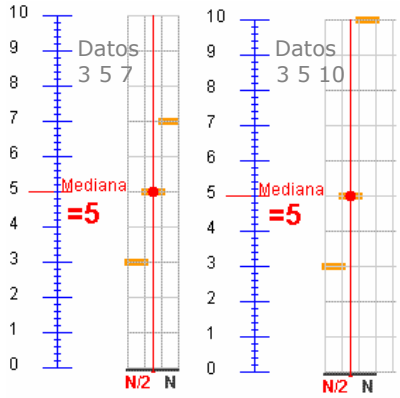
Media. Evolución de esta al añadir y/o cambiar un dato.

1 Para los datos 5 y 5 la media es 5. Si añadimos un 5 se mantiene en 5. Si añadimos un 8 la media pasa a ser 6. (Figura derecha).
2 Si tenemos 9 datos con media 5, necesitamos añadir un 6 para que la media pase a ser 5,1. Si tenemos 19 datos con media 5, necesitamos un dato de valor 7 para que la media suba a 5,1. (Figura derecha).
3 Para un conjunto de datos con media 5, si añadimos otro con media 5, por ejemplo 6 y 4, el nuevo conjunto conserva la media.



Mediana. Evolución de esta al añadir y/o cambiar un dato.

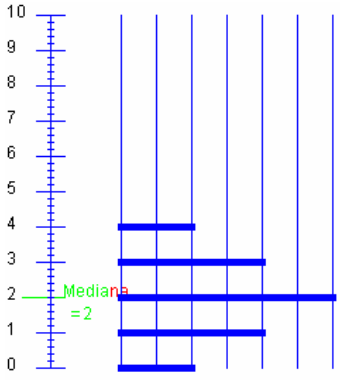
1 La mediana, para los datos 2, 3 y 4 es Me=3. Si cambiamos el 4 por 5 o por 6 o por cualquier otro valor mayor sigue siendo Me = 3.
2 En cambio, si añadimos otro dato y tenemos 2, 3, 4 y 4, por ejemplo, la Me = 3,5. Y si ahora añadimos un quinto valor, un 4 o un 5 o un 6 o cualquier otro mayor que 4, la mediana en 2,3, 4, 4 y ?? pasa a ser 4. Da igual el valor ?? es 5, 10 o 25**.**



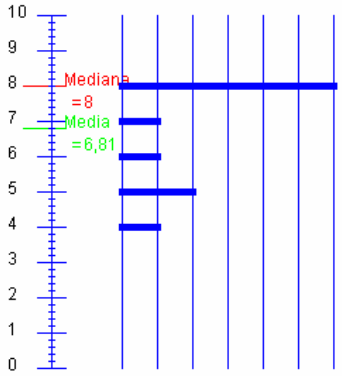
Para ver la mediana se traza una vertical desde el eje horizontal en N/2

Media y mediana comparadas

Para los datos 4 y 6 la media y la mediana coinciden en 5. Añadir un 8 o un 11 da lo mismo para la mediana, que pasa a ser en ambos casos 6. Sin embargo la media con un 8 pasa a ser 6 y con un 11 pasa a ser 7. Los valores 8 y 11 se consideran observaciones atípicas, están distanciados del resto de valores, tiran de la media y no afectan a la mediana. Si los datos estuvieran repartidos simétricamente respecto a un valor, ese valor seria a la vez la media y la mediana. En cambio, si los valores a un lado de la mediana están más alejados de ella que los del otro lado, la media se desplaza hacia esos valores alejados que tiran de ella. Hay una asimetría.



Datos simétricos
Mediana igual a la Media



Datos atípicos
Mediana distinta de la Media

Por ejemplo, si tenemos las observaciones

1. 20, 24 y 28.
   Me = 24
2. Y para 20, 24, 28 y 30
   Me = (24+28)/2 = 26
3. Para 20, 24, 28 y 100
   Me = (24+28)/2 = 26

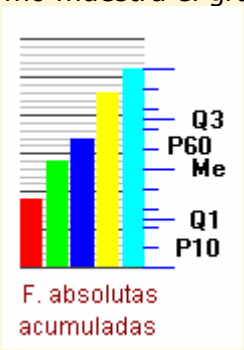
En cambio la media no se comporta de la misma forma para los mismos datos

1 $\overline{X} = 24$

2 $\overline{X} = 25{,}5$

3 $\overline{X} = 43$

También podemos hacer un diagrama de frecuencias acumuladas y dividir en partes iguales como muestra el gráfico.

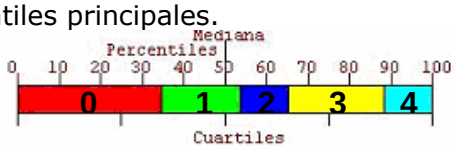


Medidas de posición: cuartiles y percentiles

Dado un conjunto de datos numéricos además de la mediana podemos considerar otras medidas de posición

- Si nos fijamos en el primer valor que supera al 25% o al 75%, estamos hablando del **primer y tercer cuartil, Q_1 y Q_3.**
- Para otros valores como el 10%, o el 80% hablamos de **percentiles, P_{10} y P_{80} .**

Ejemplo. Para la variable de valores 0, 1, 2, 3, 4, y frecuencias 0→9, 1→5, 2→3, 3→6, 4→3, dibujamos barras de longitud proporcional a las frecuencias y dividimos el total en partes iguales: en dos partes para la mediana, cuatro para los cuartiles y 10 para los percentiles principales.



EJERCICIOS resueltos

5. Calcula la media en cada caso:

a) 4, 6, 8
b) 4, 6, 8, 6
c) 100, 120, 180, 200

Soluciones: a) (4+6+8)/3 = 6
b) (4+6+8+6)= 24/4 = 6
c) (100+120+180+200)/4 = 150

6. Calcula la media en cada caso:

a

Marca	Fr
10	2
20	4
30	3
40	2

b

Marca	Fr
100	2
200	4
300	3
400	2

a) $\overline{X} = \dfrac{10\cdot2 + 20\cdot4 + 30\cdot3 + 40\cdot2}{11} = 24{,}54$

b) $\overline{X} = \dfrac{100\cdot2 + 200\cdot4 + 300\cdot3 + 400\cdot2}{11} = 245{,}45$

7. Determina la moda y la mediana

a) 5,6,6 c) 1,2,3,4,2
b) 1,1,2,3 d) 3,2,3,2,2,2

Soluciones: a) Me=6, Mo=6 c) Me=2 Mo=2
b) Me=1,5 Mo=1 d) Me=2 Mo=2

8. Calcula la moda y la mediana en cada caso:

a

Marca	Fr
10	2
20	4
30	3
40	2

b

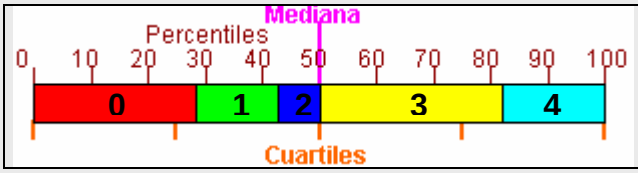
Marca	Fr
100	2
200	3
300	4
400	1

Soluciones:
a) Me=20 Mo=4
b) b) Me=250 Mo=300

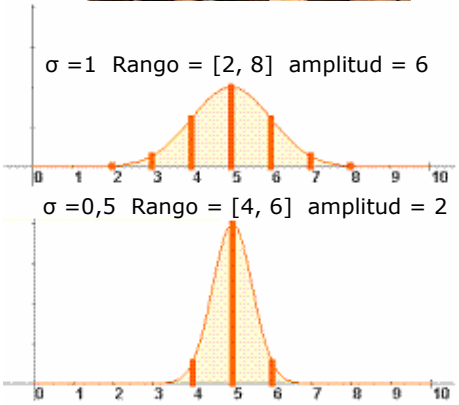
9. Calcula la mediana, cuartiles primer y 3º, y el percentil 30 60 y 90 de los datos.

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

Hacemos el recuento: 0→8, 1→4, 2→2, 3→9 y 4→5 y barras de longitud proporcional a la frecuencia para cada valor. Además partimos la longitud total de la barra en 2, 4 y 10 trozos para obtener la mediana, cuartiles y percentiles, tal y como muestra la imagen.



Vemos que la mediana está entre el azul y el amarillo, (3+2)/2 = 2.5, Q1 entre rojo y verde, Q1=0, Q3=3, p30=1, P60=3 y P90=4



En ambos gráficos la media, mediana y moda valen 5

En la práctica se suele usar la fórmula reducida para el cálculo de la desviación típica.

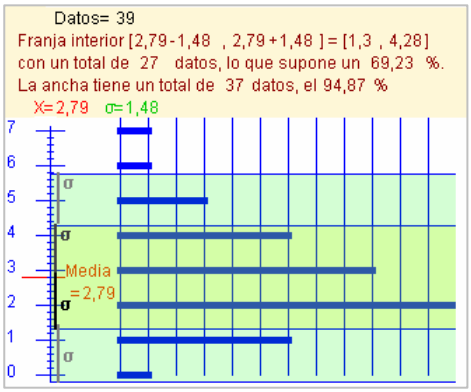
$$\sigma = \sqrt{\frac{\sum fi \cdot Xi^2}{n} - \bar{X}^2}$$

Así, para

Marca	Fr
4	3
5	3
6	2

Se tiene que la media $\bar{X} = 4,85$ y

$$\sigma = \sqrt{\frac{3 \cdot 4^2 + 3 \cdot 5^2 + 2 \cdot 6^2}{8} - 4,85^2}$$



3. Medidas de dispersión.

Varianza, Desviación típica y rango

"La estadística es una ciencia según la cual, si yo me como un pollo y tú no te comes ninguno, nos hemos comido como promedio medio pollo cada uno".

La estadística indicará que todos comen lo mismo cuando las medidas de dispersión sean todas nulas.

Rango. El intervalo definido por el menor y el mayor dato. También se llama rango a la diferencia entre el mayor y el menor de los datos.

Varianza. La media aritmética de los cuadrados de las diferencias de los datos con la media.

$$\sigma^2 = \frac{\sum fi \cdot (Xi - \bar{X})^2}{n} \text{ que equivale a } \sigma^2 = \frac{\sum fi \cdot (Xi)^2}{n} - (\bar{X})^2$$

Desviación típica. La raíz cuadrada positiva de la varianza.

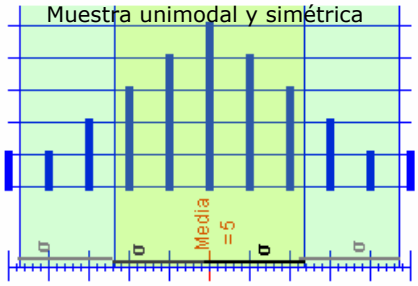
$$\sigma = \sqrt{\frac{\sum fi \cdot (Xi - \bar{X})^2}{n}} \quad \text{o} \quad \sigma = \sqrt{\frac{\sum fi \cdot Xi^2}{n} - \bar{X}^2}$$

Medir la dispersión

Ese es el objetivo de estas medidas. Por ejemplo, los datos A= {20, 20}, B={15, 20, 20, 25} tienen la misma media, moda y mediana. En todos los casos igual a 20. Sin embargo, puedes comprobar que en ninguna de las tres medidas de dispersión definidas arriba coinciden.

Media y desviación típica.

Para muestras unimodales (una sola moda) y casi simétricas, alrededor de la media podemos considerar un intervalo que contenga la mayoría de los datos. Por ejemplo, para una muestra con media 100 y desviación típica 10, la mayor parte de los datos estarán entre 90 y 110, aproximadamente el 68% ; entre 80 y 120 estará el 95% aproximadamente. Y casi todos entre 70 y 130. Hay una forma de distribución de datos llamada **normal** que cumple con lo anterior, y de una manera u otra, de todas las poblaciones grandes se pueden extraer datos que se ajustan a ella. En cursos superiores verás la importancia de estas distribuciones.



EJERCICIOS resueltos

10. Calcula la media y la desviación típica en
a) 200, 250
b) 175, 275
c) 250, 250

a) $\overline{X} = \frac{250+200}{2} = 225$ $\quad \sigma = \sqrt{\frac{(250-225)^2+(200-225)^2}{2}} = \sqrt{\frac{25^2+25^2}{2}} = 25$

b) $\overline{X} = \frac{175+275}{2} = 225$ $\quad \sigma = \sqrt{\frac{(175-225)^2+(275-225)^2}{2}} = \sqrt{\frac{50^2+50^2}{2}} = 50$

c) $\overline{X} = \frac{250+250}{2} = 250$ $\quad \sigma = \sqrt{\frac{(250-250)^2+(250-250)^2}{2}} = \sqrt{\frac{0^2+0^2}{2}} = 25$

11. Calcula la media y la desviación típica en:
a) 7, 5 , 3, 2, 4, 5
b) 20, 25, 20, 22, 21

a) $\overline{X} = \frac{7+5+3+2+4+5}{6} = \frac{26}{6} = 4,33$

$$\sigma = \sqrt{\frac{7^2+5^2+3^2+2^2+4^2+5^2}{6} - 4,33^2} = \sqrt{\frac{128}{6} - 18,75} = 1,59$$

b) $\overline{X} = \frac{20+25+20+22+21}{5} = \frac{108}{5} = 21,6$

$$\sigma = \sqrt{\frac{20^2+25^2+20^2+22^2+21^2}{5} - 21,6^2} = \sqrt{\frac{2350}{5} - 466,56} = 1,85$$

(Nota.- Observa la fórmula utilizada para la desviación)

12. Organiza los datos siguientes en intervalos de 10 cm desde 150 a 200. Amplia la tabla con dos columnas, una para el producto de las marcas con las frecuencias y otra para el producto de las frecuencias con los cuadrados de las diferencias con la media. Calcula la media y la desviación típica.

174	158	150	185	186	178	166	185	199
183	175	173	175	164	173	178	179	164
176	159	190	173	189	163	156	169	

	xi	fi	xi·fi	$fi\cdot(xi-\overline{X})^2$
[150,160)	155	5	775	1733,65
[160,170)	165	5	825	371,58
[170,180)	175	10	1750	19,02
[180,190)	185	7	1295	906,42
[190,200)	195	2	390	914,14
Total		29	5035	3944,82

Con los datos de la tabla es mas fácil, y se tiene:

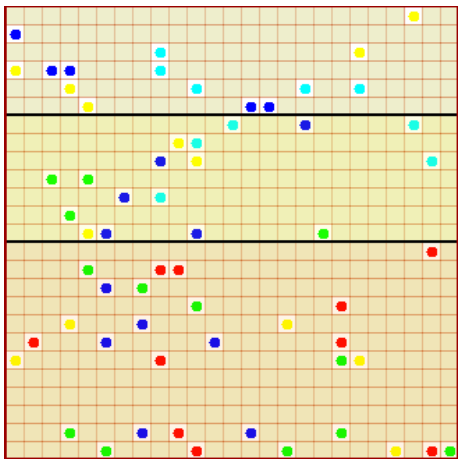
Media y Desviación típica

$\overline{X} = \frac{5035}{29} = 173,62$ $\quad \sigma = \sqrt{\frac{3944,82}{29}} = 11,66$

4. Representatividad

Representatividad. Muestreo estratificado.

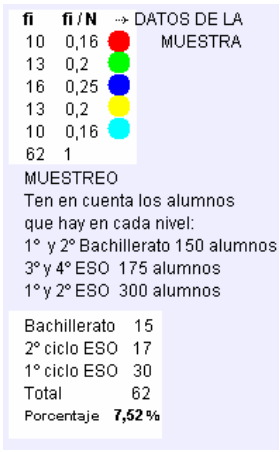
REPRESENTATIVIDAD. Una muestra es representativa de la población cuando en ella podemos encontrar las mismas proporciones de las características de estudio que en el conjunto de la población. El proceso de elegir una muestra, a qué individuos elegimos como representantes de la población, es el punto importante y de ello va a depender que el estudio sea útil o no (representativo o no).



Elegir bien la muestra no es sinónimo de representatividad, pero elegirla mal casi si es sinónimo de no representatividad.

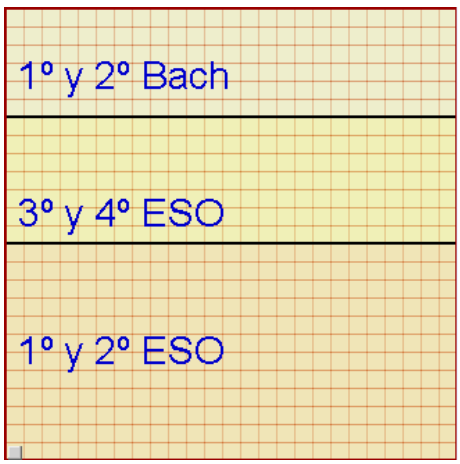
Por ejemplo, si queremos estudiar el poder adquisitivo de una población, y solo elegimos a individuos de una determinada zona, o principalmente de una determinada zona, la muestra con toda seguridad no será representativa. La muestra se ha de elegir tomando muestras de individuos proporcionales a la población de cada zona. Si hay tres zonas con 12.000, 18.000 y 20.000 habitantes, la muestra deberá tener un 24% de la primera zona, 36% de la segunda y 40% de la última.

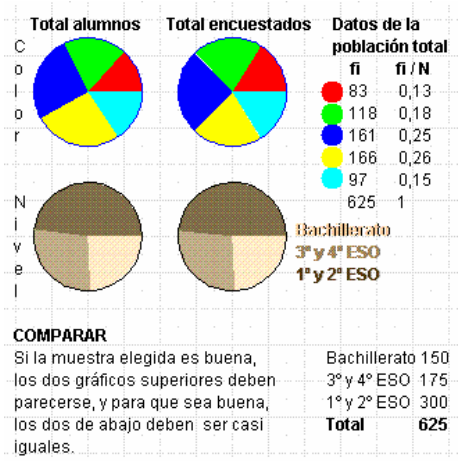
Este tipo de **muestreo**, escogiendo un reparto proporcional a los estratos, se llama **estratificado.**



Ejemplo

En la imagen tienes 625 cuadros que representan a los alumnos de un instituto ficticio





A la derecha vemos la muestra estratificada que se ha elegido y el resultado de la encuesta. Los últimos diagramas de sectores comparan la realidad con los resultados de la encuesta.

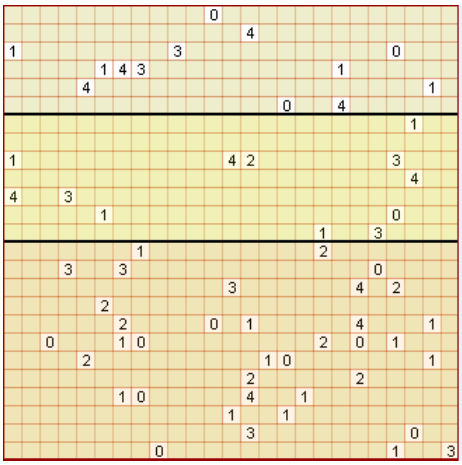
Sesgo. Muestreo aleatorio

Sesgo. Se dice que la muestra está sesgada cuando hay diferencia entre los datos de la muestra y los datos de toda la población.

Ejemplo: Llamadas telefónicas voluntarias. Estas encuestas tienen varias fuentes de sesgo. Hay familias que no tienen teléfono, el coste de la llamada no todo el mundo está dispuesto a asumirlo. Pero sobre todo, el factor de respuesta voluntaria, los encuestados se auto-seleccionan. Suelen contestar aquellos con una fuerte opinión negativa sobre el tema. El enojo les anima a participar.

Muestreo aleatorio total. A diferencia del estratificado, que guarda las proporciones, esta forma de elegir la muestra considera a toda la población y elige individuos aleatoriamente. Se considera una buena forma de proceder.

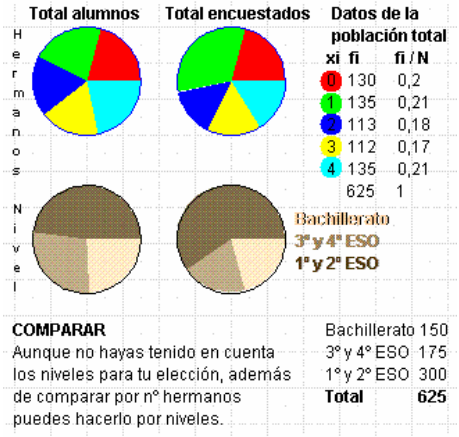
En el siguiente ejemplo se ha escogido con ordenador una muestra aleatoria total entre los 625 alumnos de un instituto, este muestreo puede salir estratificado o no, en el ejemplo no salió muy bien estratificado.



xi	fi	fi / N	→ DATOS DE LA MUESTRA
0	13	0,2	
1	20	0,32	
2	9	0,14	
3	10	0,16	
4	10	0,16	
	62	1	

EN ESTE MUESTREO
No tienes que tener en cuenta los niveles, solo que cada alumno sea elegido de entre todos aleatoriamente. Aún así habrá correlacion en los niveles entre muestra y poblacion.

Bachillerato	13
2° ciclo ESO	12
1° ciclo ESO	37
Total	62
Porcentaje	7,84%



COMPARAR
Aunque no hayas tenido en cuenta los niveles para tu elección, además de comparar por n° hermanos puedes hacerlo por niveles.

Bachillerato	150
3° y 4° ESO	175
1° y 2° ESO	300
Total	**625**

xi	fi	fi / N	→ DATOS DE LA MUESTRA
0	13	0,2	
1	20	0,32	
2	9	0,14	
3	10	0,16	
4	10	0,16	
	62	1	

EN ESTE MUESTREO
No tienes que tener en cuenta los niveles, solo que cada alumno sea elegido de entre todos aleatoriamente. Aún así habrá correlacion en los niveles entre muestra y poblacion.

Bachillerato	13
2° ciclo ESO	12
1° ciclo ESO	37
Total	62
Porcentaje	7,84%

EJERCICIOS resueltos

13. Una gran empresa tiene trabajadores en cuatro áreas. Operarios, Representantes, administración y dirección. Las condiciones de trabajo son bastantes diferentes en cada área, por lo que el grado de satisfacción no es igual en cada una de ellas. Para averiguarlo, si hay 1000, 500, 300 y 200 trabajadores en las áreas de operarios, representantes, administrativos y directivos, ¿cuántos hay que seleccionar de cada área para una muestra de tamaño?

a) 200
b) 100
c) 300

a) De un total de 2000 empleados, los porcentajes para operarios, repartidores, administrativos y directivos son del 50%, 25%, 15% y 10%. Lo cual hace que la muestra tome 100 operarios, 50 repartidores, 30 administrativos y 20 directivos.
b) 50, 25, 15 y 10.
c) 150, 75, 45 y 30

Algunos de los ejercicios propuestos a continuación están elaborados a partir de esta publicación de INE. Puedes ver artículos similares en
http://www.ine.es/prodyser/pubfolletos.htm

4/2007

cifras INE

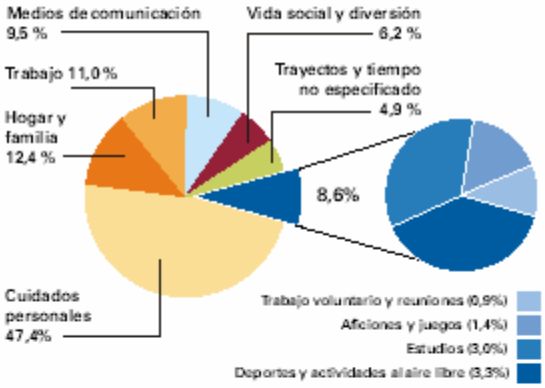
Boletín informativo del Instituto Nacional de Estadística

Encuesta de Empleo del Tiempo

Encuesta de Empleo del Tiempo

Qué hacemos y durante cuánto tiempo

Distribución del tiempo por actividades



NOTA: Los informantes de 10 y más años han anotado las actividades realizadas en un día concreto (de lunes a domingo) elegido al azar. El tiempo así estimado se refiere a un "día promedio" obtenido al concentrar todas las actividades de todos los informantes en un solo día. Los datos que aquí se presentan se refieren a toda la población investigada, salvo que se indique expresamente lo contrario.

Más información en:
www.ine.es

DEPÓSITO LEGAL: M-13543-2001 ISSN: 1579-2277 NIPO: 685-07-005-1

El Instituto Nacional de Estadística (INE) presenta en esta publicación algunos de los principales resultados de la **Encuesta de Empleo del Tiempo, primera y única encuesta de ámbito nacional sobre la utilización del tiempo**. Se realizó en España entre los años 2002 y 2003 de manera armonizada con las de otros países europeos, siguiendo las recomendaciones de la Oficina Estadística de la Unión Europea (Eurostat). Entre los años 1998 y 2004 otros países de la Unión llevaron a cabo investigaciones similares.

La encuesta facilita información, entre otras cosas, del **porcentaje de personas que realizan una determinada actividad en el transcurso del día y la duración media diaria dedicada a esa actividad por dichas personas**. Esta información primaria nos permite analizar con rigor la dimensión del trabajo no remunerado realizado por los hogares, la distribución de las responsabilidades familiares en el hogar, la participación de la población en actividades culturales y de ocio, etc. Por otra parte, la información recogida también permite comparar **datos nacionales** de uso del tiempo **en relación con los demás países europeos** que han realizado la encuesta.

Como principales resultados, cabe destacar el dato de que **las tareas domésticas y el cuidado de niños y ancianos son tareas eminentemente femeninas, ya que el 93% de las mujeres las realizan, frente al 70% de los varones.** En el contexto europeo, es de señalar **la primera posición de España en tiempo dedicado a caminar y pasear; pero también el último lugar por lo que se refiere a tiempo dedicado a la lectura.**

El INE quiere aprovechar esta ocasión para expresar su **agradecimiento a los cerca de 24.000 hogares de la muestra**, y pone a su disposición los resultados obtenidos.

Fuentes estadísticas utilizadas:
Procedentes del INE: Encuesta de Empleo del Tiempo. La información internacional procede de **Eurostat**.

Para practicar

1. Agrupa las siguientes variables: a)Peso, b)densidad, c)nº de plantas de los edificios, d)Tipo de fachada de los edificios, e)nº de ventanas, f)metros de fachada, g)nº de habitantes por edificio, h)tipo de puerta principal.

2. Escribe tres variables cualitativas que tengan que ver con embarcaciones.

3. Escribe tres variables cuantitativas discretas que tengan que ver con aviones.

4. Escribe tres variables cuantitativas continuas que tengan que ver con trenes.

5. Si las frecuencias para R, V, A y T son R→3, V→2, A→4 y T→1 ¿Cuántos grados le corresponde a cada letra en un gráfico de sectores?

6. Haz una tabla y un gráfico de sectores de los datos: R R A A R A R V N V R N

7. Haz una tabla y un gráfico de barras con los datos:
   3 3 4 5 4 5 3 2 1 2 3 4 5 4 5 4 3 3 4 4

8. Agrupa los datos siguientes en intervalos

195	194	194	182	168	179	191	154	177	189
184	187	155	167	177	187	161	171	190	162
190	152	166	180	156	186	184	167	184	162

9. Haz un histograma de los datos del ejercicio anterior

10. Calcula la media en cada caso:
    a) 4, 6, 8
    b) 4, 6, 8, 6
    c) 100, 120, 180, 200

11. Calcula la media en cada caso:

a)

Marca	Fr
1	3
2	5
3	3
4	2

b)

Marca	Fr
1000	3
2000	5
3000	3
4000	2

12. Determina la moda y la mediana
    a) 50,60,60
    b) 12,12,22,32
    c) 10,20,30,40,20
    d) 35,25,35,25,25,25

13. Calcula la moda y la mediana en cada caso:

a)

Marca	Fr
100	5
200	4
300	6
400	3

b)

Marca	Fr
100	2
200	7
300	9
400	2

14. ¿Cuál o cuáles de los datos siguientes se puede considerar una observación atípica en cada una de las dos series?
    a) 4 5 6 5 7 8 4 5 8 7 5 12 6 7 6 5 4
    b) 8 9 1 9 8 9 7 9 6 7 8

15. Calcula la mediana, primer y tercer cuartil y el percentil 90 de
    1 1 4 3 3 4 2 2 5 3 1 2 1 2 2 4 2 2 4 3 1

16. Calcula la mediana, primer y tercer cuartil y el percentil 20 de
    3 1 1 1 4 1 5 3 1 3 3 4 5 5 4 4 2 1 4 4

17. Calcula la media y la desviación típica en cada uno de los siguientes casos:
    100 y 100
    99 y 101
    110 y 90
    120 y 80

18. Completa la tabla con los datos:

190	151	193	187	158	175	165	158	184	172
197	161	157	157	183	180	150	161	182	169
162	177	160	155	188	157	189	167	186	157

Intervalo	Marca xi	Frec. fi	fi·xi	fi(X-xi)^2
[150,160)	155			
[160,170)	165			
[170,180)	175			
[180,190)	185			
[190,200)	195			

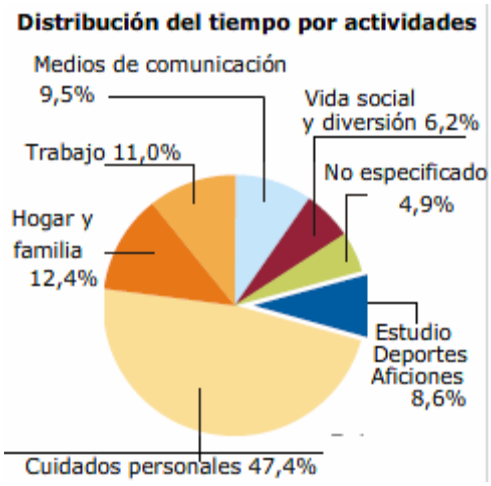
19. Determina la media y la desviación típica, de los datos de la tabla anterior.

20. Determina los intervalos $(\bar{X}-\sigma, \bar{X}+\sigma)$ y $(\bar{X}-2\sigma, \bar{X}+2\sigma)$ y el número de elementos que hay en cada uno.

Marca	Fr
0	5
1	4
2	7
3	3
4	2

21. Observa los siguientes gráficos y responde a las preguntas de cada uno

a)

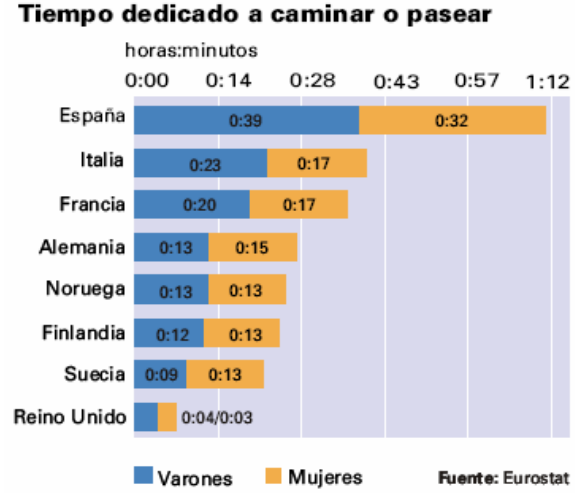


a1. ¿Cuál es la variable estudiada? ¿ y la frecuencia?

a2. ¿A qué grupo de actividades dedicamos más tiempo los españoles?

a3. Calcula cuánto tiempo dedicamos al hogar y la familia ¿cuántos grados ocupa este sector en el diagrama?

b)

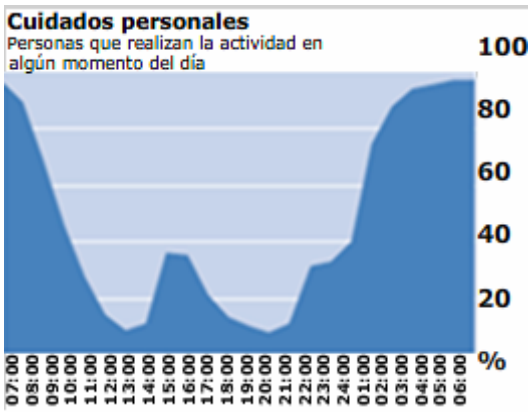


b1. ¿En qué países pasean más las mujeres que los hombres?

b2. Calcula el tiempo medio que se dedica en cada país a pasear.

b3. ¿Qué país está en el percentil 50?

c)

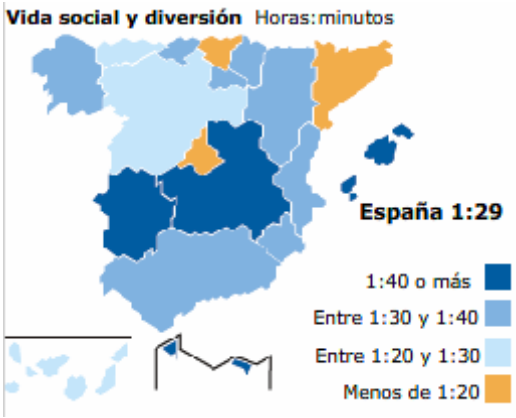


c1. ¿Crees que el dormir se ha contado como actividad de cuidado personal?

c2. A las 15:00 hay un máximo local en la gráfica ¿a qué se debe?

c3. A la hora de la comida el 38% de las personas se dedica al cuidado personal. Significa esto que un 62% de las personas no come?

d)



d1. ¿Cuáles son las comunidades en las que se dedica menos tiempo a la vida social y a la diversión

d2. ¿Cuánto tiempo dedican a la diversión o a la vida social la mayor parte de las comunidades?

d3. ¿Cuál es el tiempo medio que se dedica en España a esta actividad?

Para saber más

La profesión de enfermería.

Florence Nightingale (1820-1910), conocida por ser la fundadora de la profesión de enfermería. Durante la guerra de Crimea se percató de que la causa principal de las muertes de heridos en combate era la falta de medidas sanitarias. Al aplicarlas, la tasa de mortalidad pasó de un 42,7% a un 2,2%. Gracias a un uso eficaz de los datos consiguió modificar el sistema de atención sanitaria a su vuelta a Gran Bretaña. Cambió el sistema de registro de datos y fue una de las primeras personas en utilizar los gráficos estadísticos para representar los datos de una forma sencilla de forma que hasta los parlamentarios y generales pudieran entender.

Para Florence, los datos no eran algo abstracto, eran una forma de poder salvar vidas humanas.

El padre de la estadística.

Sir Ronald A. Fisher (1890-1962) está considerado el padre de la estadística. Los escritos de Fisher ayudaron a organizar la estadística como campo de estudio preciso cuyos métodos se aplican a problemas prácticos de muchas disciplinas. Como casi todos los pioneros en la estadística, sus trabajos nacieron de la necesidad de resolver problemas prácticos.

Inferencia estadística

La estadística desarrollada en este tema es lo que se conoce como estadística descriptiva, en ella se recoge información y se hacen cálculos que describen como están repartidos. Pongamos el caso que una muestra elegida al azar nos da una media. ¿La verdadera media está próxima a la de la muestra? Si considero un intervalo alrededor de la media muestral, la verdadera ¿con qué probabilidad estará o no en él? De estas preguntas y otras se encarga la inferencia estadística.

Principales campos de aplicación de la estadística

La estadística se aplica en muchos campos como en **Industria y empresas**. Para el control de calidad en la producción en cadena, para el análisis de mercados, para el estudio de precio de venta al público de los artículos fabricados, en gestión financiera,...

En la parte derecha se citan algunas otras de sus aplicaciones.

Algunos campos de aplicación de la estadística

Administración pública

A través de las Delegaciones territoriales y provinciales, se recogen datos para analizarlos y someterlos a procesos estadísticos. De esta forma se conocen datos referidos a nacimientos, defunciones, matrimonios, precios, salarios, trabajo, enseñanza, sanidad,... Todos estos datos se suelen publicar por el INE.

Economía.

En este campo es imprescindible, sobre todo en macro-magnitudes.

Psicología.

La mayor parte de los trabajos científicos en psicología experimental tienen como principal herramienta de trabajo la estadística.

Medicina.

En cualquier estudio experimental de estas áreas Existe una asignatura específica llamada Bioestadística para cubrir esos estudios experimentales. En Genética y antropometría encontramos dos de los campos de mayor aplicación.

Recuerda lo más importante

Población. Alumnos de un instituto ficticio.

Muestra. Alumnos encuestados

Variables estadísticas: Cualitativa, color preferido; Cuantitativa discreta, nº de hermanos y cuantitativa continua, altura.

Consideremos las dos muestras siguientes:

Nº de hermanos: 4 3 2 3 1 2 0 2 0 1 2 3 1 2 4 0 1 1 4 1 1 4 0 4 2 0 4 1

Altura: 182 172 157 194 150 166 163 196 167 199 172 185 172 168 173 160 162 173 161 192 156 164 173 180 193 172

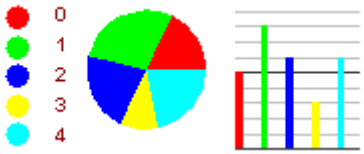
Recuento de datos:

Marca	
0	5
1	8
2	6
3	3
4	6
	28

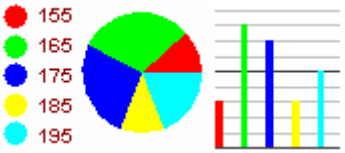
Intervalo	Xi	Fi
[150,160[	155	3
[160,170[	165	8
[170,180[	175	7
[180,190[	185	3
[190,200[	195	5
Total		26

Gráficos de **sectores** y **barras**

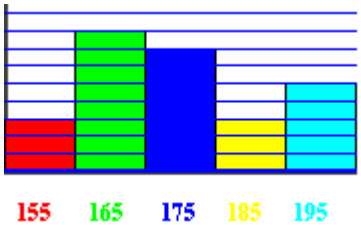
Nº de hermanos



Altura.



Histograma



Media y **moda** y **desviación típica**

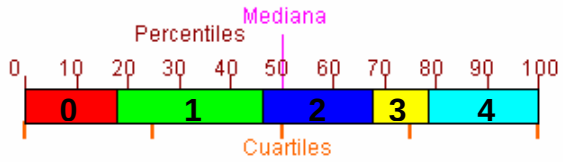
	xi	fi	xi·fi	fi·(xi-$\overline{X}$)^2
	0	5	0	0
	1	8	8	6,37
	2	6	12	0,06
	3	3	9	3,67
	4	6	24	26,64
	Total	28	53	54,67

Media = $\overline{X} = \frac{53}{28} = 1.89$

Moda=Mo=1

$\sigma = \sqrt{\frac{54.67}{28}} = 1.39$

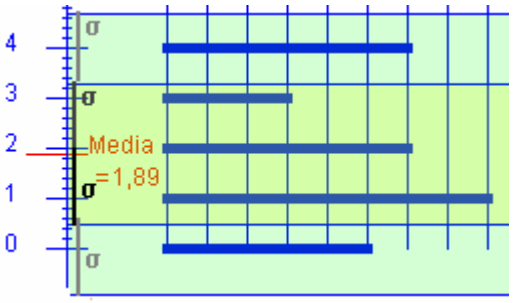
Cuartil, mediana, percentil



Me=2, Q1=1, Q3=3, P20=1, P60=2, P90=4

Recorrido. De 0 a 4, de amplitud 4

Media y desviación En nuestro ejemplo, 17 de 28 datos no se alejan de la media mas de la desviación típica, son el 60,7%, y el 100% no se alejan de la media mas de dos veces la desviación.



Representatividad

Una muestra es representativa de la población cuando en ella podemos encontrar las mismas proporciones de las características de estudio que en el conjunto de la población.

Autoevaluación

1 ¿Cuántos grados corresponden en un diagrama de sectores a la marca 2?

Xi	Fi
1	4
2	4
3	7
4	5

2 ¿La frecuencia mayor, en la tabla anterior, corresponde a la marca?

3 ¿Cuál es la moda ?

Xi	Fi
15	40
25	45
35	37
45	51

4 ¿Cuál es la mediana ?

Xi	Fi
100	4
200	4
300	7
400	5

5 ¿Cuál es el percentil 30 ?

Xi	Fi
1	4
2	4
3	7
4	5

6 ¿Cuál es la media de los datos anteriores?

7 ¿Cuál es la desviación típica del los datos del nº5?

8 ¿Cuál es la media?

Xi	Fi
180	40
200	25
220	27
240	50

9 ¿Cuál es la desviación típica de los datos anteriores?

10 ¿Cuál es el percentil 70?

Soluciones de los ejercicios para practicar

1. Cualitativas: d) h)
   Cuantitativas discretas c) e) g)
   C. continuas: a) b) f)

2. Propulsión, Carga, Tipo de travesía

3. Nº de pasajeros, nº ruedas, nº ventanas

4. Velocidad máxima, carga máxima, potencia.

5. R→108º, V→72º, A→144º y T→36º

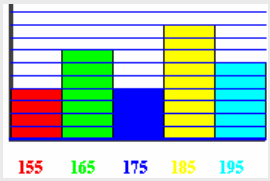
6. R→5, A→3, V→2, N→2

7. 1→1, 2→2, 3→6, 4→ 7, 5→4)

8.

Intevalo	Xi	
[150,160)	155	4
[160,170)	165	7
[170,180)	175	4
[180,190)	185	9
[190,200)	195	6

9. ---------->



10. a) 6 b) 6 c) 150

11. a) 2.3 b) 2307

12. a)Mo=60, Me=60 b)Mo=12, Me=17
    c)Mo=20, Me=20 d)Mo=25 Me=25

13. a) Mo=300, Me=250 b) Mo=300, Me=300

14. a) 12 b) 1

15. Me= 2, Q1=2, Q3=3, P90=4

16. Me=3, Q1=1, Q3=4 y P20=1

17. La media es 100 en los 4 , y la desviación 0, 1, 10 y 20.

18.

Intervalo	Marca xi	Frec. fi	fi·xi	fi(X-xi)^2
[150,160)	155	9	1395	2401
[160,170)	165	7	1155	280,77
[170,180)	175	3	525	40,33
[180,190)	185	8	1480	1494,22
[190,200)	195	3	585	1680,33
		30	5140	5896,66

19. $\bar{x} = 171,\hat{3}$ $\sigma \approx 14.02$

20. En (0.42, 2.9) hay 11, y en (-0.88, 4.14) todos

21. a1)variable:actividades. Fr:porcentaje de tiempo diario que se dedica a cada actividad
    a2)cuidados personales
    a3) 2h 58m 34s 44,64grados
    b1) Alemania, Suecia y Finlandia
    b2) E35,5 I20, F18,5 A14 N13 F12,5 S11 R3,5 en minutos
    b3) Francia
    c1) Sí. c2) Comida y Siesta
    c3) No, el pico ocupa dos horas y algunos comen en media hora
    d1) País Vasco, Cataluña y Madrid
    d2) entre 1:30 y 1:40 horas:minutos
    d3) 1:29

Soluciones AUTOEVALUACIÓN

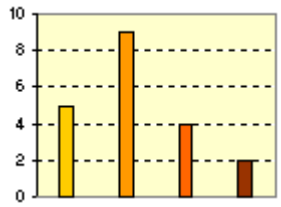
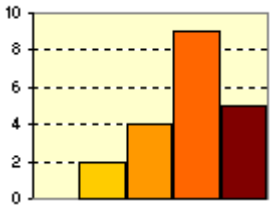
1. Sol 72º
2. Sol 3
3. Sol 51
4. Sol 300
5. Sol 2
6. Sol 2.65
7. Sol 1.06
8. Sol 212.25
9. Sol 24.53
10. Sol 240

No olvides enviar las actividades al tutor ▶

ACTIVIDADES DE ESO

Nombre y apellidos del alumno:		Curso: 4º
Quincena nº: 11	Asignatura: Matemáticas B	
Fecha:	Profesor de la asignatura:	

1. Para los datos de cada apartado elige el gráfico mas adecuado a ellos.

1) Trucha, merluza, sardina, emperador, merluza, merluza, sardina, emperador, sardina y merluza.	2) 1 2 1 2 2 2 3 4 1 2 1 2 3 1 2 2 3 4 2 3	3) 28,1 25,2 28,7 25,7 26 26,4 23,7 22,5 27,5 25,4 27,1 25,9 24,2 28,1 26,3 25,5 24,3 21,8 25,5 24
a) 	b) 	c)

2. Determina la moda en la distribución 1, la mediana en la 2 y la media en la 3.
a) b) c)

3. Determina la desviación típica en la distribución 1.

4. Calcula los cuartiles 1º y 3º de la distribución 2.

cidead@mec.es

http://cidead.cnice.mec.es

C/ TORRELAGUNA, 58
28027 MADRID
TEL: 91 3778300
FAX: 91 3778314