

Objetivos

En esta quincena aprenderás a:

- Recordar los conceptos de población, muestra, individuo y carácter.
- Valorar la importancia del concepto de variable estadística y distinguir entre los diferentes tipos.
- Resumir mediante una tabla de frecuencias cualquier serie de datos.
- Asociar e interpretar gráficos estadísticos valorando su utilización en diferentes áreas de conocimiento.
- Calcular, valorar e interpretar la media, mediana y moda en variable discreta.

Antes de empezar

1. Vocabulario estadístico pág. 4
Población, muestra, individuo y carácter
2. Carácter. Variable estadística pág. 6
Carácter cualitativo. Atributos
Variables discretas
Variables continuas
3. Ordenación de datos. Tabulación. pág. 10
Para variable discreta
Para variable cualitativa
4. Gráficos para variable cualitativa. pág. 12
Diagrama de barras
Diagrama de sectores
5. Gráficos para variable discreta pág. 14
Diagrama de barras
Polígonos de frecuencias
Diagrama de sectores
6. Medidas de centralización pág. 17
Media
Mediana
Moda
6. Medidas de dispersión pág. 20
Rango
Desviación media

RESUMEN

Autoevaluación

Antes de empezar

La Estadística ha penetrado en múltiples aspectos de la vida cotidiana haciendo familiares términos como población, muestra, media, mediana, moda...

Puede asegurarse que cualquier persona informada de hoy en día posee un vocabulario básico de estadística, lo entiende, lo utiliza y valora.

Prácticamente todas las ciencias, tanto científico tecnológicas como sociales utilizan en aspectos fundamentales de las mismas a la estadística.

El deporte no es una excepción. En todos ellos y en particular en el baloncesto el manejo de los datos estadísticos constituye un aspecto a estudiar y manejar tan importante a veces como las tácticas y la técnica implícitas del propio juego.

En el ejemplo siguiente simula un saque de fondo en baloncesto, se representa con puntos rojos los jugadores atacantes y los verdes como los defensores.

El estudio que realizan los cuerpos técnicos de los equipos se encarga de calcular que estadística de tiro tiene cada jugador, de esta manera si se deja desmarcado al jugador que tenga peor estadística; el balón irá hacia él.



1. Vocabulario estadístico

Población, muestra, individuo y carácter.

Las primeras definiciones necesarias para el inicio de cualquier estudio estadístico son población, individuo, muestra y carácter.

Comencemos con un ejemplo que nos haga intuir dichos conceptos.

Ejemplo

Estudio sobre la posible existencia de vida en otras estrellas.

¿Existen sistemas planetarios semejantes al nuestro que, quizás puedan albergar algún tipo de vida?

Hasta hace muy poco tiempo, los astrónomos no tenían pruebas de la existencia de planetas fuera del Sistema Solar.

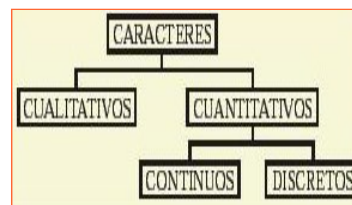
En la actualidad se han descubierto algunos centenares de planetas gigantes, que difícilmente podrían encerrar vida, pero que sí serían una especie de señal de que en esa estrella puede existir un sistema con órbitas y tamaños mas acordes con las posibilidades de vida en el sentido que conocemos.

- La población está constituida por todas las estrellas del universo visible.
- La muestra está constituida por todas las estrellas escogidas y observadas en el proyecto.
- El individuo es cada estrella del universo observable.
- El carácter es la presencia o no de perturbaciones que indiquen la existencia de planetas gigantes.

Podemos definir los conceptos anteriores de la siguiente forma:

- **Población:** Conjunto de todos los elementos que verifican una característica que será objeto de estudio.
- **Individuo:** Cada uno de los elementos de la población.
- **Muestra:** Cualquier subconjunto de la población. Este subconjunto es muy importante que sea *representativo* de la población.
- **Carácter:** Cada una de las propiedades que poseen los individuos de la población y que pueden ser objeto de estudio.

La definición de carácter debe ir acompañada de la siguiente clasificación:



Recordemos entonces que ante cualquier estudio estadístico debemos tener en cuenta la identificación de los elementos, de esta forma evitaremos errores en las conclusiones finales.

Identificación de elementos

- Población
- Muestra
- Individuo
- Carácter



Ejemplo

Estudio sobre la evolución de la talla en la juventud española.

Los españoles igualan la estatura a la mayoría de los europeos, pero evolucionan hacia la obesidad norteamericana.

Un estudio antropométrico conjunto entre varios hospitales españoles, revela que la estatura de los españoles se ha igualado en los últimos treinta años respecto a la mayoría de los países europeos.

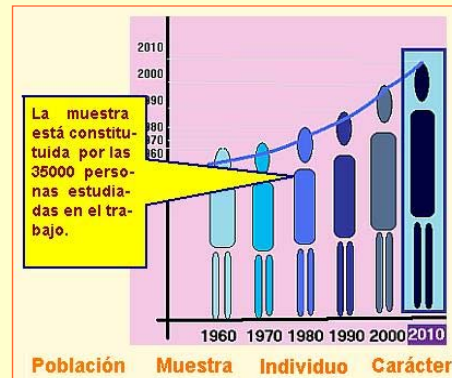
El mismo estudio también alerta sobre la preocupante tendencia hacia la obesidad en niveles similares a la población norteamericana.

El trabajo, llevado a cabo mediante la medición de casi 35000 sujetos entre los años 2000 y 2004, también demostró que las diferencias entre las distintas comunidades autónomas dentro de España son casi inexistentes.

Población



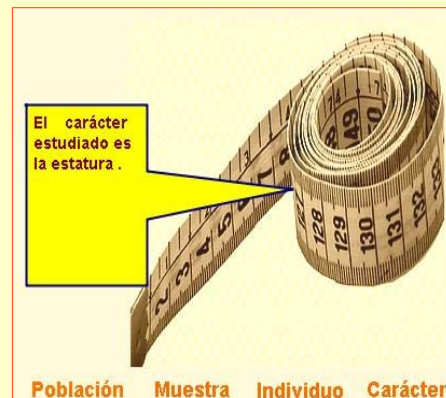
Muestra



Individuo



Carácter



2. Carácter. Variable estadística.

Trabajo de campo.

El trabajo de campo es la etapa de la investigación en la que se establece contacto directo con la población o muestra para recabar los datos que se necesiten.

La planificación es fundamental y su desarrollo depende del método de obtención de la información que se utilice.

El empleo del ordenador permite una simulación de situaciones que hace que realicemos un trabajo de campo virtual sin desgaste físico.

Carácter cualitativo. Atributos.

Comencemos nuevamente con un ejemplo que nos ilustre.

Ejemplo

Afición al fútbol.

Preguntamos a una serie de personas sobre sus preferencias en cuanto a afición futbolística.

La muestra que consideramos será de 9 personas de distintas ciudades españolas.

Los datos son F.C. Barcelona, Sevilla C.F., At. Bilbao, R. Madrid, R. Madrid, At. Madrid, Valencia C.F., F.C. Barcelona y Deportivo de la Coruña.

Las características de estos valores son:

- No son medibles con números.
- No tiene sentido la ordenación.
- Las distintos valores se identifican con el nombre del equipo elegido.

Todos los individuos de la población que vamos a estudiar tienen una serie de propiedades o cualidades que en estadística reciben el nombre de caracteres.

Los caracteres pueden ser de dos grandes tipos:

- a) CUALITATIVOS
- b) CUANTITATIVOS

Un carácter cualitativo se caracteriza porque sus diferentes modalidades no pueden expresarse con números.

Ejemplo

Tu color preferido.

Preguntamos a una serie de personas sobre sus preferencias en cuanto a colores.

En este caso la simulación de la población y color elegido se puede realizar mediante el ordenador, existen programas que permiten generar muestras aleatorias que simulan el trabajo de campo.

La muestra sobre la que actuamos será de 10 personas de una ciudad cualquiera.

DATOS OBTENIDOS	
blanco	
blanco	
verde	
amarillo	
verde	
verde	
naranja	
amarillo	
rojo	
rojo	

CARACTERÍSTICAS:

- Los valores que toma no son medibles numéricamente
- No tiene sentido la ordenación
- No tiene sentido hablar de valores CONSECUTIVOS
- Las distintas modalidades del carácter no representables numéricamente, se identifican mediante el nombre del color elegido.

VARIABLES DISCRETAS. CARÁCTER CUANTITATIVO DISCRETO.

Se denomina así al carácter cuyas modalidades se pueden representar con números.

Dentro de los caracteres cuantitativos se distinguen dos tipos: Discreto y continuo.

Es discreto si toma valores aislados, de manera que entre dos consecutivos no existe otro intermedio.

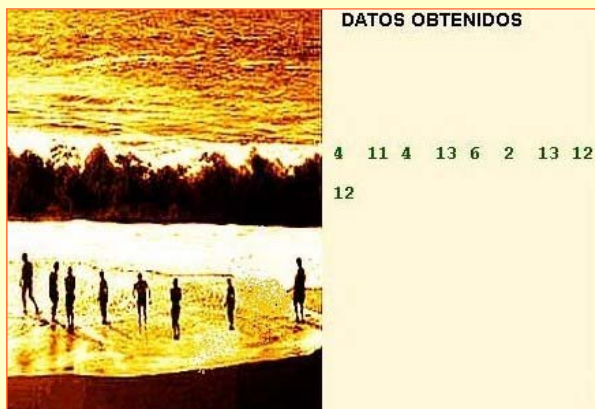
Ejemplo

¿Cuánta gente hay en la playa?

Realizamos una fotografía de una determinada zona de playa a distintas horas del día y anotamos las personas que aparecen en ella.

En este caso disponemos de un banco virtual de fotos y un procedimiento totalmente aleatorio que simula las distintas situaciones.

La muestra sobre la que actuamos es de 9 fotografías.



CARACTERÍSTICAS:

- Los valores que toma son aislados; entre 1 y 13.
- Los valores se pueden ordenar y contar
- Entre dos valores CONSECUTIVOS no existen valores intermedios
- ESQUEMA

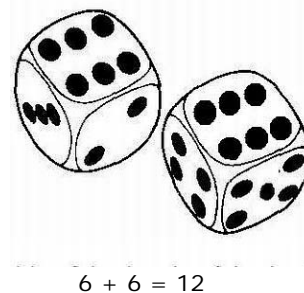
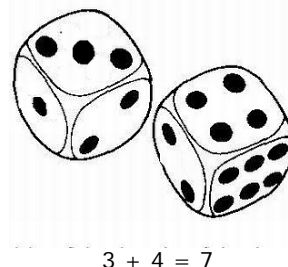


Ejemplo

¿Cuánto suman las caras superiores de dos dados previamente lanzados?

Lanzamos dos dados perfectos anotando la suma de los resultados de las caras superiores.

La muestra que consideramos será la suma de 8 pares de lanzamientos.



Los datos obtenidos son: 7, 6, 9, 2, 8, 1, 8 y 7.

Las características de estos valores son:

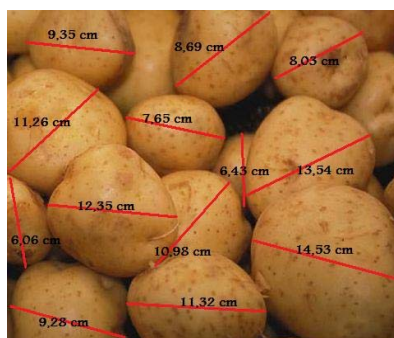
- Los valores que toman son aislados; entre 2 y 12.
- Los valores se pueden ordenar y contar.
- Entre dos valores consecutivos no existen valores intermedios.

VARIABLES CONTINUAS. CARÁCTER CUANTITATIVO CONTINUO.

Cuando las modalidades de un carácter cuantitativo pueden tomar valores de un conjunto de números reales o un intervalo, al menos teóricamente, se dice que estamos ante un carácter cuantitativo.

Ejemplo

Midiendo patatas. Diámetro medio.



Para una posterior clasificación de calidad se realiza un estudio sobre el diámetro medio en distintas producciones de patatas. La muestra que consideramos será de 8 producciones.

Todo el, en este caso, costoso procedimiento de recogida de las muestras lo sustituimos por una simulación por ordenador.

Los datos obtenidos son: 14,6 ; 6,7 ; 9,8 ; 13,2 ; 8,1 ; 9,3; 13,8 y 10,1.

Las características que se tienen sobre los valores son:

- Entre dos valores siempre existe la posibilidad de otro.
- No tiene sentido hablar de valores consecutivos.
- Toma valores dentro de un intervalo.

Ejemplo

Pesando recién nacidos.

Vamos a preguntar el peso de los recién nacidos en una determinada ciudad.

Existen programas informáticos que permiten generar muestras aleatorias que simulan el trabajo de campo.

La muestra sobre la que actuamos es de 40 bebés.

DATOS OBTENIDOS

2,28	3,22	3,95	2,43	3,59
2,47	3,41	3,50	2,07	2,13
3,55	2,61	3,21	2,29	4,07
2,33	2,51	3,38	3,67	4,10
3,84	2,31	2,63	4,03	2,95
3,86	4,03	3,02	2,99	4,00
3,72	2,98	3,88	3,07	3,28
3,84	3,37	3,55	4,07	2,94

CARACTERÍSTICAS:

- Entre dos valores siempre existe la posibilidad de otro
- No tiene sentido hablar de valores CONSECUTIVOS
- Toma valores dentro del intervalo [2,07 ,4,10]
- INTERVALO DE VALORES



Ejercicios resueltos

1. Clasifica las siguientes variables: cualitativas, discreta o continua, escribiendo una X en el recuadro correspondiente.

Nº de hijos varones, tipo de música preferida, nº de hijos, peso de recién nacidos, páginas de un libro, estatura.

SOLUCIÓN:

	CUALITATIVA	DISCRETA	CONTINUA
Nº de hijos varones		X	
Tipo de música preferida	X		
Nº de hijos		X	
Peso de recién nacidos			X
Páginas de un libro		X	
Estatura			X

2. Clasifica las siguientes variables: cualitativas, discreta o continua, escribiendo una X en el recuadro correspondiente.

Raza de perros. nº de hijos, longitud del pie, asignaturas pendientes, perímetro craneal, cantante favorito.

SOLUCIÓN:

	CUALITATIVA	DISCRETA	CONTINUA
Raza de perros	X		
Nº de hijos		X	
Longitud del pie			X
Asignaturas pendientes	X		
Perímetro craneal			X
Cantante favorito	X		

Observa:

La variable "asignaturas pendientes" hace referencia al nombre de las asignatura y por eso es variable cualitativa, mientras la variable "nº de asignaturas pendientes" sería discreta.

3. Ordenación de datos.

Al final podrás comprobar que...

PROPIEDADES INTERESANTES DE LAS TABLAS ESTADÍSTICAS

- La suma de todas las frecuencias absolutas es igual al tamaño de la población o de la muestra.
- La suma de las frecuencias relativas es siempre igual a 1.

Si se ha realizado algún redondeo en las frecuencias relativas es usual que la suma de las mismas no sea exactamente igual a uno debido a los errores cometidos..

Tabulación para variable discreta.

El paso siguiente al trabajo de campo es la disposición de los datos de manera ordenada, concisa y visualmente atractiva. En estadística, este proceso recibe el nombre de tabulación.

Los valores obtenidos se ordenan, especifican y agrupan de tal forma que sea fácil la información y búsqueda.

Las primeras columnas que deben aparecer serán:

- Valores de la variable, X_i .
- Frecuencias absolutas, f_i .
- Frecuencias relativas, h_i .
- Frecuencias absolutas, acumuladas, F_i .
- Frecuencias relativas acumuladas, H_i .

En algunos casos se puede utilizar el porcentaje en lugar de las frecuencias relativas o además de las frecuencias relativas.

Ejemplo

Edad de los estudiantes.

Las edades de 30 estudiantes de un instituto de enseñanza secundaria da los valores que posteriormente tabulamos como sigue:

16	16	16	13	14	13	13	13	16	16	13	14	12	17	15
16	15	17	17	15	13	17	12	16	14	16	16	16	14	14

X_i	f_i	h_i	F_i	H_i
12	2	$\frac{2}{30}$	2	$\frac{2}{30}$
13	6	$\frac{6}{30}$	8	$\frac{8}{30}$
14	5	$\frac{5}{30}$	13	$\frac{13}{30}$
15	3	$\frac{3}{30}$	16	$\frac{16}{30}$
16	10	$\frac{10}{30}$	26	$\frac{26}{30}$
17	4	$\frac{4}{30}$	30	$\frac{30}{30}$

Tabulación para variable cualitativa.

En los casos de carácter cualitativo, la tabulación de los datos es muy simple. Las tres columnas que tienen sentido hacen referencia a:

- El valor de los atributos.
- La frecuencia absoluta
- La frecuencia relativa

Ejemplo

La práctica de deporte.

Recogida de datos sobre deportes practicados, tabulada:

fútbol, tenis, balonmano, tenis, voleibol, atletismo, baloncesto, fútbol, fútbol, balonmano, fútbol, voleibol, balonmano, fútbol, balonmano, fútbol, fútbol, tenis, atletismo.

$X = \text{deporte}$	f	h
baloncesto	2	2/20
balonmano	4	4/20
voleibol	2	2/20
tenis	3	3/20
atletismo	2	2/20
fútbol	7	7/20
	20	

Ejercicios resueltos

3. Para un estudio de accesibilidad, durante 30 días anotamos el número de plazas libres de aparcamiento a las 5 de la tarde.

1 2 1 2 0 1 3 2 1 5 0 2 2 1 3
3 2 1 1 5 0 5 3 0 3 3 2 2 3 1

Realiza una tabulación de los datos en la que aparezcan las columnas correspondientes a las frecuencias absolutas, relativas, acumuladas absolutas y relativas.

SOLUCIÓN:

Nº de plazas de aparcamiento	f	h	F	H
0	4	0,14	4	0,14
1	8	0,28	12	0,41
2	8	0,28	20	0,69
3	7	0,24	27	0,93
4	0	0	27	0,93
5	3	0,1	30	1,03
	30			

4. Preguntamos a 20 estudiantes elegidos aleatoriamente por el tipo de música que prefieren escuchar.

Los resultados son: disco, rock, rock, clásica, rock, latina, pop, rock, latina, rock, flamenco, flamenco, flamenco, latina, rock, clásica, disco, disco, latina, rock.

Realiza una tabulación de los datos en la que aparezcan las columnas correspondientes a las frecuencias absolutas y relativas.

SOLUCIÓN:

Tipo de música	f	h
Disco	3	0,1
Rock	7	0,24
Latina	4	0,14
Clásica	2	0,07
Flamenco	3	0,1
Pop	1	0,03
	20	

4. Gráficos para una variable cualitativa.

Diagrama de barras.

El diagrama de barras es junto al de sectores el gráfico más utilizado para variable cualitativa.

Se utiliza como complemento a la tabla de frecuencias o incluso en sustitución de ésta.

En el eje de abscisas se sitúan a igual distancia los distintos atributos.

A partir de cada atributo se levantan barras de igual grosor y cuya altura sea la de la correspondiente frecuencia absoluta.

Diagrama de sectores.

El diagrama de sectores en variables cualitativas es uno de los recursos estadísticos más utilizados.

Es Especialmente útil en los casos en que existen pocas modalidades del carácter. Se suele utilizar junto a la tabla de frecuencias o sustituyendo a ésta.

Para calcular el ángulo del sector que corresponde a cada valor de la frecuencia:

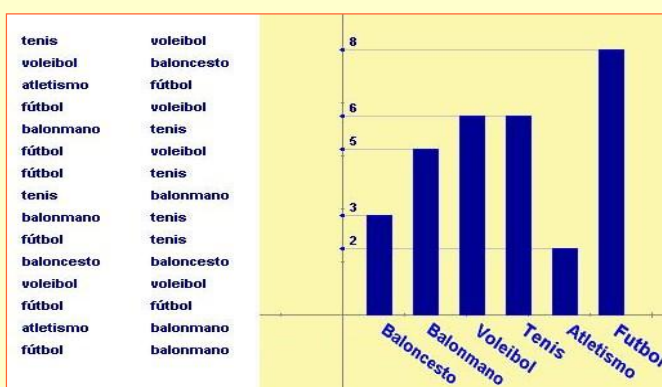
$$x = \frac{f_i \cdot 360}{\text{Total Datos}}$$

Ejemplos

Deporte practicados. (Diagrama de barras)

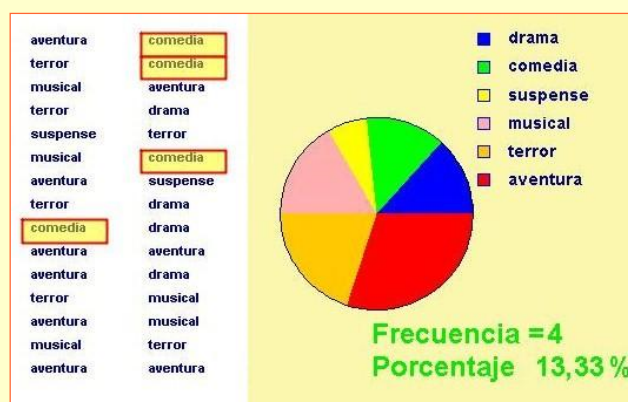
Los datos corresponden a las contestaciones realizadas por 30 estudiantes sobre el deporte que practicaban con mayor frecuencia en el instituto.

Si queremos tener una rápida visión de los datos, una forma de organizarlos es a través de una representación de diagrama de barras. En este ejemplo puedes ver la diferencia entre hacer un análisis sobre el listado o sobre la gráfica. ¿Cuál te resulta más fácil?.



Tipo de película. (Diagrama de sectores)

Hemos vuelto a preguntar a nuestros estudiantes sobre el tipo de película que les gusta ver. Otra forma de organizarlos de forma más fácil de ver es el diagrama de sectores. ¿Serías capaz de recordar alguna otro ejemplo? (Ayuda: ocurre cada cuatro años).



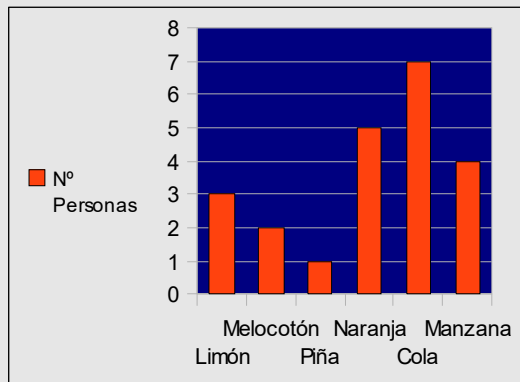
Ejercicios resueltos

5. Los datos corresponden a las contestaciones realizadas por 22 personas elegidas aleatoriamente, acerca del sabor preferido en los refrescos de una determinada marca.

Naranja, manzana, cola, naranja, limón, cola, melocotón, cola, limón, cola, cola, manzana, limón, naranja, cola, piña, manzana, naranja, cola, naranja, manzana y melocotón.

Dibuja el diagrama de barras que representa los datos anteriores.

SOLUCIÓN:

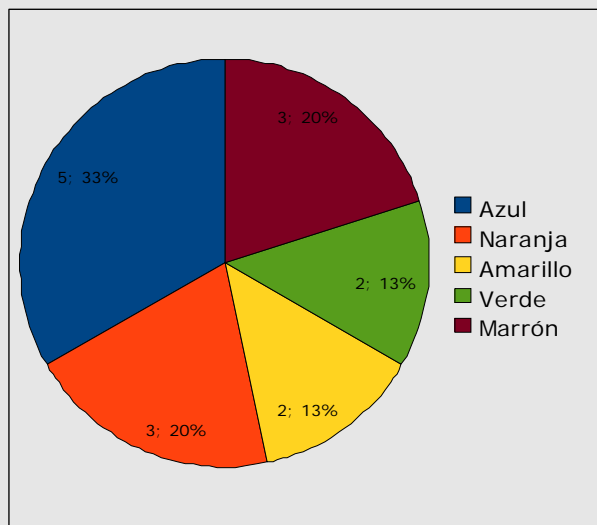


6. Los resultados corresponde a las contestaciones realizadas por 15 estudiantes acerca de cuál es su color preferido.

Las respuestas que dieron son: azul, marrón, naranja, amarillo, azul, naranja, verde, verde, azul, marrón, azul, naranja, amarillo, marrón, y azul.

Dibuja el diagrama de sectores que representa los datos anteriores.

SOLUCIÓN:



5. Gráficos para una variable discreta.

Diagrama de barras.

Es el gráfico estadístico más utilizado para variables discretas. Para elaborar el diagrama, se sitúan en el eje de abscisas los valores correspondientes de la variable.

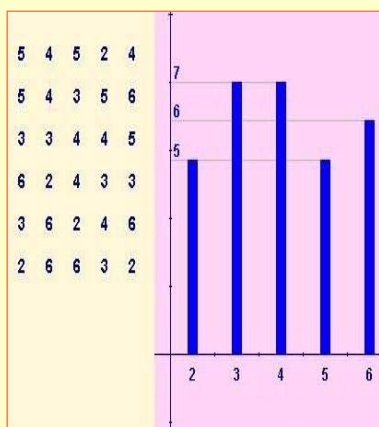
A partir de cada valor se levantan barras del mismo grosor y cuya altura será la correspondiente a cada frecuencia.

Ejemplo

Faltas de ortografía.

La profesora ha anotado el número de faltas de ortografía de sus estudiantes.

Quiere una representación que le permita ver los datos rápidamente, sabiendo cuántos estudiantes comenten un número determinado de faltas de ortografía.



Polígonos de frecuencias.

El polígono de frecuencias se construye a partir del diagrama de barras, uniendo los puntos medios de la base superior de los rectángulos que constituyen las barras.

Si se construye un diagrama de barras considerando en lugar de las frecuencias las frecuencias acumuladas y unimos los puntos medios de las bases superiores mediante segmentos, obtenemos una poligonal creciente que denominamos *polígono de frecuencias acumuladas*.

Ejemplo

Número de llamadas.

Una empresa de telecomunicaciones quiere hacer un estudio sobre sus clientes, viendo el número de llamadas que recibe un grupo de estos.

El estudio se realiza sobre 30 personas, anotando el número de llamadas recibidas en un día.

X_i	f_i	F_i
0	7	7
1	5	12
2	4	16
3	1	17
4	3	20
5	5	25
6	5	30

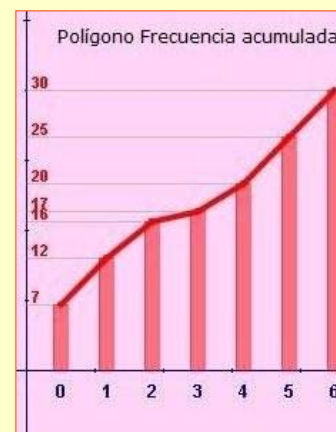
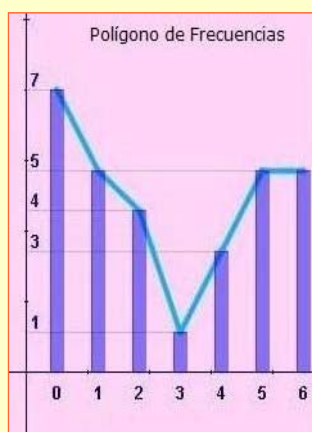
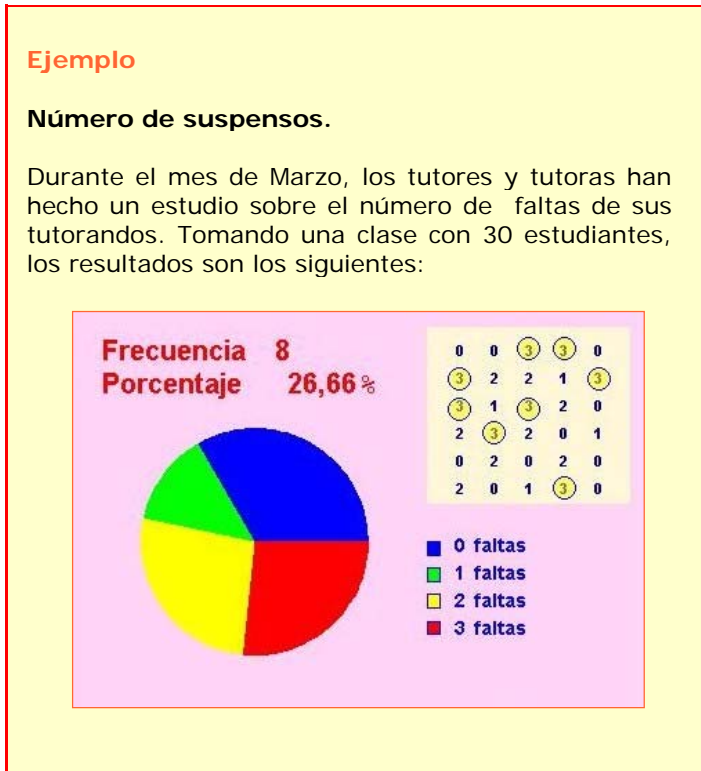


Diagrama de sectores.

Veamos primero un ejemplo.



Como complemento a la tabulación y a veces sustituyendo a ésta, en Estadística es muy habitual recurrir a gráficos cuyo efecto visual directo capta las primeras características de una distribución estadística.

Para variables cuantitativas discretas, así como para las cualitativas, los gráficos que se utilizan con mas frecuencia son:

1. El diagrama de barras
2. El diagrama de sectores

Sin embargo, depende del tipo de información que queramos obtener, a veces, resulta útil realizar el polígono de frecuencias, y el polígono de frecuencia acumuladas.

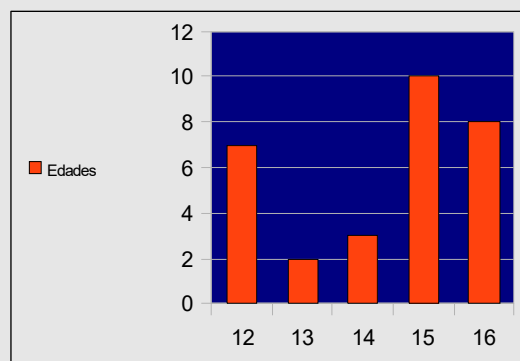
Ejercicio resuelto

7. Las edades de 30 estudiantes de un instituto de enseñanza secundaria son las siguientes:

15 15 16 15 16 16 16 16 16 12 13 12 15 16 14
12 14 12 15 13 14 16 15 15 12 15 12 15 15 12

Representa el diagrama de barras correspondiente:

SOLUCIÓN:



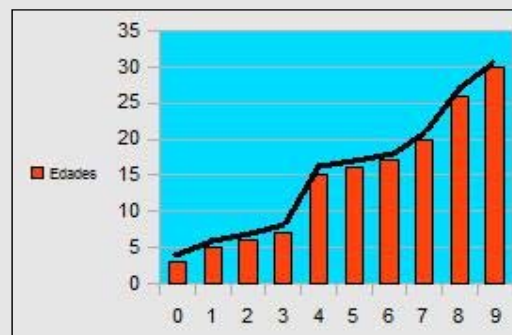
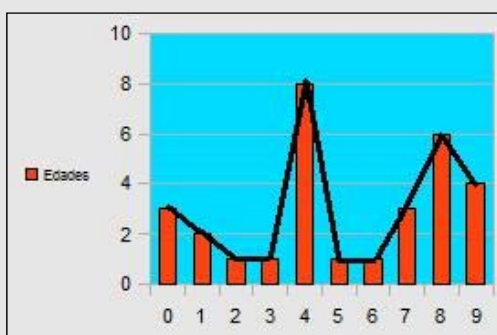
Ejercicios resueltos

8. Los datos corresponden al número de llamadas telefónicas que reciben al día 30 personas.

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

Dibuja el diagrama los polígonos de frecuencia y de frecuencia acumuladas que representa los datos anteriores.

SOLUCIÓN:

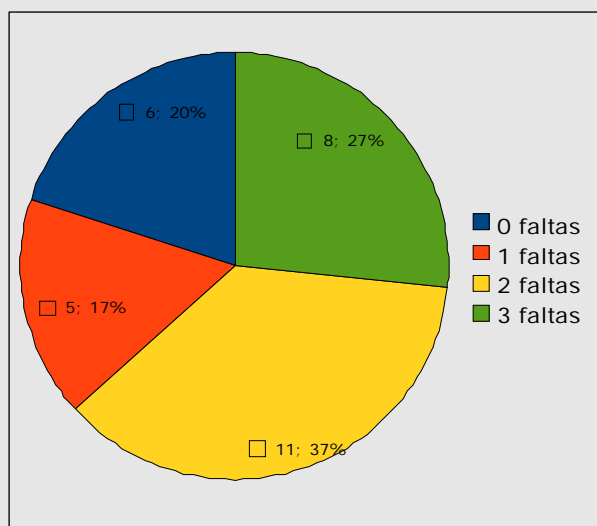


9. Los datos corresponden al número de faltas de ortografía en el mismo texto de 30 estudiantes.

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

Representa el diagrama de sectores correspondiente.

SOLUCIÓN:



6. Medidas de centralización

Media aritmética.

A los parámetros o medidas estadísticas que informan sobre la tendencia habitual o central de los datos de una distribución se les denomina en estadística *medidas de tendencia central*. La más utilizada es la media aritmética.

La media aritmética se define como la suma de todos los datos dividida entre el número total de estos. Como habitualmente dispondremos de una tabla de datos con sus frecuencias, aplicaremos:

$$\bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_n \cdot f_n}{N}$$

1. La media no tiene porqué ser un valor propio de la variable.
2. Es muy sensible a valores extremos en los datos.
3. Se comporta de forma natural en relación a las operaciones aritméticas.

Ejemplo

Faltas de asistencias. (Muchos datos)

Cuando tenemos muchos datos, para evitar realizar una cuenta con gran cantidad de números, primero organizamos una tabla.

Veamos el ejemplo en que se tienen anotados las faltas de asistencia de un grupo de 27 estudiantes.

Hay 6 estudiantes que han faltado 0 veces, 4 que faltaron 1 vez,...

X_i	f_i	$x_i \cdot f_i$
0	6	0
1	4	4
2	4	8
3	0	0
4	6	24
5	3	15
6	4	24
	27	75

Después de tabular los datos, construimos la columna correspondiente a los productos $x_i \cdot f_i$

En la última casilla, calculamos la suma total de la columna $\sum_{i=1}^n x_i \cdot f_i$

La media se obtiene dividiendo $\sum_{i=1}^n x_i \cdot f_i = 75$ entre el total de datos 27,

media = $\frac{75}{27} = 2,77$

Ejemplo

Faltas de asistencias. (Pocos datos)

Las faltas de asistencia de 4 estudiantes en un mes vienen recogida por los siguientes valores: 0, 3, 2 y 1.

La media aritmética se calcula:

$$\frac{0 + 3 + 2 + 1}{4} = 1,5$$

Mediana.

La mediana es aquel valor de la variable estadística que deja el 50% de observaciones inferiores a él; así pues, la mediana divide en dos partes iguales a la distribución estadística.

Dentro de las propiedades de la mediana se pueden destacar:

1. Como medida descriptiva no se ve tan afectada como la media por la presencia de valores extremos.
2. Es de cálculo rápido y de fácil interpretación.
3. Tiene propiedades matemáticas complicadas que hacen que se utilice poco en inferencia estadística.

Caso de pocos datos y en número impar.

En este caso se procede a ordenar los datos de menor a mayor, se considera el valor de la mediana el que corresponde al lugar central.

Caso de pocos datos y en número par.

En este caso se procede a ordenar los datos de menor a mayor, se considera el valor de la mediana el correspondiente a la semisuma de los dos lugares centrales.

Ejemplo

La mediana del número de suspensos. (Muchos datos)

Entramos en una clase de 25 estudiantes y preguntamos el número de suspensos en la última evaluación, hay 4 estudiantes con 0 suspensos, 2 con 1 suspensos,...

Como tenemos muchos datos, los organizamos en la siguiente tabla para calcular la mediana.

Para localizar la mediana, en primer lugar calculamos la mitad de los datos:

$$\frac{N}{2} = 12,5$$

Ahora buscamos en la columna de frecuencias acumuladas la primera vez que se supera a la mitad de los datos.

El valor correspondiente de X_i es la mediana de la distribución estadística. En este caso:

$$Me = 3$$

X_i	f_i	F_i
0	4	4
1	2	6
2	3	9
3	7	16
4	4	20
5	5	25
	25	25

Moda

Se define la moda como el valor de la variable estadística que tiene la frecuencia absoluta más alta.

Si existen varios valores con esta característica, entonces se dice que la distribución tiene varias modas (*plurimodal*).

Esta medida de centralización es sin duda la de más fácil cálculo. Se suele utilizar como complemento a la media aritmética y mediana ya que por sí sola no aporta una información determinante de la distribución.

No es tan sensible como la media aritmética a valores extremos.

Ejemplo

Número de llamadas.

En un grupo de 20 personas se recogen el número de llamadas que realizan durante un día.

Resultando los siguientes valores: 4 personas hacen 1 llamada, 3 personas hacen 2 llamadas, 2 personas hacen 3 llamadas...



Observa que en este ejemplo tenemos que la distribución es bimodal, ya que $X_1 = 1$ y $X_5 = 5$ corresponden con $f_1 = 4 = f_5$. Siendo ambas el máximo número de llamadas.

Compara dicho dato con lo ya aprendido de la media aritmética y la mediana.

Ejercicios resueltos

10. Las edades de un grupo de 9 amigas son: 12, 14, 13, 16, 13, 15, 15, 17 y 13. Calcula la media, mediana y moda.

SOLUCIÓN:

X = Edad	f	F	X·f
12	1	1	12
13	3	4	39
14	1	5	14
15	2	7	30
16	1	8	16
17	1	9	17
			128

Media: $\frac{128}{9} = 14,22$

Mediana 14 (si ordenamos los datos, aparece en la posición 5).

Moda: 13 (aparece 3 veces).

11. El número de llamadas telefónicas que reciben al día los 9 integrantes de una familia son:

7, 8, 15, 12, 13, 5, 10, 4, 8

Calcula la media, mediana y moda.

SOLUCIÓN:

X = N° de llamadas	f	F	X·f
4	1	1	4
5	1	2	5
7	1	3	7
8	2	5	16
10	1	6	10
12	1	7	12
13	1	8	13
15	1	9	15
			73

Media: $\frac{73}{9} = 8,11$

Mediana 8 (si ordenamos los datos, aparece en la posición 5).

Moda: 8 (aparece 2 veces).

La media aritmética nos informa que si tú te has comido dos pollos y yo ninguno, de promedio ambos nos hemos comido uno. Pero es obvio que tú te has dado un atracón y yo he pasado hambre.

Las medidas de centralización aportan una información, pero es necesario completarlas con otras medidas que denominaremos de dispersión. Así pues, en el ejemplo anterior:

- El rango de valores va de 0 a 2 pollos, hay una amplitud de $2-0=2$.
- Tú te has tomado un pollo más del promedio y yo uno menos. Ambos nos hemos desviado una unidad respecto a la media.

Ejemplo

X_i	f_i	$X_i \cdot f_i$	$ \bar{x} - X_i \cdot f_i$
5	2	10	29,44
10	1	10	9,72
15	3	45	14,16
20	4	80	1,11
25	6	150	31,66
30	2	60	20,55
18	355	106,66	

$$\text{Rango} = 30 - 5 = 25$$

$$\text{DM} = \frac{106,66}{18} = 5,92$$

7. Medidas de dispersión

- Las medidas de **dispersión** indican si los datos están más o menos agrupados respecto de las medidas de centralización. En particular respecto a la media aritmética.

Rango

El rango o recorrido es la diferencia entre el mayor y el menor valor de la variable. Indica la longitud del intervalo en el que se hallan todos los datos.

Desviación

La desviación respecto a la media es la diferencia en valor absoluto entre cada valor de la variable estadística y la media aritmética.

$$D_i = |x_i - \bar{x}|$$

Desviación media

Es la media aritmética de las desviaciones de los diferentes datos.

$$D_{\bar{x}} = \frac{|x_1 - \bar{x}|f_1 + |x_2 - \bar{x}|f_2 + \dots + |x_n - \bar{x}|f_n}{N}$$

EJERCICIOS resueltos

12. Calcula el rango y la desviación media de los datos:

	8	8	6	10	9	6	7	8	9	7
	7	6	6	7	9	5	5	7	10	7
x_i	f_i	$x_i \cdot f_i$	$ x_i - \bar{x} \cdot f_i$							
5	2	10	4,8							
6	4	24	5,6							
7	6	42	2,4							
8	3	24	1,8							
9	3	27	4,8							
10	2	20	5,2							
20	147	2940	24,6							

Solución:

El rango oscila entre 5 y 10 con una amplitud de 5.

Hacemos el recuento.

$$\text{La media: } \bar{x} = \frac{147}{20} = 7,4$$

Calculamos la desviación de cada dato respecto a la media, en valor absoluto. La media de las desviaciones:

$$\text{DM} = \frac{24,6}{20} = 1,23$$

13. Calcula la desviación media de los datos tabulados siguientes:

x_i	f_i	$x_i \cdot f_i$	$ x_i - \bar{x} \cdot f_i$
100	7	700	2831,82
300	8	2400	1636,36
500	13	6500	59,09
700	9	6300	1759,09
900	7	6300	2768,18
Total:	44	22200	9054,55

Solución:

Calculamos la media:

$$\bar{x} = \frac{22200}{44} = 504,55$$

Completamos la última columna:

$$\text{DM} = \frac{9054,55}{44} = 205,59$$

Para practicar



A. Clasifica según el carácter de la variable las siguientes situaciones:

1. Situaciones:

Cantante favorito
Longitud de espárragos
Marca de refresco favorita
Tipo de música preferida
Raza de perros
Nº días soleados al mes

2. Situaciones:

Nº días de vacaciones
Autor literario favorito
Nº hermanos
Nota media en selectividad
Temperatura media ciudad
Nº días falta a clase

3. Situaciones:

Nº días lluviosos al mes
Tiempo de espera autobús
Nº faltas en un dictado
Color de ojos
Películas vista al mes
Nota media en selectividad

B. Realiza una tabulación que incluya la frecuencia absoluta, relativa y sus acumuladas, cuando sea necesario aproxima hasta las centésimas, de los datos que se corresponden con las situaciones siguientes:

4. El número de veces que han cambiado de domicilio 23 personas.
2, 2, 0, 2, 4, 2, 4, 4, 3, 4, 3, 3, 3, 3, 0, 1, 0, 4, 0, 3, 0, 3 y 5.

5. El número de hermanos que tienen 20 estudiantes de un centro.
1, 4, 0, 2, 3, 1, 0, 3, 4, 1, 1, 3, 3, 3, 4, 1, 1, 2, 1 y 1.

6. El número de dormitorios de 28 viviendas de una ciudad.
3, 5, 0, 4, 2, 3, 0, 0, 1, 1, 3, 0, 2, 4, 1, 3, 3, 3, 1, 4, 4, 0, 3, 3, 1, 4, 3 y 1.

7. El número de faltas de ortografía en el mismo texto de 30 estudiantes son:
0, 0, 2, 1, 4, 6, 6, 5, 0, 4, 6, 5, 5, 1, 0, 0, 3, 5, 1, 2, 5, 1, 0, 5, 2, 0, 4, 3, 6 y 4.

C. Efectúa una tabulación de los datos en la que aparezcan las columnas de frecuencias absolutas y relativas. Cuando sea necesario aproxima hasta las centésimas.

8. El sabor preferido en los refrescos de una determinada marca de 22 personas.
Naranja, cola, naranja, limón, cola, melocotón, cola, limón, cola, cola, manzana, limón, naranja, cola, piña, cola, naranja, manzana, naranja, cola, naranja y manzana.

9. Las actividades realizadas en por 20 estudiantes en sus tiempos libres.
Deporte, amigos, idiomas, música, idiomas, idiomas, amigos, música, deportes, baile, baile, música, deportes, idiomas, cine, amigos, deportes, amigos, música, y cine.

10. El tipo de programa de televisión que prefieren ver en su tiempo libre. Ficción, infantiles, deportivos, espectáculo, documentales, infantiles, ficción, culturales, espectáculo, infantiles, ficción, deportivos, deportivos, espectáculo, ficción, documentales, culturales, ficción, deportivos y espectáculo.

D. Dibuja el diagrama de barras correspondiente a las situaciones que aparecen.

11. Preguntamos a 25 estudiantes elegidos aleatoriamente por el tipo de música que prefieren escuchar. Los resultados son: disco, disco, rock, clásica, rock, latina, pop, rock, pop, latina, rock, flamenco, flamenco, latina, flamenco, latina, rock, clásica, disco, disco, latina, rock, disco, latina y rock.

12. Los datos corresponden a las contestaciones realizadas por 25 personas elegidas aleatoriamente, acerca del tipo de película que prefieren ver. Los datos son los siguientes: comedia, terror, suspense, comedia, aventura, drama, aventura, aventura, comedia, musical, terror, musical, suspense, aventura, comedia, terror, musical, terror, terror, comedia, suspense, suspense, comedia, aventura y aventura.

13. Los resultados siguientes corresponden a las contestaciones realizadas por 25 estudiantes acerca de las actividades realizadas en sus tiempo libre. Deporte, amigos, amigos, idiomas, música, idiomas, deporte, música, idiomas, amigos, música, deportes, baile, música, baile, música, deportes, idiomas, cine, amigos, deportes, cine, amigos, música, y cine.

14. Las edades de 30 estudiantes de un instituto de enseñanza secundaria son las siguientes:
12, 13, 12, 15, 12, 15, 13, 14, 15, 12, 12, 12, 15, 15, 13, 14, 14, 16, 13, 12, 13, 14, 15, 16, 15, 13, 14, 15, 15 y 12.

15. Número de asignaturas suspensas de 30 estudiantes son:
2, 0, 3, 2, 4, 0, 1, 3, 4, 2, 5, 0, 3, 2, 5, 4, 1, 1, 0, 1, 1, 1, 3, 4, 2, 0, 5, 5, 3 y 2.

16. El número de llamadas telefónicas que reciben un día un grupo de 20 amigos son:
4, 5, 1, 9, 5, 3, 6, 3, 7, 8, 3, 4, 1, 0, 9, 7, 6, 2, 1 y 5.

17. Para un estudio de accesibilidad, durante 30 días anotamos el número de plazas libres de aparcamiento a las 5 de la tarde.
1, 1, 3, 5, 4, 0, 1, 3, 4, 2, 5, 0, 3, 2, 5, 4, 3, 1, 0, 1, 4, 1, 3, 4, 2, 3, 5, 4, 3 y 0.

E. Dibuja el diagrama de sectores correspondiente a las situaciones que aparecen en los ejercicios D.11, D.12, D.16 y D.17

F. Realiza el polígono de frecuencia y el de frecuencia acumulada de los ejercicios del apartado D.14 y D.15

G. Halla las medidas de centralización de los ejercicios del apartado B.6 y B.7

H. Halla las medidas de dispersión de los ejercicios del apartado B.6 y B.7

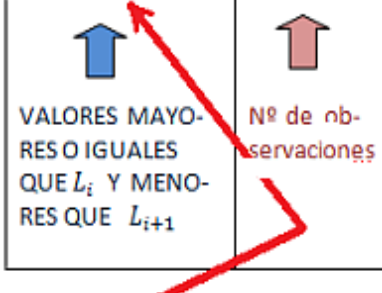
Para saber más 

En estadística en muchas ocasiones una variable discreta toma tal variedad de valores que para que la tabulación sea efectiva, debe realizarse mediante intervalos. La variable queda de esta manera dividida en clases (intervalos, generalmente de la misma amplitud). Esta es la técnica que se utiliza para variables continuas.

Además del rango y la desviación media otras medidas de dispersión que se utilizan normalmente son:

- Varianza
- Desviación típica

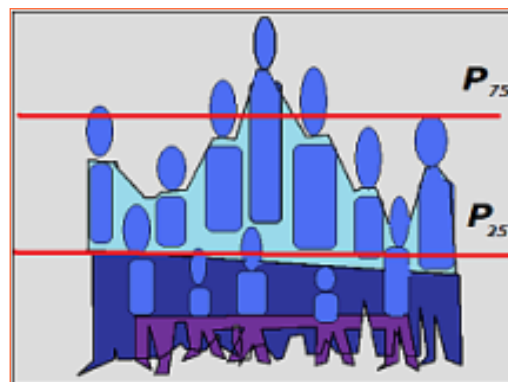
Intervalos	frecuencia
...	...
...	...
$[L_i, L_{i+1})$	f_i



VALORES MAYORES O IGUALES QUE L_i Y MENORES QUE L_{i+1}

Nº de observaciones

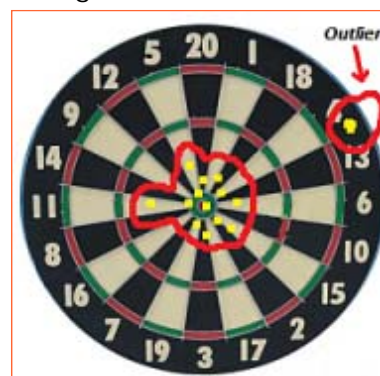
Como representante de todo el intervalo se considera el valor central (MARCA DE CLASE)



El principal parámetro estadístico y el más utilizado es la *media aritmética*, sin embargo una característica importante de la media es que se ve muy afectada por valores extremos en la distribución.

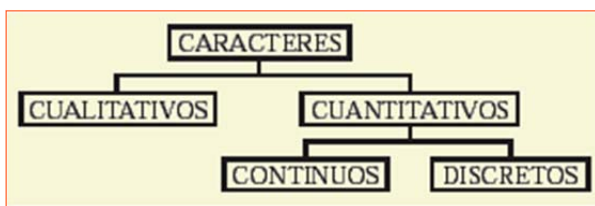
La variación de un simple dato en la distribución afecta a las medidas de tendencia central aunque no de la misma forma. En estadística a un valor especialmente anómalo se le denomina **Outlier**. Decidir si en un estudio estadístico se depuran estos valores extremos, es una de las primeras acciones que debe realizar un investigador.

La *mediana* es la medida de posición que más se utiliza, sin embargo es muy habitual en la mayoría de los estudios estadísticos hacer referencia a otras medidas de posición como los **cuartiles**, **deciles** o **percentiles**.



Recuerda lo más importante

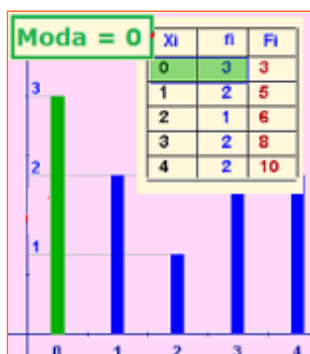
Las primeras definiciones necesarias para el inicio de cualquier estudio estadístico son: **Población**, **Individuo**, **Muestra** y **Carácter**.



CUALITATIVOS: No expresables numéricamente.
CUANTITATIVOS: Se puede expresar mediante número.

Medidas de tendencia central

Moda: Valor que tiene la frecuencia absoluta más alta.



Es la única que puede calcularse para variable cualitativa.

No es tan sensible como la media aritmética a valores extremos.

Media aritmética: suma de todos los datos dividida entre el número total de estos.

Xi	fi	xi.fi
1	5	5
2	2	4
3	2	6
4	5	20
5	0	0
6	0	0
7	5	35
8	5	40
9	0	0
	24	110

Muy sensible a valores extremos en los datos

No tiene por qué ser un valor propio de la variable

$$\text{media} = \frac{\sum_{i=1}^n x_i \cdot f_i}{\sum_{i=1}^n f_i} = \frac{110}{24} = 4,58$$

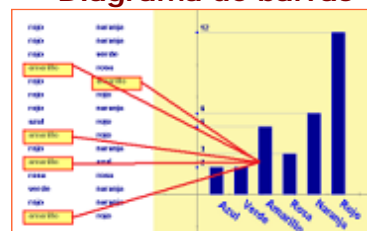
Mediana: divide en dos partes iguales a la distribución estadística.

Xi	fi	Fi
0	8	8
1	7	15
2	5	20
3	16	36
4	9	45

Me = 3
 22,50
 45 / 2

De cálculo rápido y de fácil interpretación. No es tan sensible como la media aritmética a valores extremos.

Diagrama de barras



Polígono de frecuencias



Diagrama de sectores



Autoevaluación



1. Dados los datos: 7, 5, 7, 5, 6, y 8. Calcula la media aritmética con dos cifras decimales.
2. La nota media obtenida en cinco exámenes ha sido 6,8. Si cuatro de las notas han sido 4,7; 9,5; 8,3 y 9,2. ¿Cuál es la quinta?
3. La nota media de cuatro notas es 4,2. Si he sacado ahora un 8,0. ¿Qué nota media tendré ahora?
4. En una prueba de gimnasia la puntuación de cada atleta se calcula eliminando la peor y la mejor nota de los jueces. Si las puntuaciones obtenidas han sido: 8,1; 9,0; 9,3; 9,6; 8,2; 8,7 y 9,5. ¿Qué nota corresponde?
5. Calcula la mediana de estos datos: 9, 15, 19, 22, 31, 38 y 43.
6. Calcula la desviación media de estos datos: 18, 15, 28, 22 y 36.
7. En una distribución de 63 datos, la frecuencia absoluta de un valor de la variable es 21. ¿Cuántos grados corresponderían a ese valor en un diagrama de sectores?
8. Para obtener la nota final de curso nos dan a elegir entre la media, la mediana y la moda de las nueve notas obtenidas. ¿Cuál elegirías? Las notas son: 6, 3, 3, 4, 6, 8, 7, 9 y 3.
9. Calcula la mediana de estos datos: 1, 17, 26, 5, 11 y 24.
10. Indica si la variable es discreta, continua o cualitativa: Perímetro craneal.

Soluciones de los ejercicios para practicar

A.1 Cualitativa, continua, cualitativa, cualitativa, cualitativa, discreta.

A.2 Discreta, cualitativa, discreta, continua, continua, discreta.

A.3 Discreta, continua, discreta, cualitativa, discreta, continua.

B.4

Nº de cambios de domicilio	f	h	F	H
0	5	0,17	5	0,17
1	1	0,03	6	0,21
2	4	0,14	10	0,34
3	7	0,24	17	0,59
4	5	0,17	22	0,76
5	1	0,03	23	0,79
	23			

B.5

Nº de hermanos	f	h	F	H
0	2	0,07	2	0,07
1	8	0,28	10	0,34
2	2	0,07	12	0,41
3	5	0,17	17	0,59
4	3	0,1	20	0,69
	20			

B.6

Nº de dormitorios	f	h	F	H
0	5	0,17	5	0,17
1	6	0,21	11	0,38
2	2	0,07	13	0,45
3	9	0,31	22	0,76
4	5	0,17	27	0,93
5	1	0,03	28	0,97
	28			

B.7

Nº de faltas de ortografía	f	h	F	H
0	7	0,23	7	0,23
1	4	0,13	11	0,37
2	3	0,1	14	0,47
3	2	0,07	16	0,53
4	4	0,13	20	0,67
5	6	0,2	26	0,87
6	4	0,13	30	1
	30			

Actividades tiempo libre	f	h
Deporte	4	0,14
Amigos	4	0,14
Idiomas	4	0,14
Baile	2	0,07
Cine	2	0,07
Música	4	0,14
	20	

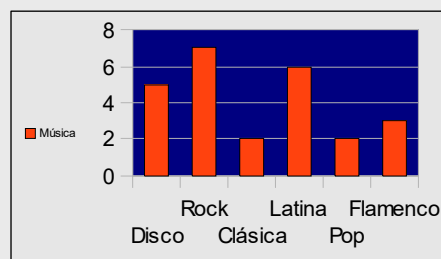
C.9

Sabor preferido	f	h
Naranja	6	0,21
Cola	8	0,28
Limón	3	0,1
Piña	1	0,03
Melocotón	1	0,03
Manzana	3	0,1
	22	

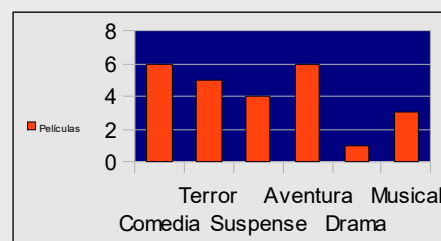
C.10

Programa de televisión	f	h
Ficción	5	0,17
Infantiles	3	0,1
Deportivos	4	0,14
Espectáculo	4	0,14
Documentales	2	0,07
Culturales	2	0,07
	20	

D.11

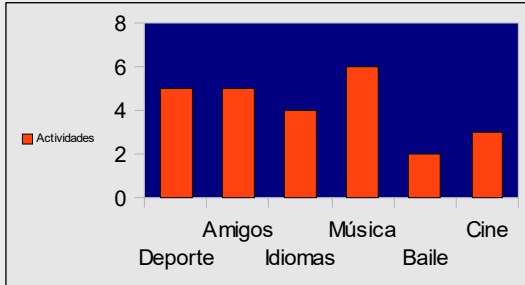


D.12

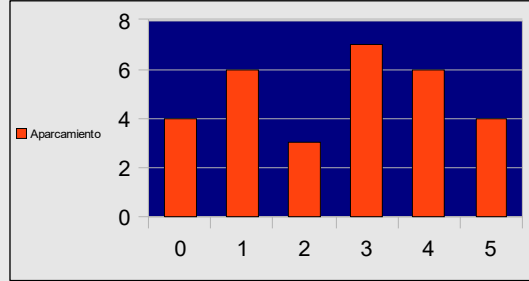


C.8

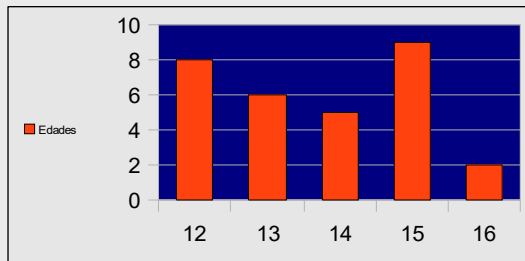
D.13



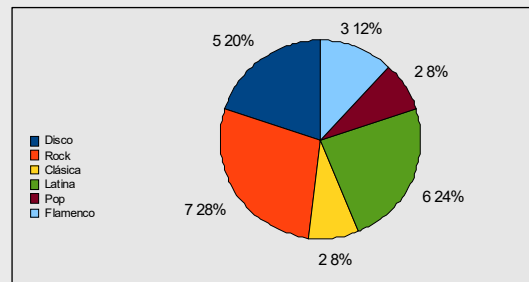
D.17



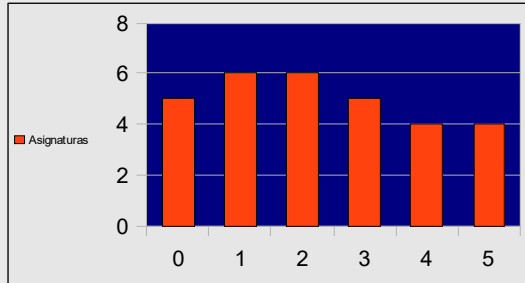
D.14



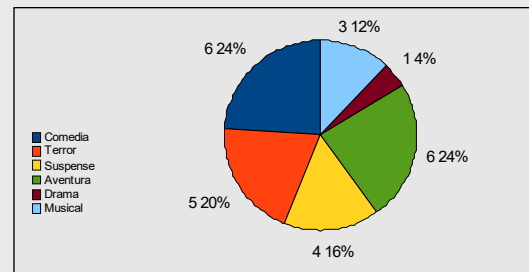
E.11



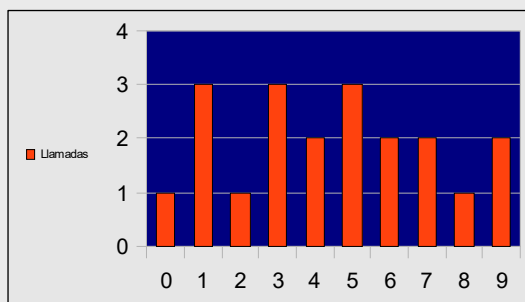
D.15



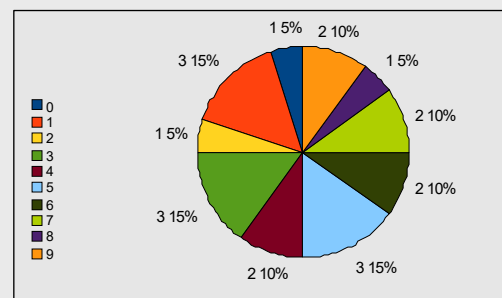
E.12



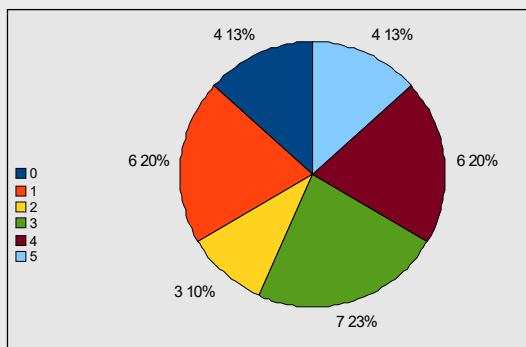
D.16



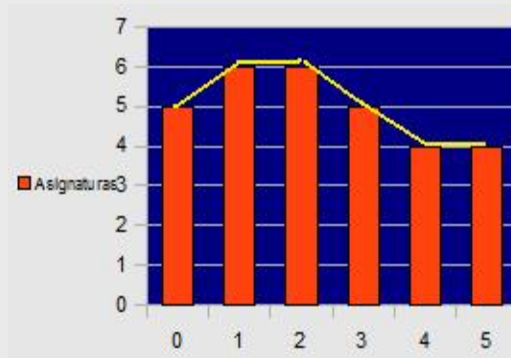
E.16



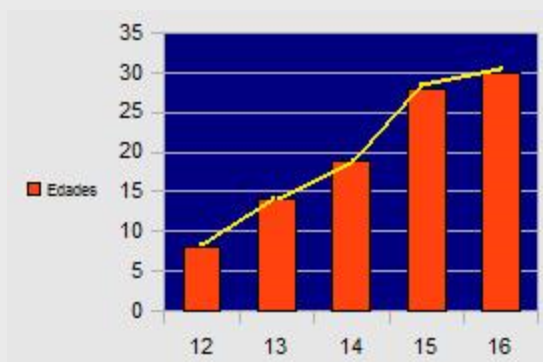
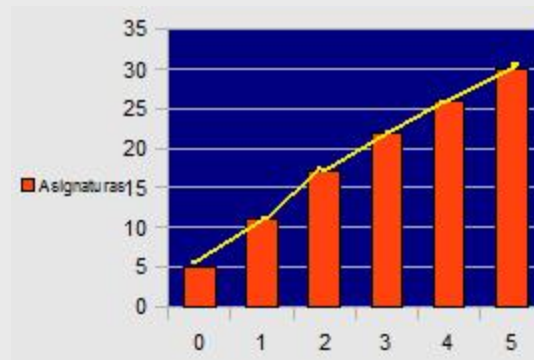
E.17



G.15



G.14



G.6

Media = 2,21
 Mediana = 3
 Moda = 3

G.7

Media = 2,86
 Mediana = 3
 Moda = 0

H.6

Rango = 5
 Desviación media = 1,34

H.7

Rango = 6
 Desviación media = 2,01

Soluciones AUTOEVALUACIÓN

1. 6'33
2. 2'3
3. 5'8
4. 8'94
5. 22
6. 6,56
7. 120°
8. La mediana
9. 14
10. Cualitativa