

REPASO DE ESTADÍSTICA BÁSICA

En Estadística, los datos deben ser ciertos; los resultados, claros y precisos.

• Caracteres y escalas de medición

Al hacer un trabajo estadístico hay que decidir los caracteres (las propiedades) que desean estudiarse. Un carácter puede ser cuantitativo o cualitativo.

Los valores que toma un carácter pueden medirse en distintas escalas: nominal, ordinal, de intervalo, o de proporción.

La escala nominal consiste en situar a cada individuo o elemento en una u otra clase dada (por ejemplo, hombre/mujer; lugar de nacimiento). Pertenecer a una u otra clase no significa ser mejor o peor, indica que son distintos.

La escala ordinal sitúa los posibles valores en orden (primero, segundo, ...), sin que la *distancia* entre dos posiciones consecutivas sea necesariamente constante, fija. En esta escala puede distinguirse, además, entre mayor y menor. Por ejemplo, la posición de los equipos de fútbol en el Campeonato de Liga; o las categorías profesionales en una empresa.

Las escalas nominal y ordinal son apropiadas para caracteres cualitativos.

La escala de intervalo permite asignar a cada individuo un número para así indicar su posición exacta a lo largo de una escala continua. Por ejemplo, la temperatura medida en grados Celsius, donde 10 °C significa más calor que 5 °C, pero no el doble de calor.

La escala de proporción (o proporcional) es la más perfecta. En ella existe un cero absoluto y, además, tiene sentido hablar de doble o mitad (un ejemplo de esta medida sería la longitud). Las escalas de intervalo y proporcional se usan para medir caracteres cuantitativos.

• Tablas de frecuencias

Se utilizan para facilitar la lectura e interpretación de grandes conjuntos de datos. Los datos suelen agruparse, indicando su frecuencia absoluta o relativa; simple (f_i) o acumulada (F_i).

La agrupación puede hacerse también en intervalos de clase. El punto medio de cada uno de esos intervalos sería el valor que representa a todos; se llama marca de clase (M.c.).

Ejemplos:

x_i	f_i	F_i
0	1	1
1	5	6
2	12	18
3	10	28
4	15	43
5	17	60
6	11	71
7	7	78
8	0	78
9	2	80
Totales	80	80

La tabla de la izquierda indica el número de errores cometidos por 80 personas al realizar un determinado test.

La tabla de abajo es la apropiada para variables continuas. En ella se indica el tiempo de espera de 80 personas que han tomado un autobús

Intervalo	M.c.	f_i	%	F_i	%a
[0, 2)	1	4	5	4	5
[2, 4)	3	15	18,75	19	23,75
[4, 6)	5	26	32,5	45	56,25
[6, 8)	7	21	26,25	66	82,5
[8, 10)	9	14	17,5	80	100
Totales		80	100	80	100

Diagrama de tallo y hojas

Es otro método de organización y visualización de un conjunto numérico de datos. Cada número del conjunto se representa por una *hoja* y un *tallo*. La hoja es la cifra de las unidades; el tallo es la cifra de las decenas. Una raya vertical separa el tallo de las hojas.

```

2 | 4
4 | 2 3 5 5 7 7 7
5 | 1 1 5 8
6 | 0 7
8 | 2
    
```

Ejemplo: El diagrama adjunto es el asociado a los datos

45 47 24 42 43 51 58 60 45 47 47 67 82 51 55

2 | 4 representa 24

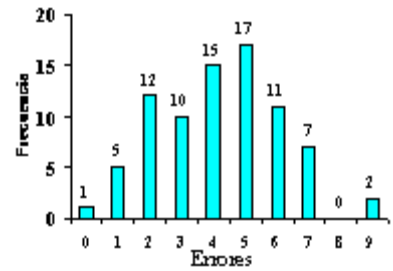
Gráficos estadísticos

Diagramas de barras

Son gráficos que representan cada valor de la variable mediante una barra proporcional a la frecuencia con que se presenta. Las barras deben estar separadas, como en la figura del ejemplo, que se corresponde con la primera tabla del ejemplo anterior.

Los diagramas de barras son apropiados para datos medidos en escala nominal u ordinal.

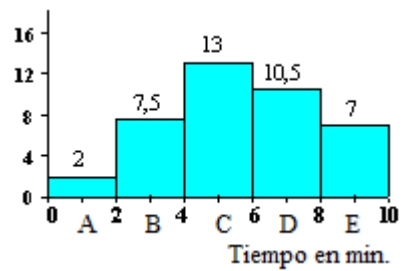
Este gráfico puede ser el “resumen visual” del número de errores cometidos por 80 personas al realizar un determinado test.



Histogramas

Se usan para variables agrupadas en intervalos, asignando a cada intervalo un rectángulo de superficie proporcional a su frecuencia. Por tanto, como la base, la amplitud del intervalo es 2, la altura de cada rectángulo se halla dividiendo la frecuencia que representa entre 2.

Intervalo	Mc	fi
[0, 2)	1	4
[2, 4)	3	15
[4, 6)	5	26
[6, 8)	7	21
[8, 10)	9	14
Totales		80

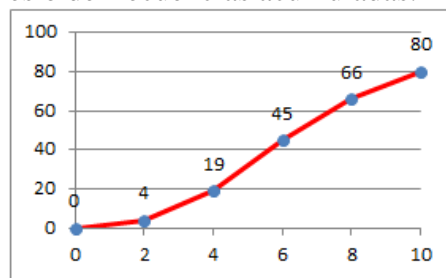
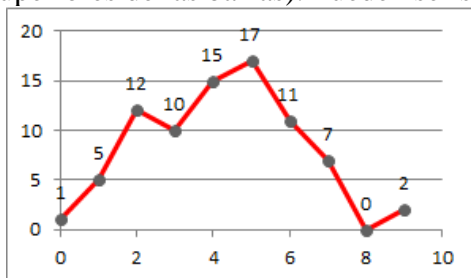


→ En la figura adjunta, el segundo rectángulo [B], que representa una frecuencia de 15, tiene altura $7,5 = 15/2$; el 3º, [C], $13 = 26/2$; ... Este histograma puede resumir el tiempo de espera de 80 personas que han tomado un autobús. El intervalo [C] = [4, 6), indica que 26 personas han esperado entre 2 y 4 minutos.

Los histogramas son apropiados para variables continuas (medidas en escala de intervalo o de proporción); por eso las barras van unidas y tienen la anchura indicada por el intervalo.

Poligonal de frecuencias

Los histogramas, y algunos diagramas de barras, también se pueden representar por una poligonal de frecuencias, que es la línea que une los puntos correspondientes a las frecuencias de cada valor (extremos superiores de las barras). Pueden ser simples o de frecuencias acumuladas.



La poligonal (simple) de la izquierda representa los datos del diagrama de barras de arriba; la de la derecha, acumulada, se corresponde con el histograma anterior.

→ Muchos de estos gráficos pueden hacerse con Excel: véase el Problema n. 2.

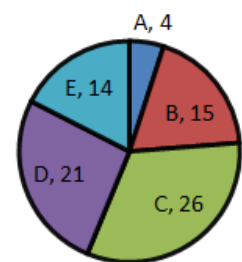
Diagrama de sectores

En estos gráficos, cada suceso viene representado por un sector circular de amplitud proporcional a su frecuencia. La amplitud de cada sector se halla mediante una regla de tres.

Este ejemplo se corresponde con los datos del histograma de arriba. La amplitud de cada sector, en grados, es:

A, 4: 18° ; B, 15: $67,5^\circ$; C, 26: 117° ; D, 21: $94,5^\circ$; E, 14: 63° .

→ La amplitud, por ejemplo, de B, se halla así: $\frac{360}{80} \cdot 15 = 4,5 \cdot 15 = 67,5$.



• Medidas de centralización

Están relacionadas con el promedio de los datos estudiados, y dan una idea de los valores más representativos para todo el conjunto.

La media aritmética

Se calcula sumando el valor de todos los datos y dividiendo por el número de ellos. Esto es:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}; \quad \text{o más breve: } \bar{x} = \frac{\sum x_i}{n}.$$

→ Para datos agrupados: $\bar{x} = \frac{\sum x_i f_i}{\sum f_i}$, donde f_i es el número de veces que se repite el valor x_i .

Ejemplos:

a) Las calificaciones de una alumna a lo largo del curso en los distritos exámenes de Matemáticas han sido: 7, 9, 6, 8, 10, 9, 5, 7 y 7.

→ Si todos los exámenes tienen el mismo peso, su nota media será:

$$\bar{x} = \frac{7+9+6+8+10+9+5+7+7}{9} \approx 7,56.$$

b) Las calificaciones de 85 alumnos en un examen fueron las que se dan en la siguiente tabla:

Nota: x_i	1	2	3	4	5	6	7	8	9	10
N.º de alumnos: f_i	2	3	6	10	22	15	12	7	5	3

La nota media será:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{1 \cdot 2 + 2 \cdot 3 + 3 \cdot 6 + 4 \cdot 10 + 5 \cdot 22 + 6 \cdot 15 + 7 \cdot 12 + 8 \cdot 7 + 9 \cdot 5 + 10 \cdot 3}{2 + 3 + 6 + 10 + 22 + 15 + 12 + 7 + 5 + 3} = \frac{481}{85} \approx 5,66.$$

Media ponderada: $\bar{x}_p = \frac{\sum x_i p_i}{\sum p_i}$, siendo p_i el peso del dato x_i .

Se calcula igual que la media para datos agrupados: el peso (p_i) sustituye a la frecuencia (f_i).

Ejemplos:

En un concurso oposición, para cada uno de los participantes, se valoran los siguientes aspectos, con los pesos que se indica:

méritos, 25 %; examen teórico: 35 %; examen práctico: 40%.

Cada uno de esos apartados se valoran de 0 a 10 puntos.

Si los opositores A, B y C obtuvieron los puntos que se indican (respectivamente en los aspectos considerados), ¿en qué orden se clasificaron?

Concursante A: 4, 7, 10. Concursante B: 7, 7, 8. Concursante C: 10, 7, 6.

(Observa que los puntos directos suman, respectivamente, 21, 22 y 23).

→ Las calificaciones ponderadas son:

$$\bar{x}_p(A) = \frac{4 \cdot 25 + 7 \cdot 35 + 10 \cdot 40}{25 + 35 + 40} = \frac{745}{100} = 7,45. \quad \bar{x}_p(B) = \frac{7 \cdot 25 + 7 \cdot 35 + 8 \cdot 40}{25 + 35 + 40} = \frac{740}{100} = 7,40.$$

$$\bar{x}_p(C) = \frac{10 \cdot 25 + 7 \cdot 35 + 6 \cdot 40}{25 + 35 + 40} = \frac{735}{100} = 7,35.$$

El orden es el dado: 1º, A; 2º, B; 3º, C.

La mediana: es el valor del dato que ocupa el lugar intermedio. Los datos deben estar ordenados.

La moda: es el valor que se presenta con mayor frecuencia. (La moda, en muchos casos, no pasa de ser un simple dato anecdótico).

• **Medidas de posición**

Indican la situación, en términos porcentuales, de algunos elementos de la distribución. Los datos deben estar ordenados de menor a mayor.

Amplitud, rango o recorrido. Es la diferencia entre los valores de los datos máximo y mínimo. La información que proporciona es imprecisa, pues sólo tiene en cuenta los valores extremos.

Cuartiles, deciles y percentiles

Cuartiles: Valores de las posiciones correspondientes al 25 %, al 50 % y al 75 % de los datos.

Deciles: Valores correspondientes al 10 %, 20 %, ... y 90 % de los datos.

Percentiles (o centiles) dan el valor de la posición correspondiente a cualquier porcentaje.

En muchos casos, su cálculo requiere aplicar la interpolación. Véase el problema propuesto 8.

• **Medidas de dispersión**

Dan una idea del *alejamiento* de los datos respecto de la media.

La varianza:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n}; \quad \sigma^2 = \frac{\sum x_i^2}{n} - \bar{x}^2.$$

→ Para datos agrupados: $\sigma^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}; \quad \sigma^2 = \frac{\sum x_i^2 f_i}{\sum f_i} - \bar{x}^2.$

La desviación típica es la raíz cuadrada de la varianza. En consecuencia:

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}; \quad \sigma = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}; \quad \sigma = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i}}.$$

La desviación típica es una medida de la desigualdad de los datos estudiados: a mayor desigualdad corresponde mayor desviación típica.

Ejemplo:

a) Para hallar desviación típica de los valores 1, 4, 5, 6, 14:

1) Se halla su media, $\bar{x} = \frac{1+4+5+6+14}{5} = 6.$

2) Aplicando la fórmula $\sigma = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2} \rightarrow$

$$\sigma = \sqrt{\frac{1^2 + 4^2 + 5^2 + 6^2 + 14^2}{5} - 6^2} = \sqrt{\frac{274}{5} - 36} = \sqrt{18,8} \approx 4,34.$$

x_i	x_i^2
1	1
4	16
5	25
6	36
14	196
$\sum x_i = 36$	$\sum x_i^2 = 274$

b) Datos agrupados.

Las calificaciones de 85 alumnos en un examen fueron las que se dan en la siguiente tabla:

Nota: x_i	1	2	3	4	5	6	7	8	9	10	Sumas
N.º de alumnos: f_i	2	3	6	10	22	15	12	7	5	3	85
Producto $(x_i \cdot f_i)$	2	6	18	40	110	90	84	56	45	30	481
Valores x_i^2	1	4	9	16	25	36	49	64	81	100	—
Producto $(x_i^2 \cdot f_i)$	2	12	54	160	550	540	588	448	405	300	3059

Por tanto, aplicando las fórmulas anteriores:

$$\bar{x} = \frac{481}{85} \approx 5,66; \quad \sigma = \sqrt{\frac{3059}{85} - (5,66)^2} = \sqrt{3,95} \approx 1,99.$$

Uso de la calculadora

Cada calculadora puede presentar pautas diferentes: consulta el manual de la tuya. (También puedes buscar en Internet tecleando el nombre y modelo de tu calculadora).

- Las pautas de uso para una de las calculadoras más frecuentes son:

(1) Hay que poner la calculadora en el “modo estadístico”: SD.

–Pulsar **MODE** **2** (en la pantalla aparecerá SD).

–Borrar los datos de memoria: Pulsar **SHIFT** **AC**.

(2) Introducir los datos, sumando con la tecla **M+**

Por ejemplo, para datos simples: x_1 **M+** x_2 **M+** ... x_n **M+**.

(3) Los parámetros se obtienen pulsando:

SHIFT **2** **1** → \bar{x} (media); **SHIFT** **2** **2** → σ_n (desv. típica)

- Pulsando **SHIFT** **1** se obtienen otros resultados: ($\sum x_i^2$, con 1; $\sum x_i$, con 2; n , con 3).

→ **Ensayo 1:** Con la calculadora, halla la media y la desviación típica de los valores 1, 4, 5, 6, 14.

(Sol. $\sum x_i^2 = 274$; $\sum x_i = 30$; $n = 5$; $\bar{x} = 6$; $\sigma = 4,335896678$).

–Para datos agrupados:

(1) Poner la calculadora en el modo SD y borrar los datos de memoria: **MODE** **2** **SHIFT** **AC**.

(2) Introducir los datos como sigue:

x_1 **SHIFT** **,** f_1 **M+** x_2 **SHIFT** **,** f_2 **M+** ... x_n **SHIFT** **,** f_n **M+**.

(3) Los parámetros se obtienen pulsando:

SHIFT **2** **1** → \bar{x} (media); **SHIFT** **2** **2** → σ_n (desv. típica)

- Pulsando **SHIFT** **1** se obtienen otros resultados: ($\sum x_i^2$, con 1; $\sum x_i$, con 2; n , con 3).

→ **Ensayo 2:** Con la calculadora, halla la media y la desviación típica de los valores:

Nota: x_i	1	2	3	4	5	6	7	8	9	10
N.º de alumnos: f_i	2	3	6	10	22	15	12	7	5	3

(Sol. $\sum x_i^2 = 3059$; $\sum x_i = 481$; $n = 85$; $\bar{x} = 5,658823529$; $\sigma = 1,991469698$).

Nota: Utilizando **GeoGebra** también se pueden obtener esos parámetros:

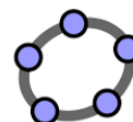
Para datos simples (los mismos del Ensayo 1 de arriba), teclear:

media(1,4,5,6,14) → sale 6; de(1,4,5,6,14) → sale 4,34. (de ≡ desv. estándar)

Para datos agrupados (los mismos del Ensayo 2 anterior), teclear:

media({1,2,3,4,5,6,7,8,9,10}, {2,3,6,10,22,15,12,7,5,3}) → sale 5,66.

de({1,2,3,4,5,6,7,8,9,10}, {2,3,6,10,22,15,12,7,5,3}) → sale 1,99.



• **El coeficiente de variación**

Es una medida de la *dispersión relativa* del conjunto de datos. Se define como: $CV = \frac{\sigma}{\bar{x}}$. (Sirve

comparar dos conjuntos: es más disperso el que tiene un CV mayor).

El coeficiente de variación suele darse en porcentajes: $CV = \frac{\sigma}{\bar{x}} \cdot 100$.

→ Véase el problema propuesto 12.

Ejemplo:

El coeficiente de variación de los conjuntos de datos anteriores es:

$$CV(a) = \frac{4,34}{6} \approx 0,723 \rightarrow 72,3\%; \quad CV(b) = \frac{1,99}{5,66} \approx 0,352 \rightarrow 35,2\%.$$

• **Recursos informáticos: confección de gráficos**

Con frecuencia, y no solo en Matemáticas, hay necesidad de realizar trabajos que requieran una presentación vistosa adjuntando, por ejemplo, gráficos adecuados. Aquí se harán algunos gráficos utilizando Excel.



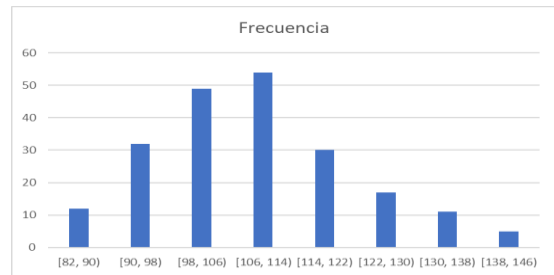
Construcción de un histograma y de una poligonal acumulada

→ Supongamos que los datos del cociente intelectual de 210 alumnos de un centro de bachillerato son los dados en la siguiente tabla.

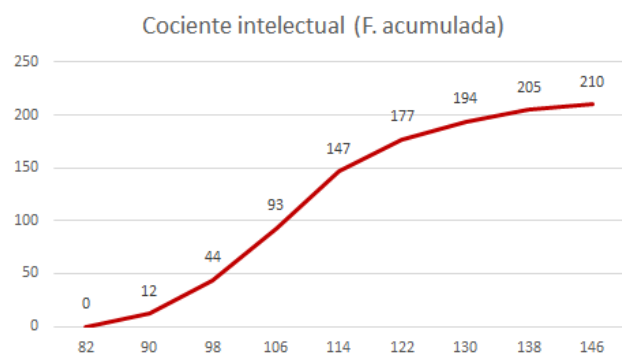
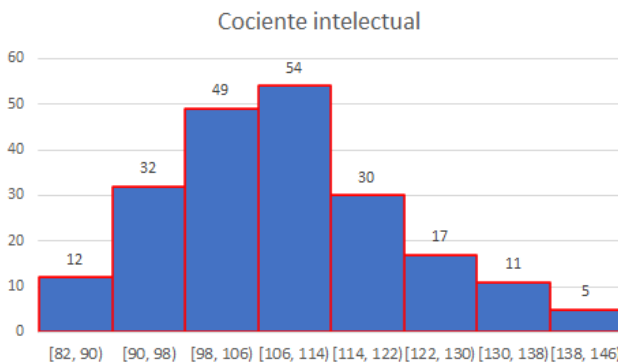
Intervalo	[82, 90)	[90, 98)	[98, 106)	[106, 114)	[114, 122)	[122, 130)	[130, 138)	[138, 146)
Frecuencia	12	32	49	54	30	17	11	5

Estos datos son apropiados para construir un histograma. En Excel puede hacerse como sigue:

- 1) Teclear los datos del problema. (Si hay posibilidad pueden copiarse directamente: tal como se dan aquí: marcar, copiar, abrir Excel y pegar).
- 2) Marcando la tabla completa ir a Insertar gráfico → gráficos recomendados (elegir columnas en 2-D). Aparece el gráfico adjunto.



- 3) Pulsando sobre las barras con el botón derecho del ratón aparece un cuadro con distintas opciones: ir a dar formato a los datos; con ancho de rango 0 % las barras se juntan.
- 4) En Estilo de gráfico se puede cambiar el título, colores, tamaño de letra... Así se obtiene el gráfico que sigue (izquierda).



El gráfico de la derecha es la poligonal acumulada para los mismos datos.

Para hacerla es sido necesario modificar la forma de dar los datos, escribiéndolos como sigue:

Intervalo	82	90	98	106	114	122	130	138	146
Frecuencia	0	12	32	49	54	30	17	11	5

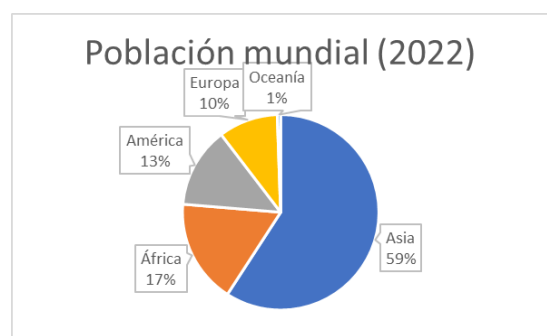
Marcando esta nueva tabla completa: ir a Insertar → gráficos recomendados (elegir la opción correspondiente). Aparece la poligonal de arriba (en Estilo de gráfico hacer los ajustes necesarios).

Construcción de un gráfico de sectores

→ Los datos de la población mundial se teclean como se indica.

Año 2022	Millones (ONU)
Asia	4504
África	1310
América	1014
Europa	747
Oceanía	42

Eligiendo el gráfico de sectores aparece la figura de la derecha. (Hay que hacer los ajustes necesarios).



PROBLEMAS PROPUESTOS

1. En la siguiente tabla se dan los datos correspondientes a las notas de Matemáticas de 60 alumnos de 1º Bachillerato.

Notas	IN: [1, 5)	SF: [5, 6)	BI: [6, 7)	NT: [7, 9)	SB: [9, 10]
N.º de alumnos	20	13	12	10	5

- Haz una tabla de frecuencias y porcentajes, simple y acumulada.
- Dibuja el correspondiente histograma.
- Representa los datos mediante un diagrama de sectores y mediante una poligonal acumulativa.

2. El número de turismos matriculados en España, para el período 2007/2018, se da en la siguiente tabla:

Año	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Miles de turismos	1634	1185	971	1000	818	711	742	890	1094	1230	1342	1425

- Tomando como base 100 el número de turismos matriculados en el año 2007 expresa en números índices la variación de la serie.
- (Opcativo). Representa los datos mediante una poligonal simple usando Excel.

3. La precipitación (P) y la temperatura media mensual (T) registradas en Soria a lo largo del año son:

Mes	E	F	M	A	M	J	J	A	S	O	N	D
P (mm)	44	45	48	47	62	55	32	31	47	46	49	55
T (°C)	1,3	3,1	5,6	7,5	10,6	15,6	18,1	18,1	15	9,4	5,6	3,1

Representa gráficamente estos datos mediante un climograma.

4. Siete estudiantes han leído este curso el siguiente número de libros: 3 4 5 6 5 7 5

Para estos datos, determina:

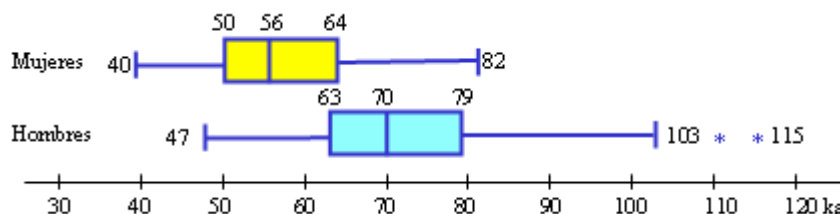
- La media.
- La mediana.
- La moda.
- El rango.

5. En una empresa hay 3 directivos, 50 operarios y 8 vendedores. Los sueldos mensuales, en euros, de cada categoría son los siguientes: directivos, 4000 €; operarios, 1400 €; vendedores, 2000 €.

- Halla la moda, la mediana y la media de los sueldos.
- ¿Qué medida es más representativa del promedio?

6. En primero de bachillerato de un centro escolar hay tres grupos, A, B y C, con 30, 35 y 25 alumnos, respectivamente. La nota media en Matemáticas fue, también respectivamente, de 5,3, 6,5 y 5,6. Halla la nota media de Matemáticas de todos los alumnos de primero.

7. El gráfico siguiente representa los pesos (en kg) de un grupo similar de hombres y mujeres.



- Indica los valores de las medianas respectivas.
- ¿Cuánto vale en cada caso el rango intercuartílico?
- ¿Hay algún elemento extraño? ¿Cuál es su peso?
- ¿Qué porcentaje de mujeres pesa entre 40 y 50 kg?
- ¿Dónde se da más homogeneidad de pesos, entre los más flacos o entre los más pesados?

8. El cociente intelectual de los 210 alumnos de un centro de bachillerato se da en la tabla adjunta:

Intervalo	[82, 90)	[90, 98)	[98, 106)	[106, 114)	[114, 122)	[122, 130)	[130, 138)	[138, 146)
Frecuencia	12	32	49	54	30	17	11	5

- a) Calcula los cuartiles y el rango intercuartílico.
- b) Halla la diferencia entre los deciles 3 y 6.
- c) Calcula la puntuación necesaria para pertenecer al 15 % de alumnos con mayor cociente intelectual.
- d) Teniendo en cuenta lo hecho en el punto de **Recursos informáticos** (Gráfico de Cociente intelectual, F. acumulada), da una explicación gráfica para cada una de las preguntas anteriores.

9. Se ha preguntado a 50 mujeres sobre su número de hijos, obteniéndose los resultados:

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

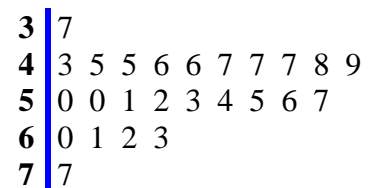
Construye la tabla de frecuencias y calcula la media, varianza y desviación típica.

10. Se ha realizado una encuesta a los 40 empleados de una empresa para saber cuánto tiempo tardan en llegar desde su casa hasta su puesto de trabajo. Las respuestas, en minutos, son las siguientes:

30 42 37 50 15 35 90 65 38 45 30 12 78 20 35 41 25 32 85 25
 41 28 50 30 20 60 14 36 48 32 27 30 76 30 51 28 25 22 17 10

- a) Construye la tabla de frecuencias agrupando los datos en intervalos.
- b) Calcula la mediana, la moda, la media y la desviación típica.

11. Halla la media y la desviación típica de los datos correspondientes al diagrama de tallo y hojas adjunto.



12. Los rendimientos medios (en kilogramos por hectárea) en España, para los cereales que se indican, fueron:

Año	2010	2011	2012	2013	2014
Trigo	2150	3100	2300	2830	2840
Maíz	9450	9220	9720	9510	9110

3 | 7 representa 37 kilos

Halla los rendimientos medios para el quinquenio de cada cereal. ¿Qué cereal es más fiable?

13. A un congreso asisten seis mujeres cuyas edades (en años) son: 27, 34, 38, 42, 33, 36.

- a) Calcula la media y varianza de sus edades.
- b) Cinco años después coinciden las mismas mujeres. A partir de los cálculos anteriores, halla la nueva media y varianza de sus edades.

14. El siguiente gráfico representa un total de 600 elementos. ¿Cuál es la frecuencia de cada categoría?

