

Fútbol.....	108
Baloncesto.....	99
Bádminton.....	81
Natación.....	63
Balónvolea.....	54
Tenis.....	45
Atletismo.....	36
Otros deportes.....	81

ESTADÍSTICA

Una de las más bellas leyendas de las mil y una noches cuenta que el califa Harún al Rasid salía disfrazado de mercader de su palacio para conocer la opinión de los habitantes de Bagdad. También el escritor estadounidense Mark Twain, muchos años después, hace del príncipe Eduardo un mendigo, que de este modo llegará a conocer cómo vivían y pensaban sus súbditos. Hoy, a poco del siglo XXI, los

métodos han cambiado de manera radical y en la actualidad la importancia de la estadística es enorme: los métodos estadísticos son fundamentales para estimar, planificar, predecir y decidir en problemas relacionados con la política, la sociología, la investigación, la industria, la economía y una larga lista de actividades.

UN POCO DE HISTORIA:

El estudio estadístico más antiguo que se conoce se realizó en China cuando el emperador Yao encargó la confección de un censo del imperio. Esto sucedía a finales del tercer milenio antes de Cristo. También se tienen noticias de que en el antiguo Egipto se realizaron estadísticas y trabajos censales de tipo agrícola. Otro censo famoso, según relataba el historiador Tácito, fue el que recogía las propiedades y los ejércitos del imperio de Roma en tiempos del emperador Octavio Augusto.

No obstante, hemos de esperar algún tiempo hasta que aparezcan los trabajos de John Graunt, un comerciante de mercería inglés, considerado como el precursor de la actual estadística. Graunt, entre los años 1604 y 1661, realizó un estudio sobre los nacimientos y defunciones de Londres y, a partir de los datos obtenidos, extrajo consecuencias formulando leyes demográficas y comportamientos sociológicos.

Treinta años más tarde publica el astrónomo Halley un estudio sobre las tasas de mortalidad, sentándose las bases de los estudios sobre esperanza de vida.

PRESENTACIÓN DE DATOS EN TABLAS

En un Instituto el Departamento de Orientación realiza una encuesta entre los estudiantes al comienzo del curso.

A instancias del seminario de E. F. se le preguntó a los alumnos que indicaran el deporte que más practicaban, obteniéndose los siguientes resultados :



Normas, no necesariamente estrictas, para realizar esta operación:

- Es costumbre tomar los intervalos con la misma amplitud.
- La amplitud suele ser múltiplo de los números 2,3,5,10 y 20.
- El número de intervalos no ha de ser menor que 5 ni mayor que 15 (estos límites varían según los autores).
- No debe haber muchos intervalos con pocos individuos, ni con demasiados (relativamente hablando).

Discute en grupo cuál de las tres tablas te parece la más indicada.

Observarás que si damos a conocer los datos en una tabla, por ejemplo la 1, sólo informaremos de que hay 24 individuos que leyeron entre 9 y 11 libros anuales, pero se desconocerá el número de libros que, exactamente, leyó cada uno de ellos. Es decir, al agrupar se pierde información, aunque la lectura fácil de la tabla, su comprensión y la manipulación más cómoda de los datos, hace que dicha pérdida sea poco significativa.

1. Los datos de la tabla adjunta corresponden a la audiencia de un informativo nocturno de televisión durante los 30 primeros días del verano. (En

978	650	831	506	940	788	632	975	1025	890	652	590	544
789	870	986	754	822	655	779	621	592	835	1034	1043	
976	723	785	691	833								

miles)

Agrupar los datos en intervalos.

TERMINOLOGÍA ESTADÍSTICA

La estadística no siempre estudia a colectivos humanos. Puede ser interesante realizar trabajos estadísticos acerca de la producción de las fincas de una determinada comarca, de la resistencia de un tipo de vigas para la construcción o de la vida media de diferentes modelos de vehículos.

Por tradición se llama **población** al conjunto de personas, o de cosas, de que es objeto una estadística.

Hay que tener presente que una misma población puede ser sometida a distintas estadísticas: del conjunto de vehículos vendidos anualmente por un concesionario de una marca nos pueden interesar varias características como el color, los años de financiación o la edad del comprador.

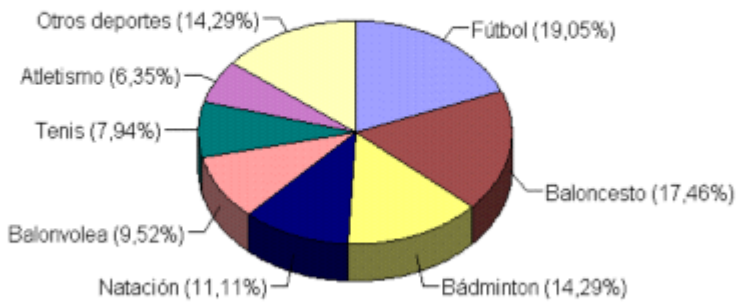
A estas características las llamaremos **caracteres o variables estadísticas**.

Los años de financiación y la edad del comprador se describen mediante números, diremos que son **variables cuantitativas**.

El color es una cualidad que expresa sus modalidades mediante palabras como verde, rojo..., diremos que se trata de una **variable cualitativa**.

Conviene no confundir una variable cuantitativa (X) con la frecuencia (N) con que se dan sus valores o modalidades.

DEPORTE MÁS PRACTICADO



2. En las siguientes estadísticas determina la población y la variable estadística. Si la variable es cuantitativa, indica si conviene agrupar sus valores en intervalos:

Número de horas perdidas por huelgas en las empresas de más de 50 trabajadores.

Número de tarjetas amarillas sacadas en cada partido de una jornada futbolística.

Estudio del turismo de un año por nacionalidades.

Número de hermanos de los estudiantes de una clase.

Velocidad de saque de un jugador de tenis.

3. Describe varias variables, de diferentes tipos, que se puedan estudiar en la población compuesta por la viviendas de una ciudad.

REPRESENTACIÓN GRÁFICA DE DATOS ESTADÍSTICOS

Las gráficas tienen la ventaja de que dan una información visual que resulta más cómoda que la lectura de una tabla. Veamos a continuación cómo se realizan y las características que presentan.

Diagramas de sectores:

Se trata de asignar a cada modalidad un sector circular que sea directamente proporcional a su frecuencia.

Si 108 alumnos prefieren el Fútbol, de un total de 567, el sector asignado a este deporte vendrá dado por la proporción

$$x/360 = 108/567; x = 69^\circ, \text{ aproximadamente,}$$

De manera análoga se calculan los demás sectores.

Este tipo de representación es muy utilizada en variables cualitativas.

Pictogramas:

Es otra forma habitual de representación que anticipamos en el tema de proporcionalidad. Se trata de dibujar un objeto característico de la variable, de manera que los tamaños sean directamente proporcionales a las frecuencias: en el dibujo A la altura es proporcional a la frecuencia; en el B lo es el área.

Gráfico B: Nº de médicos por 1000 habitantes

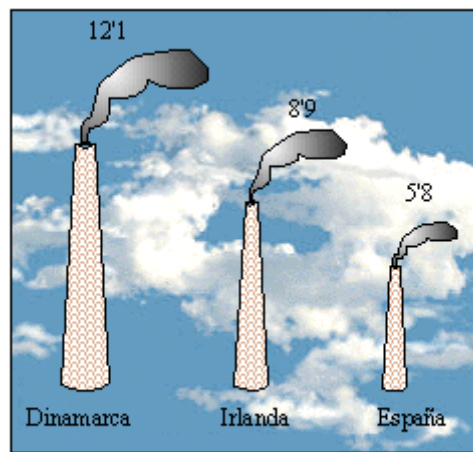
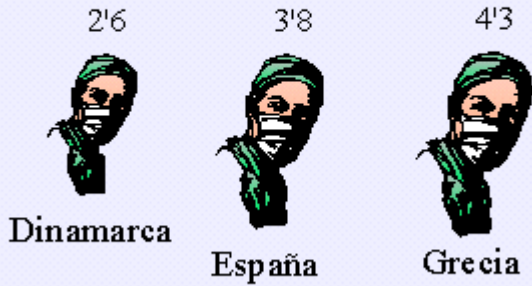
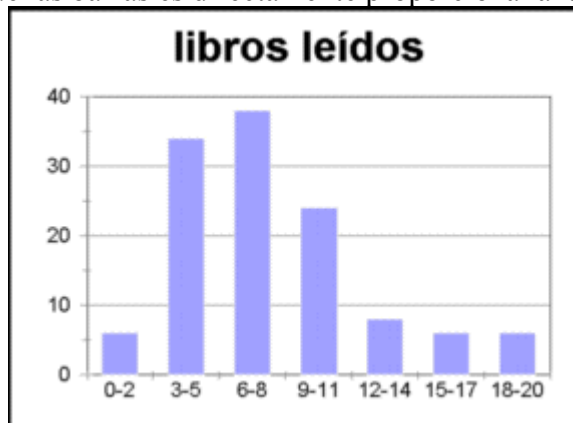


Gráfico A: Emisión de CO₂ per cápita

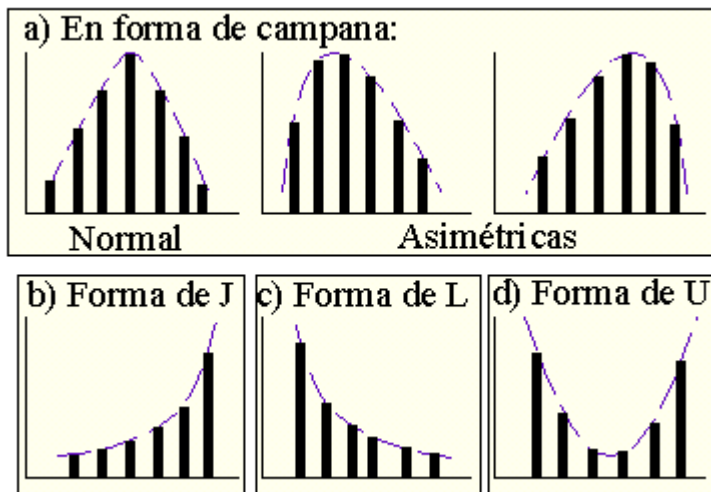
Diagramas de barras:

La altura de las barras es directamente proporcional a la



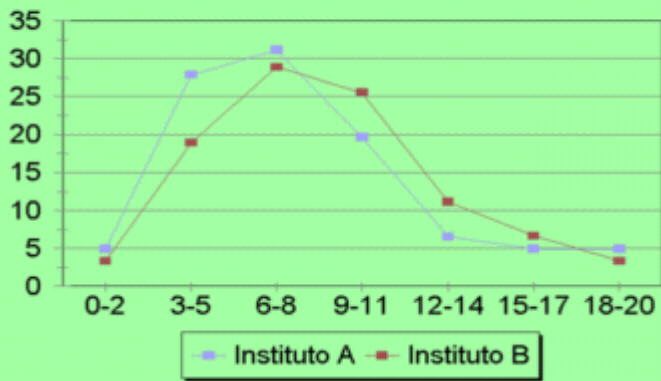
frecuencia.

Algunas distribuciones suelen clasificarse en virtud de la forma que adopta su representación por barras:



Libros leídos (en %)

Alumnos de 4º de los Institutos A y B



Son particularmente interesantes las distribuciones cuyo diagrama se asemeja a la campana normal, pues son numerosos los ejemplos en los que la mayoría de la población toma valores centrales, siendo pocos los valores extremos.

4. De las siguientes variables estadísticas, escribe los posibles valores y señala las que adoptan un tipo similar a los reseñados:

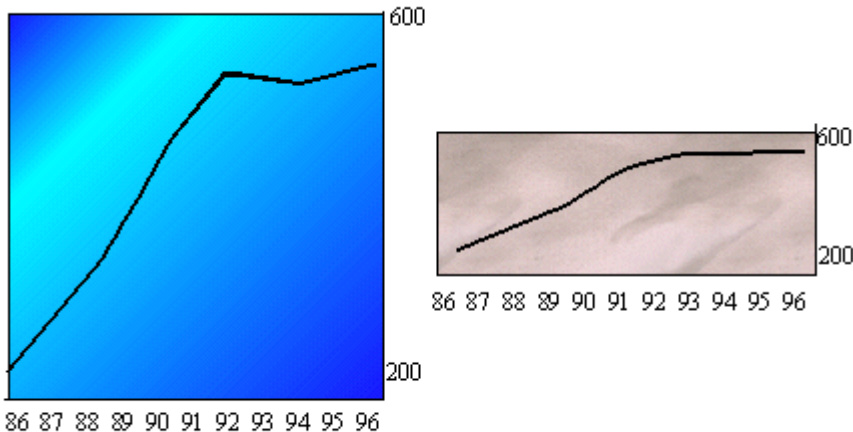
Número de suspensos.

Número de hijos por familia.

Tamaño del pie de los estudiantes de la clase.

Altura de las mujeres de 3º de E.S.O.

Duración de una pila o batería.



86 87 88 89 90 91 92 93 94 95 96

86 87 88 89 90 91 92 93 94 95 96

Evolución de los gastos en Investigación y Desarrollo en España (Datos del INE)

Polígonos de frecuencias:

Son similares en su fundamento a los diagramas de barras y suelen utilizarse para representar la evolución de un fenómeno a lo largo del tiempo o de una serie de etapas. También son muy cómodos para realizar comparaciones de dos poblaciones en un mismo gráfico:

Como el número de alumnos de 4º no era el mismo en cada Instituto, los hemos representado en tanto por ciento para poder compararlos.

¡Ojo con los gráficos!

En el capítulo de funciones comprobamos que tomando una escala inadecuada se puede desvirtuar la imagen del crecimiento real de un fenómeno. Algunas gráficas estadísticas también pueden ser desvirtuadas como ocurría con las funciones si la elección de la escala es exagerada o disminuida. Para evitar el problema algunas personas adoptan la norma de que la altura máxima sea el 69% de la anchura de la gráfica.

5. Las dos gráficas anteriores corresponden a un mismo fenómeno. Una la realizó el Gobierno y la otra la Oposición, ¿cuál corresponde a cada uno?

Diagramas de tronco:

159	180	186	136	181	145	183	188	138	189
181	180	162	181	148	177	154	148	186	183
149	159	190	127	156	193	182	143	138	188
159	179	150	161	182	177	152	178	190	183
178	146	162	154	154	145	186	148	178	184

Tronco (Decenas)	Hojas (Unidades)
12	7
13	6 8 8
14	3 5 5 6 8 8 8 9
15	0 2 4 4 4 6 9 9 9
16	1 2 2
17	7 7 8 8 8 9
18	0 0 1 1 1 1 2 2 3 3 3 4 6 6 6 8 8 9
19	0 0 3

En algunos casos resulta conveniente

Para ver en considerar

emplear este tipo de representación de datos. que consiste vamos a los siguientes ejemplos:

Tronco millones	Hojas centenas de millón
3	8 9
4	1 3 4 8
5	2 2
6	2 4 6
7	
8	3 8
9	3
10	
11	5 7

A un tenista se le mide la velocidad de los primeros 50 saques válidos en un partido de competición. Las velocidades se muestran en la tabla siguiente (En Km/h):

Una primera inspección visual muestra que los datos varían entre ciento veinte y ciento noventa aproximadamente. Como es lógico, las velocidades menores corresponden a un segundo servicio. Si miramos con más detalle observamos que el saque menos veloz fue de 127 Km/h y el más rápido de 193 Km/h.

El diagrama de tronco presenta los datos anteriores de manera más clara y legible y se realiza como se muestra a continuación:

El doce de la columna tronco señala los saques realizados a una velocidad mayor o igual que 120 y menor que 130. Mirando la columna correspondiente a hojas, vemos que sólo se sacó una vez en esa banda y que la velocidad fue de 127 Km/h.

La siguiente fila indica que se realizaron tres saques con velocidades mayores o iguales a 130 y menores que 140. En concreto fueron un saque a 136 Km/h y dos a 138 Km/h.

El resto de la tabla se interpreta de igual manera.

Estos diagramas tienen la ventaja de que permiten conocer con detalle los valores registrados y presentan, además, el aspecto de un diagrama de barras.

A veces no interesa reseñar en el diagrama la totalidad de las cifras, sólo se apuntan las más significativas con el objeto de hacer más legible el diagrama:

Los dieciséis partidos más vistos de la final del campeonato del mundo de Fútbol de 1994 tuvieron los siguientes espectadores (en miles):

6685, 5203, 4481, 8311, 3975, 4842, 11599, 6251, 4336,

3830, 4119, 5276, 11708, 8889, 9345, 6468.

Adoptaremos la norma de no escribir más de una cifra en la columna de hojas para hacer cómoda su lectura.

MEDIDAS DE POSICIÓN CENTRAL

La media

177, 180, 180, 174, 183, 171, 180, 181, 185, 184, 180, 181, 172, 182, 180, 182, 179, 181, 180, 183, 181, 179, 182 y 184.

En un circuito se mide la rapidez con la que un determinado piloto negocia una de sus curvas. Después de contabilizar su paso las primeras 24 veces, se obtienen las siguientes velocidades (En Km/h):

Velocidad	Nº de pases	X·n
X	n	X·n
171	1	171
172	1	172
174	1	174
177	1	177
179	2	358
180	6	1080
181	4	724
182	3	546
183	2	366
184	2	368
185	1	185
TOTAL	24	4321

X	n	c	cn
171-173	2	172	334
174- 176	1	175	175
177- 179	3	178	534
180- 182	13	181	2353
183- 185	5	184	920
TOTAL	24		4328

La media o promedio, que notaremos por \bar{x} , se calcula sumando todos los valores obtenidos y dividiendo esta cantidad entre el número total N de valores contabilizados:

$$S = 177 + 180 + 180 + 174 + \dots + 182 + 184 = 4321.$$

$$\bar{x} = S/N = 4321 / 24 = 180'04$$

La mayoría de las veces se presentan los datos en una tabla:

Esto supone una ventaja para su cálculo pues no será necesario sumar un valor cada vez que aparezca: se multiplican los valores por sus frecuencias (columna x*n) y al sumar los resultados obtenemos la suma total S.

La media es la suma de la columna x·n, dividida por el total:

$$\bar{x} = \frac{\sum x \cdot n}{N} ; \bar{x} = 4321/24$$

La letra griega Σ (*sigma* mayúscula) indica la suma de los valores x*n; N representa el total de veces que medimos su paso por la curva.

Imaginemos que la información anterior nos la hubieran presentado agrupando los datos por intervalos:

Como decíamos al principio, si nos dan los datos en intervalos estamos perdiendo información: sólo sabremos que realizó dos pases con una velocidad comprendida entre 171 y 173, pero no conoceremos la velocidad real a la que lo hizo. Lo que haremos en estos casos, es suponer que dicha velocidad fue de 172, que es el centro o marca de clase del intervalo. La misma operación se realiza con el resto de intervalos.

Si la elección de intervalos ha sido buena, la media estimada tendrá una diferencia insignificante con

la media real. En este caso,

$$\bar{x} = \frac{\sum c \cdot n}{N} = 4328/24 = 180'33$$

Concretando:

La media es la suma de todos los valores observados, dividida por el número total de estos

Dicho parámetro resume de alguna manera el comportamiento de la población puesto que cuanto más altos (bajos) son los valores observados, más alta (baja) es la media.

Otras características de la media:

a) Es fácil de calcular.

b) A partir de ella, conociendo el tamaño de la población, se puede calcular el total:

6. En un partido de Baloncesto entre los equipos de 3º y 4º de E.S.O., el primero empleó 10 jugadoras y el otro a 8. Por término medio, cada jugadora de 3º anotó 6'3 puntos mientras que la media de puntos que consiguió 4º fue de 7'5. ¿Cómo terminó el partido?

c) Cualquier variación en un dato influye en la media, es decir, la media depende de todas las observaciones y utiliza toda la información recogida. Lo anterior suele ser una ventaja, aunque algunas veces se presenta el inconveniente de que un valor raro, muy grande o pequeño, influye mucho en ella y desvirtúa el comportamiento general:

7. Un grupo de amigos disponen para sus gastos mensuales de 36, 48, 60, 63, 30 y 270 euros respectivamente. Calcula la media de dichas asignaciones. ¿Da una información significativa?

d) Cuando los valores de una distribución están concentrados en los extremos, la media no da, por sí sola, una imagen real del comportamiento de la población.

Así puede suceder con este peculiar equipo de Baloncesto de 3º de E.S.O., cuyas notas de matemáticas fueron:

0,10,10,0,10,0,0,0,10 y 10.

La media, 5, por sí sola, no informa de los comportamientos radicales de estos jugadores.

e) Dos poblaciones con igual media no tienen por que tener un comportamiento similar:

8. La ciudad de Shangai tiene una temperatura media anual de 15'1º y la de Montevideo de 16º. ¿Se pueden considerar parecidos dichos climas?. ¿Viajarías en el mes de Julio a los dos lugares con el mismo tipo de ropa?.

f) La media no tiene por que ser un valor observable:

9. El número de hijos de 10 familias son 0, 2, 1, 0, 2, 2, 1, 3, 4 y 2. Cada familia tiene, por término medio, 1'7 hijos. Lo anterior es, literalmente, imposible. Hay que saber interpretar dicho resultado en el sentido de que entre las 10 familias cuentan 17 hijos.

Medias compensadas

En un Instituto, el curso de 2º de Bachillerato se ha dividido en dos grupos : A y B.

2	.1
6	.4 .8 .8
7	.4 .4
8	.2 .4
9	.2 .6 .6
10	.4 .8 .8
11	.2 .4
12	.2 .6 .6 .6 .8
13	.4 .4 // .4 .6 .8 .8 .8
14	.2 .4 .6 .6 .6
15	.4 .4 .6 .8
16	.2 .4 .4
17	.2 .6
18	.2 .8
19	.2
20	.0

X	n	N
160-164	3	3
165-169	4	7
170-174	2	9
175-179	9	18
180-184	8	26
185-189	4	30

En la clase de 2º A, con 30 alumnos, se ha obtenido una nota media en Lengua de 6'3. En 2º B, con 19 alumnos, la media fue 5'9. ¿Cuál es la nota media de la asignatura en el curso de 2º?

En total hay 49 alumnos, la media vendrá dada por la suma de todas sus notas, dividida entre 49.

Como la media de 2º A es 6'3, la suma de las notas correspondientes a ese curso es $6'3 \times 30 = 189$. De manera análoga, la suma de las notas de 2º B vale $5'9 \times 19 = 112'1$.

La suma total de notas es $189 + 112'1 = 301'1$

La nota media es $301'1 : 49 = 6'14$, aproximadamente.

Hemos visto que, en determinados casos, la media no resume muy bien el comportamiento de una población. A continuación vamos a ver otra forma de describirla mediante un número al que no le afecta la presencia de algún valor grande o pequeño.

La mediana

Al primer ejercicio de una oposición, para optar a seis puestos de trabajo en una empresa, se presentan cuarenta y seis aspirantes. El tribunal decide eliminar al 50% en el primer ejercicio para que su labor de selección resulte más cómoda en las siguientes pruebas.

Si las notas obtenidas, puntuadas de 0 a 20, fueron las que se indican en el recuadro, ¿a partir de qué puntuación se superaría el ejercicio?.

Si contamos los 23 opositores que tienen peor nota, vemos que el listón del 50% lo marca la nota 13'4. Como es obvio, el tribunal no podrá promocionar al 50% exacto y tendrá que decidir si aprueba a partir de 13'4 o de 13'6.

¿Se aprueba sacando una nota superior a la media?

Diremos que 13'4 es la mediana, que notaremos por Me, de la distribución de notas.

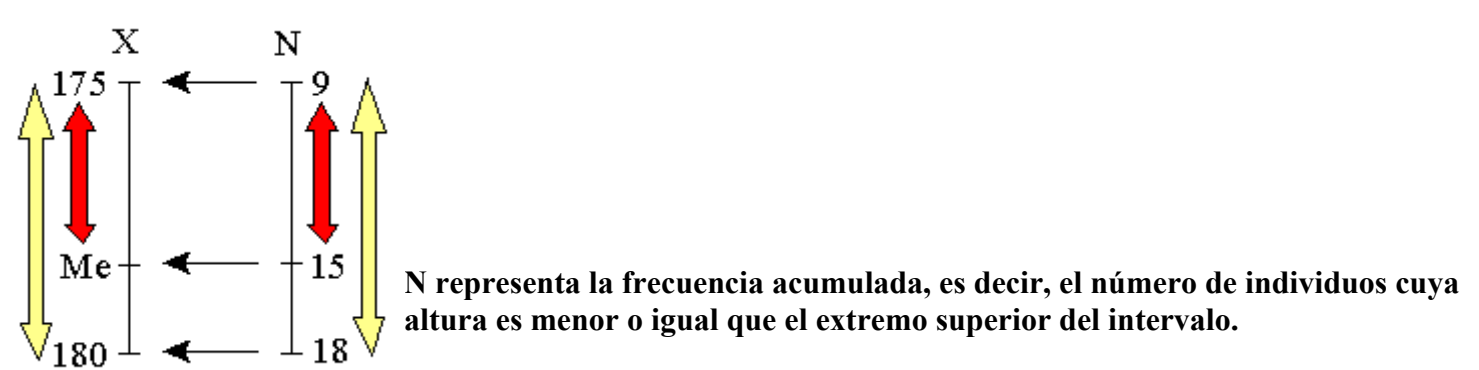
Si ordenamos de manera creciente los valores de una variable estadística, la mediana es el valor que divide a la población en dos mitades iguales

Consideremos a continuación la distribución de las tallas de 30 alumnos de un curso de 1º que, una vez ordenadas, son las que siguen: 160, 162, 162, 166, 167, 168, 168, 170, 171, 175, 175, 177, 177, 178, 178 // 179, 179, 179, 180, 181, 182, 182, 183, 183, 183, 184, 185, 185, 187, 189.

En este caso, el corte del 50% se da entre los valores 178 y 179. Diremos, cuando así ocurra, que la mediana es el valor intermedio entre ambos: 178'5: por debajo de esta talla encontramos a quince alumnos que corresponden al 50% de la población, y por encima de él al otro 50%.

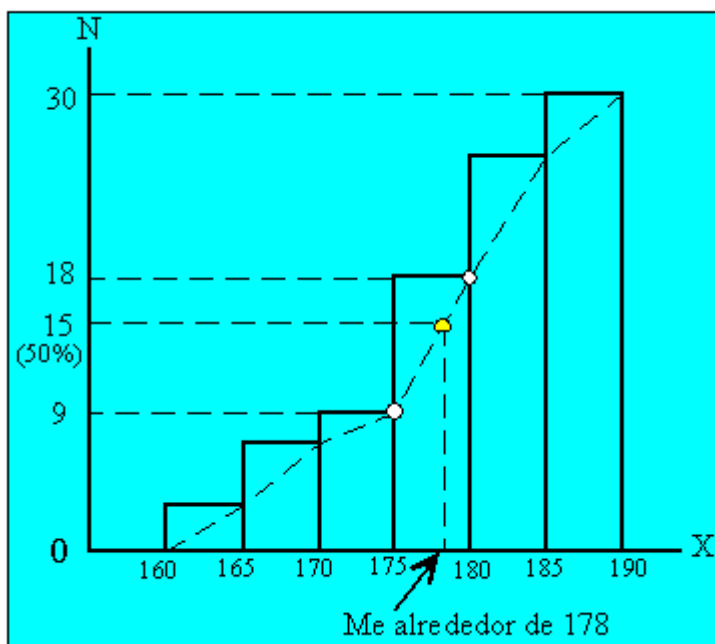
Estimación gráfica de la mediana:

Supongamos que los datos anteriores nos los hubiesen dado a conocer en una tabla:



Realizamos el histograma correspondiente a la tabla anterior y procedemos como se sugiere en el gráfico:

Observa que la Me se obtiene mediante una estimación visual, pero la diferencia en cm con la mediana real es poco significativa. Se puede obtener mayor precisión si trabajamos con papel milimetrado.



Otra forma de estimar la mediana sin tener que dibujar el histograma es la siguiente:

a) Se localizan los intervalos cuya frecuencia acumulada es más próxima a 15 y se establece la siguiente proporción:

Debajo de 175 hay 9 individuos

Debajo de la Me hay 15 individuos

Debajo de 180 hay 18 individuos:

$$\frac{Me - 175}{180 - 175} = \frac{15 - 9}{18 - 9}$$

X	n	N
160-164	3	3
165-169	3	6
170-174	5	11
175-179	7	18
.....		
.....		

b) Despejando obtenemos una estimación de 178'3 aproximadamente.

En realidad, lo que hemos hecho ha sido interpolar: hallar el valor de x correspondiente a $y = 15$, de la recta que pasa por los puntos (175,9) y (180,18)

10. En otro curso de 1º, el profesor, después de recoger los datos, escribió la tabla en la pizarra dejándola incompleta porque tocó el timbre de salida.

Los datos escritos fueron:

¿Se puede calcular la altura mediana sabiendo que en la clase hay 28 personas?

Características de la mediana:

a) No es necesario para su cálculo conocer el valor de todas las observaciones. No se utiliza toda la información recogida.

b) No se calcula mediante una expresión matemática. (Conociendo las medianas de dos grupos no se puede determinar la conjunta).

c) Es independiente de los valores raros o extremos:

11. Un laboratorio está ensayando un medicamento y desea conocer el efecto que se causa al añadir cierta cantidad de una sustancia a un determinado compuesto. Debido a la imprecisión de la balanza, deciden realizar varias pesadas con la misma cantidad y dar por bueno el valor mediano (ya que una pesada mal hecha, que arrojará un valor extraño, influiría en la media pero no así en la mediana). Los resultados obtenidos fueron (en mg):

337, 330, 345, 340, 335, 321, 329, 360, 333, 346, 331, 375, 335, 332, 339 y 339.

Determina el peso asignado a la sustancia.

Cuartiles

Otro investigador utiliza el método consistente en ordenar las pesadas, despreciar el 25% de las más bajas y el 25% de las más altas, evitando de éste modo el problema de las pesadas mal hechas. Con el 50% central calcula la media.

12. Halla el peso que asignaría el último investigador.

Llamaremos **primer cuartil**, C_1 , al valor que deja por debajo al 25% por ciento de la población. Al número que deja por encima el 25% de la misma, le llamaremos **tercer cuartil** y lo escribiremos como C_3 . El cálculo gráfico de los cuartiles se realiza igual que el de la mediana con sólo cambiar el porcentaje de referencia.

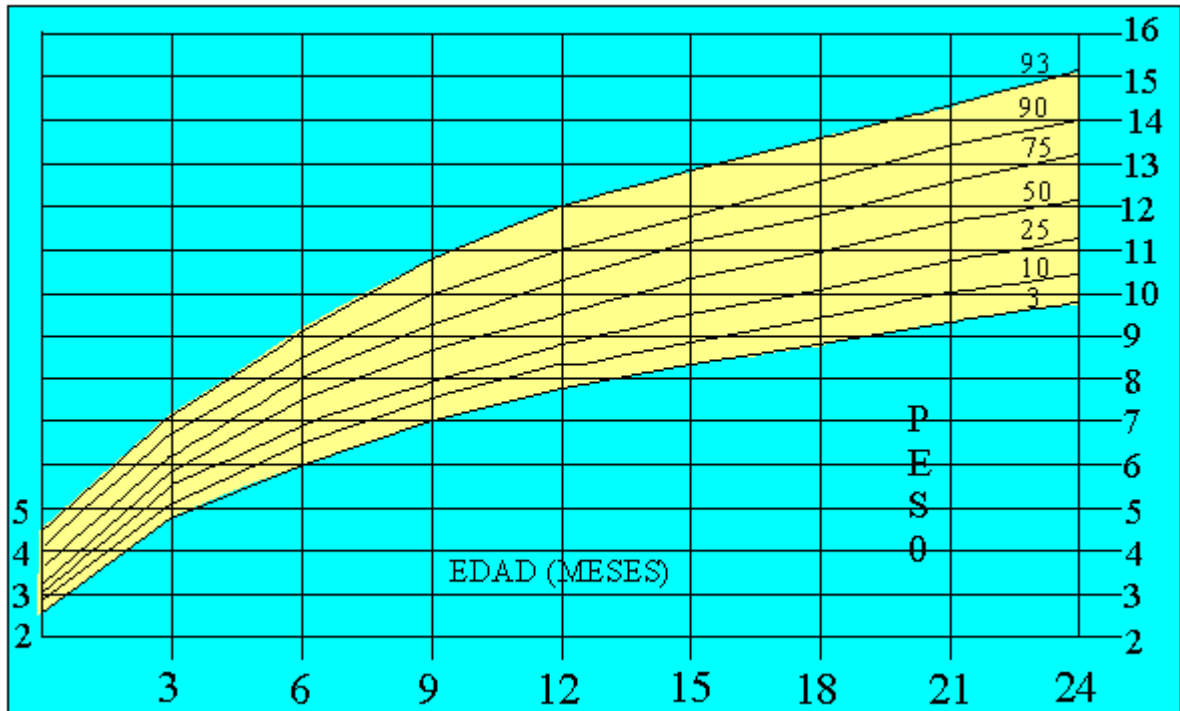
A la diferencia $C_3 - C_1$ le llamaremos **rango intercuartílico**.

13. Calcula los cuartiles en los ejemplos en que hallamos la mediana.

Percentiles

En algunas ocasiones el rango de valores que toma una variable es muy amplio y el hecho de que un valor se aleje de la mediana no supone ninguna anomalía. Otro tipo de medidas, como son los percentiles, están más indicados para comparar el comportamiento de un individuo frente al del colectivo.

En el siguiente gráfico tienes los pesos significativos de las niñas hasta los dos años de edad:



14. Señala los "límites de normalidad" de las niñas de un año, sabiendo que no se considera normal a quien se halle por debajo del 3% o por encima del 97%.

A dichos límites se les conoce como *percentil 3* y *percentil 97* respectivamente.

15. Calcula el percentil 10 (P_{10}) correspondiente al peso de los 14 meses.

16. Halla la altura mediana al año y seis meses.

La moda

Es el valor de la variable que tiene mayor frecuencia

En el ejemplo de los deportes más practicados la moda es el Fútbol, en el test de intuición espacial la moda sería obtener una puntuación de 6.

Cuando trabajemos con intervalos hablaremos de intervalo modal: en el ejemplo de los libros leídos, si nos dan la información mediante la tabla 1, la moda es leer entre 6 y 8 libros al año.

MEDIDAS DE DISPERSIÓN

Las marcas obtenidas por dos atletas de decatlon se dan en la siguiente tabla:

PRUEBAS	Atleta A	Atleta B
100 m	831	872
Salto de longitud	821	917
Lanzamiento de peso	812	703
Salto de altura	802	809
400 m	822	820
110 m vallas	820	881
Lanzamiento de disco	796	699
Salto de pértiga	817	843
Jabalina	804	705
1500 m	693	767

Es obvio que el atleta A es más regular que el B, que compensa las malas actuaciones en las pruebas de lanzamientos con su mayor velocidad. La cuestión que surge es cómo medir las irregularidades, o en otros términos, cómo expresar mediante un número la variabilidad o dispersión de unos datos.

La puntuación media obtenida por el atleta A fue de 801.8 puntos y la del B de 801.6

Observemos los datos siguientes:

Atleta A (Puntos)	Diferencia respecto a la media	Atleta B (Puntos)	Diferencia respecto a la media
831	29.2	872	70.4
821	19.2	917	115.4
812	10.2	703	-98.6
802	0.2	809	7.4
822	20.2	820	18.4
820	18.2	881	79.4
796	-5.8	699	-102.6
817	15.2	843	41.4
804	2.2	705	-96.6
693	-108.8	767	-34.6

Los números que señalan la variación de los datos son las diferencias respecto a la media: independientemente del signo, cuanto mayores sean más irregularidad indican.

Interesa resumir todas las variaciones en un sólo número. Si calculamos la media de las mismas, obtendremos que, al sumar, las que tienen signo + se compensan con las de signo - y la suma se anula perdiéndose toda la información (compruébalo).

Caben dos opciones:

a) Considerar positivas todas las diferencias, que llamaremos desviaciones, calculando a continuación el promedio de las mismas:

Desviaciones de A (con signo +)	Desviaciones de B (con signo +)
29.2	70.4
19.2	115.4
10.2	98.6
0.2	7.4
20.2	18.4
18.2	79.4
5.8	102.6
15.2	41.4
2.2	96.6
108.8	34.6
Promedio: 22.92	Promedio: 66.48

Al número así obtenido se le llama **Desviación media** y presenta los inconvenientes de la media.

b) Elevar al cuadrado todas las diferencias, calcular el promedio de los cuadrados y, para compensar de alguna manera, extraer la raíz cuadrada del valor obtenido. Obsérvese que al elevar al cuadrado potenciamos las desviaciones mayores, que proceden de los valores extremos, y damos menos relevancia a las pequeñas. La desviación media otorgaba la misma importancia a todas las desviaciones resaltando menos la presencia de valores alejados del promedio.

Desviaciones al cuadrado de A	Desviaciones al cuadrado de B
852.64	4956.16
368.64	13317.16
104.04	9721.96
0.04	54.76
408.04	338.56
331.24	6304.36
33.64	10526.76
231.04	1713.96
4.84	9331.56
11837.44	1197.16
Promedio: 1417.16	Promedio: 5746.24
Raíz del promedio: 37.64	Raíz del promedio: 75.8

Al valor promedio, así obtenido, se le llama Varianza de la distribución y se representa por V . A la raíz de la varianza se la llama Desviación típica y se nota como σ (la letra griega *sigma* en minúscula).

Veamos la forma de proceder para calcular la desviación típica cuando los datos vienen presentados en una tabla. Consideremos la estadística de los libros leídos durante el curso y añadamos a la tabla inicial los cálculos puestos en rojo:

x	n	c	$(c - \bar{x})^2$	$(c - \bar{x})^2 \cdot n$
0-2	6	1	47'33	284
3-5	34	4	15'05	511'7
6-8	38	7	0'77	29'26
9-11	24	10	4'49	107'76
12-14	8	13	26'21	209'68
15-17	6	16	65'93	395'58
18-20	6	19	123'65	741'9
Total	122			2279'88

Varianza: $2279'88/122=18.68$

Desviación típica: 4'32

Podemos resumir la fórmula de la varianza mediante la expresión:

$$V = \frac{\sum (x - \bar{x})^2}{N} = \frac{2279'88}{122} = 18'68$$

Se puede demostrar que la varianza también se calcula mediante esta otra

expresión:

$$V = \frac{\sum x^2 \cdot n}{N} - (\bar{x})^2$$

Es menos pesado utilizar la segunda fórmula, para ello la tabla se confecciona añadiendo la siguiente columna:

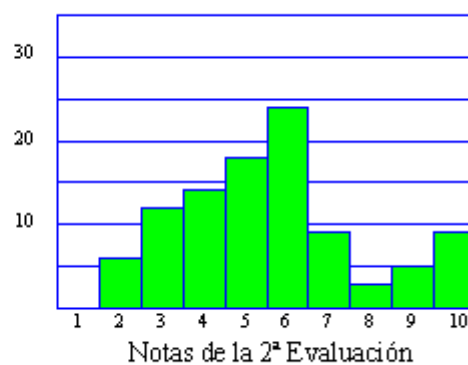
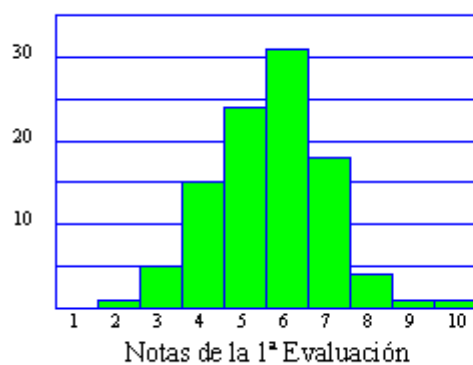
X	n	c	c^2	$c^2 \cdot n$
0-2	6	1	1	6
3-5	34	4	16	544
6-8	38	7	49	1862
9-11	24	10	100	2400
12-14	8	13	169	1352
15-17	6	16	256	1536
18-20	6	19	361	2166
Total	122			9866

Varianza: $=9866/122 - 62'09=19'76$

Desviación típica: 4'44

Observarás que los valores no coinciden con los anteriores. Este hecho se debe a que hemos truncado, en el primer caso, los números a dos decimales y se ha producido un pequeño error como consecuencia del truncamiento.

17. En la clase de 3º A la nota media otorgada por un profesor ha sido de 6'5 y la desviación típica de 2'8. El mismo profesor da clase a 3º B, y en este curso, la media fue de 6'2 y la desviación típica de 1'2. ¿Dónde es más probable que halla habido superior número suspensos? ¿y que halla más empollones?
18. En los gráficos de abajo tienes los resultados de los curso de 1º de Bachillerato, en determinada materia, de las dos primeras evaluaciones:







19. Contesta a las siguientes preguntas sin realizar cálculos:

¿Son muy diferentes las notas medias de las evaluaciones?

¿Dónde hubo mayor desviación típica?

20. En uno de los exámenes se pusieron cinco preguntas fáciles y otras cinco difíciles; en el otro se pusieron las diez preguntas normales. Razona qué examen se corresponde con cada evaluación.

Cálculos estadísticos con ayuda de la calculadora

Comenzaremos borrando los números de la memoria pulsando la tecla , a continuación ponemos la calculadora en modo estadístico pulsando las teclas  . Ahora sólo tenemos que escribir los datos en la calculadora pulsando la tecla  cada vez que se introduzca uno.

Una vez realizado lo anterior obtendremos los cálculos al pulsar la función estadística deseada:

Cálculo de la media. \bar{x} :   

Cálculo de la desviación típica. σ_n :

Suma de los datos. $\sum x$:

Suma de los cuadrados de los datos. $\sum x^2$:

Coeficiente de variación

Juan mide el perímetro del instituto, de 420 m de longitud, equivocándose en 2'2 m. Antonio mide el largo de la clase, de 830 cm, equivocándose en 5 cm. A la hora de comparar las habilidades midiendo, se ha de tener en cuenta el error cometido en relación al tamaño de lo medido. Para ello se utiliza el llamado *error relativo*: Juan se equivocó 2'2 m sobre 420 m y Antonio en 5 cm sobre 830 cm.

$$E_r(\text{Juan}) = \frac{2'2}{420} = 0'0523... \quad E_r(\text{Antonio}) = \frac{5}{830} = 0'0240...$$

Si multiplicamos el error relativo por cien tendremos, en %, el grado de equivocación de cada uno: Juan se equivocó en torno al 5'23% y Antonio en el 2'4%.

Algo parecido ocurre cuando queremos comparar la dispersión de los datos de dos poblaciones diferentes. Recordemos a nuestro atleta B que tenía una media de 801.6 puntos por prueba y una desviación típica de 75.8; estaba claro que su comportamiento era más irregular que el del atleta A, que presentaba una media similar con una desviación típica notablemente menor.

Deseamos comparar de nuevo la regularidad de B con la de otro atleta C de menor categoría, cuya media por prueba es de 750 puntos y su desviación típica de 60.3. Para esto utilizamos el coeficiente

de variación que se define como $CF = \frac{\sigma}{\bar{x}}$

En el caso de B vale $\frac{75.8}{801.6} = 0.0945...$ y en el de C $\frac{60.3}{750} = 0.0804...$ B presenta una irregularidad del 9.45% y C de 8.04%

21. Una empresa tiene dos factorías similares, una Estados Unidos y otra en Portugal. El salario medio en la americana es de 45300 \$ y la desviación típica de 10230 \$. En la de Portugal la media es de 38900 euros y la desviación de 5390 euros. Determina la respuesta adecuada:

a) Las diferencias entre sueldos en la factoría americana y las diferencias entre sueldos en la portuguesa son, proporcionalmente, similares.

b) Las diferencias en USA son casi el triple que en Portugal.

c) Las diferencias en USA son, prácticamente, el doble que en Portugal.

Teorema de Chebichev

Con anterioridad hablamos de la importancia de las distribuciones en forma de campana, en particular de las que no presentaban una asimetría destacable: las distribuciones normales, que suelen ser las más frecuentes. La desviación típica tiene el inconveniente de que es menos intuitiva que la desviación media, a pesar de ello es más utilizada. Existen algunas razones que lo justifican, este teorema es una de ellas porque permite, sabiendo la media y la desviación típica, conocer a cierta escala el comportamiento de sus individuos:

Las estadísticas relativas al número de libros leídos y la prueba de razonamiento espacial se distribuían en forma normal.

En la primera, la media vale 7'885 y la desviación típica 4'324, si consideramos el intervalo $(\bar{x} - \sigma, \bar{x} + \sigma) = (3'561, 12'209)$, observamos que contiene a 79 alumnos, el 66%, cuya lectura de libros se halla entre los dos valores.

En el segundo caso, la media es 5'591 y la desviación típica 2'014. $(\bar{x} - \sigma, \bar{x} + \sigma) = (3'577, 7'606)$. El número de alumnos cuya capacidad de razonamiento espacial se encuentra entre los valores del intervalo es 81, aproximadamente el 67%.

El hecho de que ambos porcentajes sean parecidos no es casualidad, tal y como afirma el conocido Teorema de Chebichev:

En una distribución normal:

Aproximadamente el 68% de la población está dentro de los valores del intervalo

$(\bar{x} - \sigma, \bar{x} + \sigma)$

Aproximadamente el 95% se halla en $(\bar{x} - 2\sigma, \bar{x} + 2\sigma)$

Aproximadamente el 99% se halla en $(\bar{x} - 3\sigma, \bar{x} + 3\sigma)$

22. Se realizó una estadística sobre el número de vehículos que utilizaban semanalmente el trasbordador de un río. La media de vehículos fue de 375 y la desviación típica de 25.
23. El propietario estima que debe comprar un nuevo trasbordador si el número de semanas que sobrepasan los 400 vehículos es mayor del 40%. ¿Qué consejo le darías al empresario?

MUESTREO

Hasta ahora hemos trabajado con poblaciones pequeñas y en circunstancias en las que hemos podido recabar los datos correspondientes a la totalidad de los individuos. Estas condiciones tan favorables no se dan en todas las ocasiones:

a) Un periódico encarga un trabajo sobre la valoración que hacen las personas en edad electoral de los principales líderes políticos. La población que comprende la estadística es de enorme tamaño y ello hace imposible, por razones de tiempo, personal y costo, que se pueda realizar una encuesta a cada uno de los electores.

12385	03138	86105	27259
37224	12537	08201	06951
91728	36791	77466	44853
16263	32950	52861	70519
87434	28729	45751	96471
14106	14351	67392	49181
77049	70866	95319	33507
99837	24378	12446	71564
50982	57332	09240	20506
72405	74957	35921	61472
68392	28193	17436	83810
54071	96628	60918	15629
28519	82160	25237	01347
41962	61312	50493	48612
58207	58261	11150	35749
29993	90682	49275	50237
05172	93144	73922	52185

b) Unos laboratorios ensayan una nueva vacuna. El director de producto desea realizar un estudio con las primeras unidades elaboradas, para comprobar a que temperatura de almacenamiento comienzan a perder sus propiedades. En este caso, cada prueba realizada supone la destrucción del producto, con su correspondiente coste.

En situaciones como las anteriores se selecciona una parte de la población, que sea representativa de la misma, a la que llamaremos MUESTRA, y se realiza el estudio sobre ella, generalizando los resultados a la totalidad de la población.

La importancia que tiene el tomar una muestra adecuada se pone de manifiesto en este conocido caso:

En 1936, con motivo de las elecciones presidenciales americanas, la revista Literary Digest realizó entre sus lectores 2500000 encuestas y de ellas se desprendía la victoria, por clara mayoría, del candidato Landon. Por otro lado, Gallup, Crossley y Roper, realizando tan sólo 5000 entrevistas, acertaban al pronosticar la reelección de Roosvelt.

24. Intenta explicar por qué sucedió.
25. Tres estudiantes tienen que realizar un trabajo de Sociales para averiguar cómo emplea el tiempo libre la gente de su pueblo. Como suelen salir juntos los sábados por la tarde, deciden realizar una encuesta en sus lugares de reunión.
26. Al producirse en clase la presentación y el debate de su trabajo, comenzaron afirmando que el 85% de las personas del pueblo se lo pasan pipa bailando los fines de semana.

¿Te parece correcta la afirmación?

La pregunta inmediata es: ¿Cómo elegir una muestra adecuada sin caer en errores subjetivos o de otro tipo?

La respuesta no es sencilla, pues depende de cada caso y se suelen utilizar técnicas que rebasan los conocimientos de este curso. No obstante, vamos a comentar dos formas de realizar un muestreo:

Muestra aleatoria

La tabla adjunta es una tabla de números aleatorios obtenida por ordenador. Veamos cómo utilizarla para obtener una muestra de los propietarios de una urbanización de 15 bloques, con cinco plantas por bloque y cuatro pisos por planta.

En total hay 300 viviendas que podemos numerar de 0 a 299, comenzando por la letra A de la primera planta del primer bloque y terminando por la letra D de la quinta planta del bloque nº 15.

0. ¿A qué piso corresponden los números 130 y 249?

Si queremos seleccionar 40 propietarios al azar, procederemos como sigue:

Si el primer número comienza por cero, uno o dos, los tres primeros dígitos del número indicarán el propietario elegido.

Si no ocurre así, miramos si el segundo dígito cumple esa condición y en ese caso se elige el piso que corresponde a los dígitos segundo, tercero y cuarto. Si el segundo dígito fuese mayor que dos, miramos el tercero y vemos si la cumple, siendo en tal caso los tres últimos dígitos los que señalan el propietario. De fallar el tercero, se pasa al número siguiente.

También se puede hacer comenzando por la primera columna y señalando los primeros 40 números menores que 300.

Los propietarios elegidos se ordenan y se comprueba si alguno está repetido.

1. Realiza una muestra aleatoria de estudiantes del Instituto, utilizando la tabla anterior.

Muestras estratificadas

En sondeos de carácter político, la adhesión a un determinado partido depende en cierta manera de la categoría social y cultural en la que se enmarca cada individuo. Los censos recogen datos como la profesión, nivel de estudios etc, que permiten establecer la proporción, dentro de determinada población, de tales categorías. La muestra ha de ser tal que cada categoría mantenga dentro de ella el mismo porcentaje que en el total de la población.

2. Trabajo en grupo: Realizar una encuesta, sobre un tema que consideréis interesante, entre los estudiantes del centro.

Comparar los cuestionarios elaborados y los resultados obtenidos por los distintos grupos.