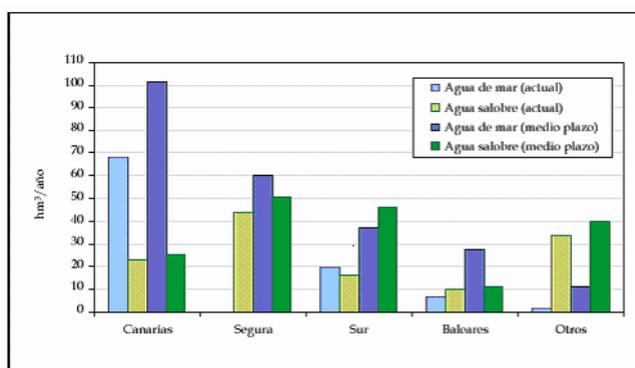


Tema 11. Estadística

1. Introducción

Cada día oímos noticias del tipo...

(PD/Agencia EFE).- El calentamiento global ha esquebrajado las plataformas de hielo en la Antártida y el aumento de icebergs ha alterado los sistemas ecológicos en torno a esas moles de hielo, según un estudio publicado en la revista Science.



(PD/Agencias).- El Fondo Mundial para la Naturaleza (WWF, por sus siglas en inglés) critica en un informe la "frenética construcción" de desalinizadoras en España y su impacto negativo en el medioambiente y el cambio climático.

Son muy importantes los **estudios estadísticos** que conducen a este tipo de informaciones, ya que basándose en estos estudios los gobiernos diseñan sus planes de gestión del medio ambiente. Son múltiples las aplicaciones de la estadística en el campo de las Ciencias Sociales, por ejemplo, para estudiar datos sobre población, inmigración, encuestas políticas, etc. Incluso hay una asignatura ofertada en distintas Universidades españolas denominada "Modelos estadísticos para el medio ambiente" y empresas dedicadas a hacer estudios estadísticos sobre temas medioambientales.

Por ejemplo si se quiere conocer la población de aves en Las Tablas de Daimiel, el nivel de contaminación de los acuíferos o cómo controlar los residuos urbanos, necesitamos:

- conocer los datos objetivos
- ordenarlos
- analizarlos
- sacar conclusiones.

Pero sobre todo se necesita desarrollar un **espíritu crítico ante las distintas noticias científicas basadas en estudios estadísticos**, que se dan como "verdades absolutas". Estamos acostumbrados a que datos objetivos sean interpretados según el interés de quien los muestra.

2. Planificamos el trabajo

Mientras Javier estaba con Luís en el parque leyó en el periódico una noticia que decía “Según un estudio estadístico, el 60,6% de los castellano-manchegos opina que principal problema relacionado con el medio ambiente son los incendios forestales”. Este titular le llevo a la siguiente reflexión “¿Cómo se puede saber lo que opinan los castellano-manchegos en general, si yo soy de Guadalajara y a mí no me ha preguntado nadie?”.

Para realizar un estudio estadístico hay que tener en cuenta una serie de pasos a seguir que podemos resumir en: ¿Qué queremos saber? Y ¿de quién lo queremos saber?

2.1. ¿Qué queremos saber?: Variables estadísticas

Muchos estudios estadísticos comienzan con una pregunta o preguntas sobre un tema concreto. En estos casos en primer lugar habrá que crear un **cuestionario**. Por ejemplo si nos planteamos un estudio sobre “Impacto medioambiental en Castilla-La Mancha” podríamos formular la pregunta:

“¿Qué problema relacionado con el medio ambiente le preocupa más?”

Las respuestas a esta pregunta pueden ser:

- **abiertas:** cada persona entrevistada puede dar tantas respuestas como le apetezca.
- **abiertas pero limitadas:** cada persona entrevistada podría dar una o dos o tres o un número predeterminado de antemano de respuestas libres.
- **cerradas:** cada persona entrevistada elige una o varias opciones sobre un listado prefijado de respuestas posibles.

Habría por tanto que decidir si se crea un **cuestionario**:

- **abierto:** cada uno puede contestar lo que quiera.
- **limitado:** con un número prefijado de posibles respuestas.
- **cerrado:** más cómodo para el entrevistado pero que puede “deformar” el estudio, ya que el listado de posibles respuestas va a depender del encuestador y su buen criterio.

Para evitar la “manipulación” en un cuestionario cerrado, siempre debería existir la opción de respuesta **“otra respuesta diferente a las propuestas”**

Al conjunto final de respuestas obtenidas le llamamos **VARIABLE ESTADÍSTICA**. Las Variables Estadísticas pueden ser de dos tipos:

CUALITATIVAS: No son números.

- Color preferido de un grupo de gente.
- Partido al que votarás en las siguientes elecciones.
- Problema relacionado con el medio ambiente que preocupa más a los castellano-manchegos.

CUANTITATIVAS: Son números, y pueden ser **discretas**, si representan valores numéricos aislados, o **continuas**, si representan valores numéricos en una escala de números real.

- La altura o el peso de un grupo de personas (continuas)
- Número de hermanos, número de casas, número de llamadas realizadas (discreta)



En el caso de que el estudio se refiera, por ejemplo, a conocer “el tamaño en milímetros de una determinada especie de mosquito” no será necesario un cuestionario, habrá en este caso que elegir un **sistema o aparato de medida** adecuado para llevar a cabo las mediciones necesarias. Los datos obtenidos en este ejemplo concreto formarán una **variable cuantitativa**.



2. 2. ¿De quién lo queremos saber?: Población y muestra

Imagina que queremos saber cuál es el problema medioambiental que más preocupa a los vecinos de Valdepeñas. Podría ocurrir que fuese muy distinto del que pueda preocupar a los vecinos de Manzanares o a la población manchega en su conjunto. Por esto es muy importante, una vez terminado el cuestionario o el método de recogida de datos adecuado al estudio, decidir a quién va dirigido:

- Si vamos a preguntar a todos los nacidos o nacidas en Castilla-La Mancha o en alguna población en concreto de nuestra geografía.
- Si se preguntaría a todas las personas censadas en Castilla-La Mancha (o lugar en concreto de ésta)
- Si preguntaremos a cualquier persona que esté en ese momento dado en Castilla-La Mancha (o en una determinada población manchega).



Al conjunto total de personas o de objetos de los que nos interesa conocer una determinada opinión o característica es a lo que llamaremos **POBLACIÓN**.

Sea cual sea la elección, *preguntar a toda la población normalmente es imposible, así que habrá que elegir un grupo que represente toda la población.*

El grupo elegido para que responda al cuestionario o del que se van a recoger determinados datos, es a lo que se denomina **MUESTRA**.

Cuanto mayor sea el número de personas que forman la muestra más fiable será el estudio estadístico. Y aquí es donde pueden empezar los problemas, porque si elegimos mal la muestra los resultados no serán reales.

2.2.1. ¿Cómo elegir entonces la muestra?

Esto es una de las partes más complejas de la estadística y hay teorías matemáticas muy complicadas al respecto.

La elección de la muestra puede ser:

- **Aleatoria:** se eligen al azar. Este método tiene como ventaja la objetividad, pero, por el contrario, es posible que la muestra no sea representativa.
- **Intencional:** el encuestador elige a los que quiere. En este caso la muestra estará caracterizada por la subjetividad de quién realiza el estudio.



EJEMPLOS. Muestras aleatorias e intencionadas

En los siguientes ejemplos vemos las diferencias de los resultados obtenidos en una recogida de datos en función del modelo elegido:

1. Dato: altura, en centímetros, de las primeras diez personas que pasan por la calle

- Aleatorio: datos tomados en una calle cualquiera de una ciudad de Castilla La Mancha.
167, 169, 165, 178, 177, 169, 181, 176, 168 y 175
- Intencionado: datos tomados en la puerta de un pabellón polideportivo a la hora en la que salen de su entrenamiento unos jugadores de un equipo de baloncesto.

174, 199, 197, 187, 206, 189, 188, 203, 188 y 178

2. Dato: color del cabello de las diez primera personas que pasan por la calle:

- Aleatorio: datos tomados en un lugar cualquiera de nuestra ciudad.
Moreno, Moreno, Rubio, Castaño, Moreno, Moreno, Castaño, Moreno, Rubio y Moreno
- Intencionado: datos tomados en la puerta de un concurso de imitadoras de Marilyn Monroe e imitadores de Robert Redford.

Moreno, Rubio, Rubio, Rubio, Castaño, Moreno, Castaño, Rubio, Rubio, Moreno

Como vemos, las diferencias son sustanciales, como era de esperar debido a la elección del lugar en el que realiza en cada caso la toma de datos o muestreo.

Muchas veces el sentido común nos dirá cuál es la mejor manera de elegir la muestra.

3. Organización de datos en tablas de frecuencias

Es muy importante la organización de los datos en forma de tabla, ya que los hace más comprensibles y facilita los cálculos. En el caso de variables cuantitativas discretas o cuantitativas continuas cuando hay poca cantidad de datos, se pone en la primera columna cada valor de la variable, de forma ordenada.

| VALORES x_i | FRECUENCIAS ABSOLUTAS n_i | FRECUENCIAS ABSOLUTAS ACUMULADAS N_i | $x_i \cdot n_i$ | $x_i^2 \cdot n_i$ |
|------------------|--------------------------------|--|------------------------|--------------------------|
| x_1 | n_1 | $N_1 = n_1$ | $x_1 \cdot n_1$ | $x_1^2 \cdot n_1$ |
| x_2 | n_2 | $N_2 = N_1 + n_2$ | $x_2 \cdot n_2$ | $x_2^2 \cdot n_2$ |
| x_3 | n_3 | $N_3 = N_2 + n_3$ | $x_3 \cdot n_3$ | $x_3^2 \cdot n_3$ |
| . | . | . | . | . |
| . | . | . | . | . |
| . | . | . | . | . |
| x_k | n_k | $N_k = N_{k-1} + n_k = n$ | $x_k \cdot n_k$ | $x_k^2 \cdot n_k$ |
| | $\sum_{i=1}^k n_i = n$ | | $\sum_{i=1}^k x_i n_i$ | $\sum_{i=1}^k x_i^2 n_i$ |

- ⊙ **Valores:** en esta columna se representan todos los valores obtenidos de forma ordenada
- ⊙ **Frecuencias absolutas:** es la repetición de cada valor de la variable.
- ⊙ **Frecuencias absolutas acumuladas:** en cada fila, es la suma de la frecuencia absoluta más la frecuencia acumulada de la fila anterior.
- ⊙ **$x_i n_i$:** producto del valor de la variable por su frecuencia absoluta
- ⊙ **$x_i^2 n_i$:** producto del cuadrado del valor de la variable por su frecuencia absoluta.



EJEMPLO: Preguntamos a 30 personas el número de hermanos que tienen, y obtenemos los siguientes resultados... Construye la tabla de frecuencias

2 1 0 2 2 1 1 0 0 1
 1 3 4 6 2 3 2 1 0 1
 4 3 3 2 5 1 0 1 0 1

La tabla se construye de la siguiente forma:

| VALORES x_i | FRECUENCIAS ABSOLUTAS n_i | FRECUENCIAS ABSOLUTAS ACUMULADAS N_i | $x_i \cdot n_i$ | $x_i^2 \cdot n_i$ |
|------------------|--------------------------------|--|-----------------|-------------------|
| 0 | 6 | 6 | 0 | 0 |
| 1 | 10 | 16 | 10 | 10 |
| 2 | 6 | 22 | 12 | 24 |
| 3 | 4 | 26 | 12 | 36 |
| 4 | 2 | 28 | 8 | 32 |
| 5 | 1 | 29 | 5 | 25 |
| 6 | 1 | 30 | 6 | 36 |
| | 30 | | 53 | 163 |

Por ejemplo, para la segunda fila, el valor de la variable es 1 (1 hermano).

- En la columna correspondiente a frecuencias absolutas el valor es 10, ya que 10 personas respondieron que tenían sólo 1 hermano.
- En la columna correspondiente a frecuencias absolutas acumuladas, el valor es 16, ya que se suman la frecuencia absoluta de esta fila (10) con el valor de la frecuencia absoluta acumulada de la fila anterior (6).
- En la columna $x_i \cdot n_i$ el valor resulta de multiplicar el valor de la variable (1) por su frecuencia absoluta (10).
- En la columna $x_i^2 \cdot n_i$ el valor resulta de multiplicar el valor de la variable al cuadrado (1^2) por su frecuencia absoluta (10).

4. Organización de los datos en tablas con intervalos o clases

A veces, cuando trabajamos con variables cuantitativas, la cantidad de resultados distintos puede ser muy elevada (por ejemplo, en un sondeo sobre la edad podemos encontrarnos con 30 ó 40 edades distintas). En estos casos, a la hora de crear la tabla, se agrupan los datos por intervalos o clases. Cada uno de ellos tendrá una amplitud, que se calcula restando al extremo mayor el extremo menor.



EJEMPLO. Se preguntan las edades de 20 personas, arrojando los siguientes datos. Construye una tabla de frecuencias con cuatro intervalos de amplitud 19.

14 17 20 65 34 23 76 54 45 45
 46 54 21 67 90 32 41 32 32 19

Para construir la tabla en sí, añadimos una columna más a la izquierda, con los intervalos, y la columna de valores (x_i) ahora se llamará **marca de la clase**. La marca de la clase corresponde al valor medio de los extremos de cada intervalo. En las frecuencias absolutas, se van recontando los datos obtenidos dentro del intervalo que les corresponda. Cuidado con los extremos de los intervalos, pues por ejemplo, según la siguiente tabla, el valor 33 se contabilizaría en la clase [33,52) y no en la [14,33).

| Intervalo | MARCA DE LA CLASE x_i | FRECUENCIAS ABSOLUTAS n_i | FRECUENCIAS ABSOLUTAS ACUMULADAS N_i | $x_i \cdot n_i$ | $x_i^2 \cdot n_i$ |
|-----------|----------------------------|-----------------------------|---|-----------------|-------------------|
| [14,33) | 23'5 | 9 | 9 | 211,5 | 4970,25 |
| [33,52) | 42'5 | 5 | 14 | 212,5 | 9031,25 |
| [52,71) | 61'5 | 4 | 18 | 246 | 15129 |
| [71,90] | 80'5 | 2 | 20 | 161 | 12960,5 |
| | | 20 | | 831 | 42091 |

5. Parámetros estadísticos centrales

Son valores que suelen situarse hacia el centro de la distribución de datos. Los más destacados son la media, la mediana y la moda, que vemos a continuación.

5.1. La media

La **media** es el valor medio de todos los datos de la muestra. Si tenemos acceso a todos los datos de la muestra, la forma de calcularla es sumarlos todos y dividir el resultado entre el número total de datos.



EJEMPLO. Se pregunta la altura a diez personas obteniendo los datos que se muestran a continuación. ¿Cuál será el valor de su media?

167 169 165 178 177 169 181 176 168 175

La media por tanto se calculará como:

$$\frac{167+169+165+178+177+169+181+176+168+175}{10} = 172,5$$

Cuando trabajamos con tablas, para calcular la media hay que dividir el sumatorio de valores $x_i \cdot n_i$ entre el número

de valores n. **Media:** $\bar{x} = \frac{\sum x_i \cdot n_i}{n}$



EJEMPLO. Cálculo de la media en una tabla de frecuencias

Por ejemplo, en la siguiente tabla...

| Intervalo | MARCA DE LA CLASE x_i | FRECUENCIAS ABSOLUTAS n_i | FRECUENCIAS ABSOLUTAS ACUMULADAS N_i | $x_i \cdot n_i$ | $x_i^2 \cdot n_i$ |
|-----------|----------------------------|-----------------------------|---|-----------------|-------------------|
| [14,33) | 23'5 | 9 | 9 | 211,5 | 4970,25 |
| [33,52) | 42'5 | 5 | 14 | 212,5 | 9031,25 |
| [52,71) | 61'5 | 4 | 18 | 246 | 15129 |
| [71,90] | 80'5 | 2 | 20 | 161 | 12960,5 |
| | | 20 | | 831 | 42091 |

... la media será $831/20 = 41,55$

5.2. La mediana

La mediana viene a ser el valor que está en el centro una vez ordenados los datos. Si trabajamos con todos los datos porque su número es pequeño, para calcular la mediana, primero los ordenamos de menor a mayor, y entonces habrá dos casos:

- Si el número de datos es impar, el dato central de la ordenación será la mediana.

Datos ordenados: 2, 2, 3, 5, 5, 7, 9

$$\text{Mediana} = 5$$

- Si el número de datos es par, la media de los dos datos centrales será la mediana.

Datos ordenados: 2, 2, 3, 4, 5, 5, 7, 9

$$\text{Mediana} = (4+5) / 2 = 4,5$$

Cuando hay que calcular la mediana a partir de datos agrupados en una tabla, dividimos el número total de datos entre 2 y se busca el resultado en la columna de las Frecuencias Absolutas Acumuladas. Si se encuentra en esta columna, lo tomaremos como indicador y si no está, se coge como indicador el número mayor más cercano. Después nos fijamos en el valor de la variable (o marca de la clase) que le corresponde a ese indicador, dicho valor es la mediana.



EJEMPLO. Cálculo de la mediana en una tabla de frecuencias

Por ejemplo, en la siguiente tabla la mediana será 42,5

| Intervalo | MARCA DE LA CLASE x_i | FRECUENCIAS ABSOLUTAS n_i | FRECUENCIAS ABSOLUTAS ACUMULADAS N_i | $x_i \cdot n_i$ | $x_i^2 \cdot n_i$ |
|-----------|----------------------------|-----------------------------|---|-----------------|-------------------|
| [14,33) | 23'5 | 9 | 9 | 211,5 | 4970,25 |
| [33,52) | 42'5 | 5 | 14 | 212,5 | 9031,25 |
| [52,71) | 61'5 | 4 | 18 | 246 | 15129 |
| [71,90] | 80'5 | 2 | 20 | 161 | 12960,5 |
| | | 20 | | 831 | 42091 |

- *En primer lugar, dividimos el número total de datos (20) entre 2, lo que nos da 10.*
- *Después, en la columna de frecuencias absolutas acumuladas, buscamos el primer número mayor o igual a 10 (14). Ese será el indicador.*
- *Finalmente, la mediana será el valor de la variable o marca de la clase en la misma fila donde encontramos el indicador.*

5.3. La moda

La moda es el valor de la variable que más se repite. Si tenemos todos los datos, aquel o aquellos valores que más se repitan serán la moda o modas (puede haber más de una).

Datos: 2, 2, 3, 5, 5, 7, 9 → Modas = 2 y 5 (ambos se repiten 2 veces)

Cuando se trabaja con tablas, buscamos el mayor valor (o mayores) en la tabla de Frecuencias Absolutas, y el valor que tome en esa fila la variable o marca de la clase será una Moda.



EJEMPLO. Cálculo de la moda en una tabla de frecuencias

Por ejemplo, en la siguiente tabla la moda será $M_o = 23,5$.

| Intervalo | MARCA DE LA CLASE x_i | FRECUENCIAS ABSOLUTAS n_i | FRECUENCIAS ABSOLUTAS ACUMULADAS N_i | $x_i \cdot n_i$ | $x_i^2 \cdot n_i$ |
|-----------|-------------------------|-----------------------------|--|-----------------|-------------------|
| [14,33) | 23'5 | 9 | 9 | 211,5 | 4970,25 |
| [33,52) | 42'5 | 5 | 14 | 212,5 | 9031,25 |
| [52,71) | 61'5 | 4 | 18 | 246 | 15129 |
| [71,90] | 80'5 | 2 | 20 | 161 | 12960,5 |
| | | 20 | | 831 | 42091 |

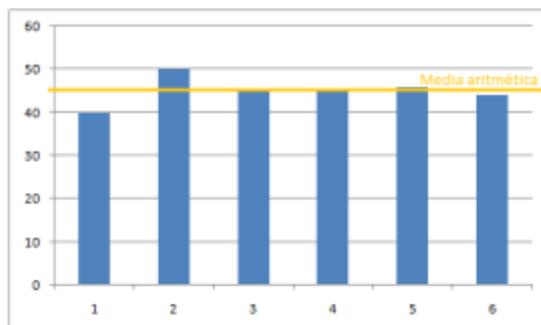
- *Se observa que el valor más alto en la columna de frecuencias absolutas es 9*
- *La moda será el valor de la variable o marca de la clase en la misma fila donde encontramos el indicador.*

6. Parámetros estadísticos de dispersión

Las medidas de dispersión, variabilidad o variación nos indican si esos datos están próximos entre sí o si están dispersos, es decir, nos indican cuán esparcidos se encuentran los datos. Estas medidas de dispersión nos permiten apreciar la distancia que existe entre los datos a un cierto valor central.



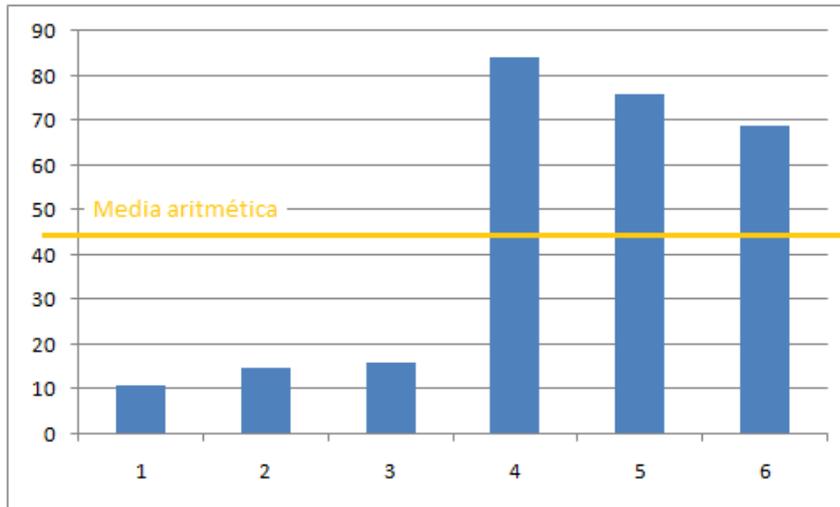
EJEMPLO 1: Si hacemos una encuesta de edades y tenemos de resultados 40, 50, 45, 45, 46 y 44, la media sería 45 años. Si nos fijamos, los datos están muy cercanos numéricamente a esa media, ya que la diferencia de edades con respecto a la media, en cada caso, sería 5, 5, 0, 0, 1 y 1 años. Realmente, ninguna edad se aleja mucho de la media.



Datos poco dispersos porque están cercanos a la media



EJEMPLO 2. Si hacemos una encuesta de edades y tenemos de resultados 11, 15, 16, 84, 76 y 69, la media sería aproximadamente 45 años. Si nos fijamos, los datos están muy dispersos con respecto a esa media, ya que la diferencia de edades con respecto a la media, en cada caso, sería 34, 30, 29, 39, 31 y 24 años. Realmente, ninguna edad es cercana a la media.



Datos muy dispersos con respecto a la media, pues sus valores reales están muy alejados de la media

6.1. La varianza

La varianza, al ser un parámetro de dispersión, sirve para identificar si los datos están cercanos a la media o no. Su valor mínimo es 0, cuando todos los datos sean iguales a la media. Cuando los datos más se acercan a la media, más pequeño será su valor. Se calcula sumando los valores que se obtienen de elevar al cuadrado la diferencia de cada dato con la media, y dividiendo este valor entre el número de datos. Para representar este parámetro se utilizan los símbolos S^2 o σ^2 .

$$S^2 = \frac{1}{n} \sum x_i^2 n_i - \bar{x}^2$$

Cuando queremos calcular la varianza sobre un conjunto de datos, primero debemos calcular la media, y después, calculamos la varianza sumando los valores que se obtienen de elevar al cuadrado la diferencia de cada dato con la media, y dividiendo este valor entre el número de datos.



EJEMPLO. Cálculo de la varianza para dos conjuntos de alturas de diez personas:

Datos 1: 167, 169, 165, 178, 177, 169, 181, 176, 168 y 175

Datos 2: 174, 199, 197, 187, 206, 189, 188, 203, 188 y 178

Calculamos, para cada caso, en una columna, el cuadrado de la diferencia entre cada dato y la media (calculada como la suma de datos entre 10 en cada caso), como se observa en las siguientes tablas:

| | Datos 1 | (Dato - Media) | (Dato - Media) ² | | Datos 2 | (Dato - Media) | (Dato - Media) ² |
|-----------|---------|----------------|-----------------------------|-----------|---------|----------------|-----------------------------|
| | 167 | -5,5 | 30,25 | | 174 | -16,9 | 285,61 |
| | 169 | -3,5 | 12,25 | | 199 | 8,1 | 65,61 |
| | 165 | -7,5 | 56,25 | | 197 | -6,1 | 37,21 |
| | 178 | 5,5 | 30,25 | | 187 | -3,9 | 15,21 |
| | 177 | 4,5 | 20,25 | | 206 | 15,1 | 228,01 |
| | 169 | -3,5 | 12,25 | | 189 | -1,9 | 3,61 |
| | 181 | 8,5 | 72,25 | | 188 | -2,9 | 8,41 |
| | 176 | 3,5 | 12,25 | | 203 | 12,1 | 146,41 |
| | 168 | -4,5 | 20,25 | | 188 | -2,9 | 8,41 |
| | 175 | 2,5 | 6,25 | | 178 | -12,9 | 166,41 |
| Total: | 1725 | | 272,5 | Total: | 1909 | | 964,9 |
| Media: | 172,5 | | | Media: | 190,9 | | |
| Varianza: | 27,25 | | | Varianza: | 96,49 | | |

- En el conjunto Datos 1, la suma de los cuadrados de la diferencia entre cada dato y la media es 272,5. Por tanto, dividiendo entre el número de datos que es 10, obtenemos una varianza de 27,25
- En el conjunto Datos 2, la suma de los cuadrados de la diferencia entre cada dato y la media es 964,9. Por tanto, dividiendo entre el número de datos que es 10, obtenemos una varianza de 96,94
- A la luz de los resultados, los datos están más dispersos en el conjunto Datos 2 pues su varianza es mayor.

En las tablas de frecuencias, se calcula de la siguiente manera:

| Intervalo | MARCA DE LA CLASE x_i | FRECUENCIAS ABSOLUTAS n_i | FRECUENCIAS ABSOLUTAS ACUMULADAS N_i | $x_i \cdot n_i$ | $x_i^2 \cdot n_i$ |
|-----------|----------------------------|-----------------------------|--|-----------------|-------------------|
| [14,33) | 23'5 | 9 | 9 | 211,5 | 4970,25 |
| [33,52) | 42'5 | 5 | 14 | 212,5 | 9031,25 |
| [52,71) | 61'5 | 4 | 18 | 246 | 15129 |
| [71,90] | 80'5 | 2 | 20 | 161 | 12960,5 |
| | | 20 | | 831 | 42091 |

Siendo la media 41,55

$$S^2 = \frac{1}{20} \cdot 42091 - (41,55)^2 = 378,15$$

6.2. La desviación típica

La desviación típica da un valor de las diferencias de los valores con respecto a la media que se obtiene haciendo la raíz cuadrada de la varianza, lo que hace que el valor sea más comprensible y manejable que el obtenido con la propia varianza.

Varianza y Desviación típica:

$$S^2 = \frac{1}{n} \sum x_i^2 n_i - \bar{x}^2 = \frac{203}{30} - (2'1)^2 = 6'7666 - 4'41 = 2'3566$$

$$S = +\sqrt{S^2} = +\sqrt{2'3566} = 1'5351$$

Por ejemplo, si

$$S^2 = \frac{1}{20} \cdot 42091 - (41,55)^2 = 378,15$$

Entonces

$$S = +\sqrt{S^2} = +\sqrt{378,15} = 19,45$$

6.3. El coeficiente de variación

Es el cociente entre la desviación típica y la media. Su fórmula expresa la desviación estándar como porcentaje de la media aritmética, mostrando una mejor interpretación porcentual del grado de variabilidad que la desviación típica o estándar. A mayor valor del coeficiente de variación mayor heterogeneidad de los valores de la variable; y a menor C.V., mayor homogeneidad en los valores de la variable. Suele representarse por medio de las siglas **C.V.**, y se calcula como:

$$C_V = \frac{\sigma}{|\bar{x}|}$$

Por ejemplo, si

Media= 41,55

$$S^2 = \frac{1}{20} \cdot 42091 - (41,55)^2 = 378,15$$

$$S = +\sqrt{S^2} = +\sqrt{378,15} = 19,45$$

Entonces

$$C_V = \frac{\text{Desviación típica}}{\text{Media}} = \frac{19,45}{41,55} = 0,47$$

Si se multiplica por cien, se expresa en tanto por ciento (en este ejemplo 47%)

7. Representación gráfica

Una gráfica estadística es la mejor forma de disponer de toda la información que se haya recogido con una simple “ojeada” y que además permite distinguir, sin dificultad alguna, que opción es la preferida por los encuestados.

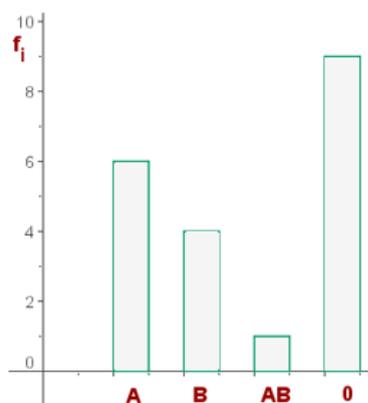


7.1. Diagramas de barras

Los diagramas de barras se caracterizan por:

- ⊙ Son los indicados para variables estadísticas cualitativas o cuantitativas discretas (sin intervalos)
- ⊙ Se representan sobre el eje de abscisas los valores de la variable y sobre el de ordenadas las frecuencias asociadas a cada valor
- ⊙ Se levanta sobre cada valor de la variable un segmento vertical de altura igual a la frecuencia con que se ha observado dicho valor.

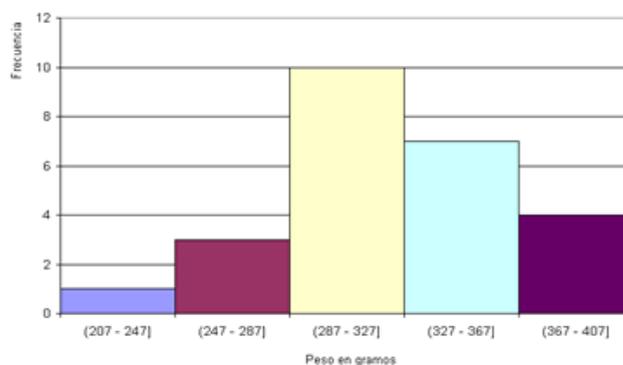
| Grupo sanguíneo | f_i |
|-----------------|-------|
| A | 6 |
| B | 4 |
| AB | 1 |
| O | 9 |
| | 20 |



7.2. Histogramas

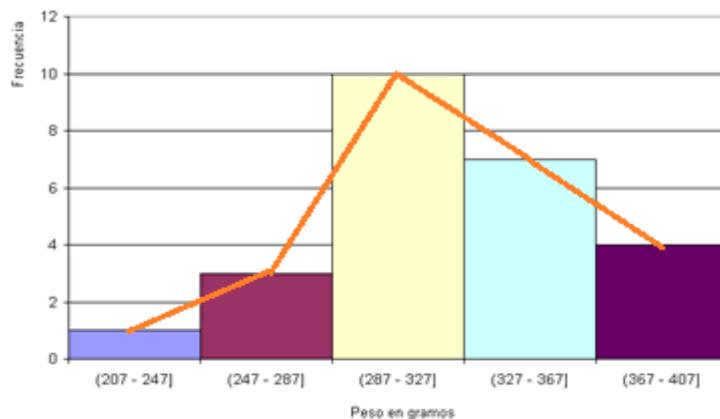
Los histogramas se caracterizan por:

- ⊙ Se utilizan para variables estadísticas cuantitativas expresadas en intervalos
- ⊙ Sobre el eje de abscisas se representan las distintas clases o intervalos en los que se han agrupado los valores de la variable, y sobre cada clase se construye un rectángulo cuya base sea el intervalo y la altura la frecuencia absoluta de dicha clase. Las barras quedarán gráficamente unidas unas a otras.



7.3. Polígonos de frecuencia

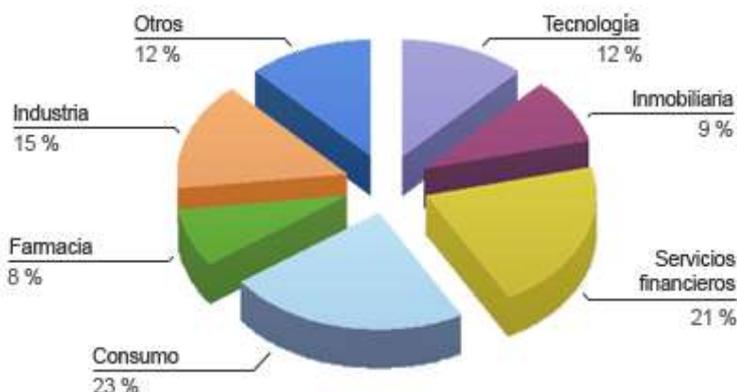
Uniando los puntos medios de las bases superiores de los rectángulos de un histograma (en variables cuantitativas continuas) se dibuja lo que se conoce como polígono de frecuencias.



7.4. Diagramas de sectores

Los diagramas de sectores se caracterizan por:

- Se utilizan para caracteres cualitativos y cuantitativos.
- Consiste en repartir el área del círculo en sectores de tamaño proporcional a la frecuencia de cada valor que ha presentado un determinado carácter.



Para hacer el diagrama de sectores, hay que hacer unos pequeños cálculos que nos dicen el ángulo de cada porción del diagrama. En concreto, para cada valor de la variable o marca de la clase, el ángulo que tomará del círculo vendrá definido por:

$$\text{Ángulo de la variable o clase} = \frac{\text{frecuencia absoluta de la clase}}{\text{número total de datos}} \cdot 360^\circ$$



EJEMPLO

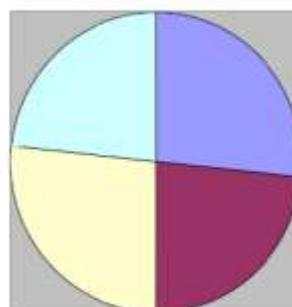
| INTERVALOS $L_{i-1} - L_i$ | MARCA DE CLASE (x_i) | FRECUENCIAS ABSOLUTAS (n_i) | FRECUENCIAS ABSOLUTAS ACUMULADAS (N_i) | $x_i n_i$ | $x_i^2 n_i$ |
|-------------------------------|------------------------------|-------------------------------------|--|-----------|-------------|
| [33,43) | 38 | 8 | 8 | 304 | 11552 |
| [43,53) | 48 | 7 | 18 | 336 | 16128 |
| [53,63) | 58 | 8 | 23 | 464 | 26912 |
| [63,73) | 68 | 7 | 30 | 476 | 32368 |
| | | 30 | | 1580 | 86960 |

$$1^a. - \frac{360^\circ}{30} = \frac{x_1'}{8} \Rightarrow x_1' = \frac{8}{30} 360^\circ = 96^\circ$$

$$2^a. - \frac{360^\circ}{30} = \frac{x_2'}{7} \Rightarrow x_2' = \frac{7}{30} 360^\circ = 84^\circ$$

$$3^a. - \frac{360^\circ}{30} = \frac{x_3'}{8} \Rightarrow x_3' = \frac{8}{30} 360^\circ = 96^\circ$$

$$4^a. - \frac{360^\circ}{30} = \frac{x_4'}{7} \Rightarrow x_4' = \frac{7}{30} 360^\circ = 84^\circ$$



8. Ejemplos de cálculos estadísticos



EJEMPLO 1. Se ha realiza una encuesta a 30 personas sobre el número de veces a la semana que comen pescado, obteniendo los siguientes resultados:

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

Calcula el número medio de veces que se come pescado a la semana, el número más frecuente, el valor mediano, el recorrido de los datos, la varianza y la desviación típica. Realiza el diagrama de barras y el diagrama de sectores de los datos anteriores.

En primer lugar calculamos la tabla de todas las cosas que nos van pidiendo y posteriormente realizamos las operaciones necesarias para obtener las medidas que nos han pedido.

| VALORES DE LA VARIABLE (x_i) | FRECUENCIA S ABSOLUTAS (n_i) | FRECUENCIAS ABSOLUTAS ACUMULADAS (N_i) | $x_i n_i$ | $x_i^2 n_i$ |
|---|---|---|-----------|-------------|
| 0 | 6 | 6 | 0 | 0 |
| 1 | 6 | 12 | 6 | 6 |
| 2 | 5 | 17 | 10 | 20 |
| 3 | 7 | 24 | 21 | 63 |
| 4 | 4 | 28 | 16 | 64 |
| 5 | 2 | 30 | 10 | 50 |
| | 30 | | 63 | 203 |

Una vez realizada la tabla se calculan las medidas pedidas:

$$\text{Media: } \bar{x} = \frac{\sum x_i \cdot n_i}{n} = \frac{63}{30} = 2'1$$

$$\text{Moda: } Mo = 3$$

$$\text{Mediana: } n/2 = 15, \text{ luego } Me = 2$$

$$\text{Recorrido: } Re = 5-0$$

Varianza y Desviación típica:

$$S^2 = \frac{1}{n} \sum x_i^2 n_i - \bar{x}^2 = \frac{203}{30} - (2'1)^2 = 6'7666 - 4'41 = 2'3566$$

$$S = +\sqrt{S^2} = +\sqrt{2'3566} = 1'5351$$

Una vez calculadas las medidas realizamos el gráfico

Diagrama de barras:

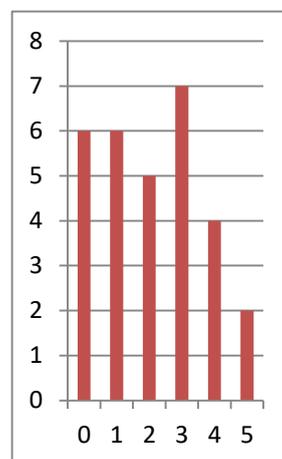


Diagrama de sectores:

Para dibujar el diagrama de sectores necesitamos realizar unas operaciones antes, en nuestro caso como los valores de la variable son seis, tenemos que realizar seis operaciones:

$$\frac{360^\circ}{n} = \frac{x_i^\circ}{n_i} \Rightarrow x_i^\circ = f_i \times 360^\circ = \frac{n_i}{n} 360^\circ$$

$$1^{\text{a}}. - \frac{360^{\circ}}{30} = \frac{x_1^{\circ}}{6} \Rightarrow x_1^{\circ} = \frac{6}{30} 360^{\circ} = 72^{\circ}$$

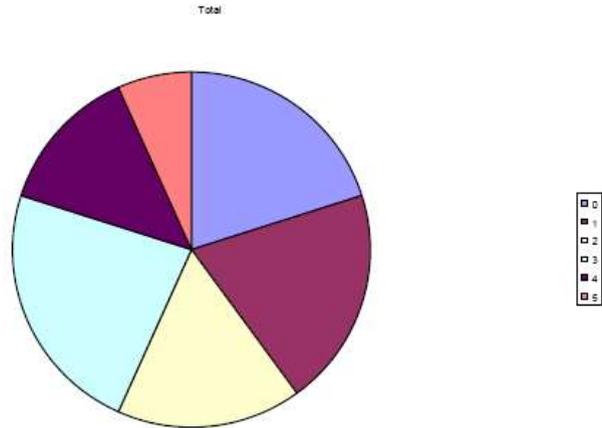
$$2^{\text{a}}. - \frac{360^{\circ}}{30} = \frac{x_2^{\circ}}{5} \Rightarrow x_2^{\circ} = \frac{6}{30} 360^{\circ} = 72^{\circ}$$

$$3^{\text{a}}. - \frac{360^{\circ}}{30} = \frac{x_3^{\circ}}{5} \Rightarrow x_3^{\circ} = \frac{5}{30} 360^{\circ} = 60^{\circ}$$

$$4^{\text{a}}. - \frac{360^{\circ}}{30} = \frac{x_4^{\circ}}{7} \Rightarrow x_4^{\circ} = \frac{7}{30} 360^{\circ} = 84^{\circ}$$

$$5^{\text{a}}. - \frac{360^{\circ}}{30} = \frac{x_5^{\circ}}{4} \Rightarrow x_5^{\circ} = \frac{4}{30} 360^{\circ} = 48^{\circ}$$

$$6^{\text{a}}. - \frac{360^{\circ}}{30} = \frac{x_6^{\circ}}{2} \Rightarrow x_6^{\circ} = \frac{2}{30} 360^{\circ} = 24^{\circ}$$



EJEMPLO 2. Se ha realizado una encuesta obteniendo los siguientes resultados:

33, 65, 34, 71, 46, 57, 39, 50, 70, 66, 54, 38, 57, 48, 39, 69, 54, 37, 46, 62, 37, 46, 55, 73, 36, 64, 53, 47, 53, 43

Agrupando los datos en intervalos de amplitud 8, calcula la media, la moda, la mediana, el recorrido, la varianza y la desviación típica. Realiza el histograma, el polígono de frecuencias y el diagrama de sectores de los datos anteriores.

En primer lugar se construye la tabla de frecuencias con los intervalos...

| Intervalo | MARCA DE LA CLASE x_i | FRECUENCIAS ABSOLUTAS n_i | FRECUENCIAS ABSOLUTAS ACUMULADAS N_i | $x_i \cdot n_i$ | $x_i^2 \cdot n_i$ |
|-----------|----------------------------|-----------------------------|---|-----------------|-------------------|
| [33,41) | 37 | 8 | 8 | 296 | 10952 |
| [41,49) | 45 | 6 | 14 | 270 | 12150 |
| [49,57) | 53 | 6 | 20 | 318 | 16854 |
| [57,65) | 61 | 4 | 24 | 244 | 14884 |
| [65,73] | 69 | 6 | 30 | 414 | 28566 |
| | | 30 | | 1542 | 83406 |

Una vez realizada la tabla estoy en condiciones de calcular las medidas pedidas:

- $Media = \frac{1542}{30} = 51,4$
- Para calcular la mediana, divido 30 entre 2 que es 15, y busco en la columna de N_i el primer valor mayor o igual que 15, que es 20 en la tercera fila. Por tanto la mediana es el valor de la marca de la clase en esa fila, que es 53.
- Para calcular la moda, busco en la columna n_i el valor más alto, que es 8 en la primera fila. Por tanto es la marca de la clase en esa fila, 37.
- $S^2 = \frac{1}{30} \cdot 83406 - (51,4)^2 = 138,24$
- $S = \sqrt{138,24} = 11,76$
- $CV = \frac{11,76}{51,4} = 0,23$, es decir, el 23%

Histograma y polígono de frecuencias:

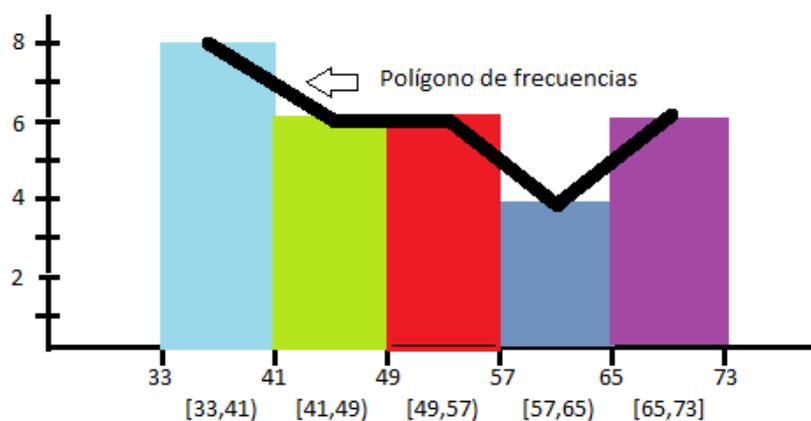


Diagrama de sectores:

Para dibujar el diagrama de sectores necesitamos realizar unas operaciones antes, en nuestro caso como los valores de la variable son seis, tenemos que realizar seis operaciones:



$$\frac{360^\circ}{n} = \frac{x_i^\circ}{n_i} \Rightarrow x_i^\circ = f_i \times 360^\circ = \frac{n_i}{n} 360^\circ$$

- 1) $x_1^\circ = 8 \times \frac{360^\circ}{30} = 96^\circ$
- 2) $x_2^\circ = 6 \times \frac{360^\circ}{30} = 72^\circ$
- 3) $x_3^\circ = 6 \times \frac{360^\circ}{30} = 72^\circ$
- 4) $x_4^\circ = 4 \times \frac{360^\circ}{30} = 48^\circ$
- 5) $x_5^\circ = 6 \times \frac{360^\circ}{30} = 72^\circ$

EJERCICIOS

1. ¿Cuál de las siguientes informaciones te parece claramente manipulada o errónea?

- Según un estudio estadístico, realizado a dos personas en un club náutico, se determina que a todos los españoles les encanta el buceo deportivo.
- Según un estudio estadístico, realizado por una compañía eléctrica, se sabe que los andaluces no aprecian que haya contaminación generada por las centrales térmicas en nuestro territorio.
- Un estudio estadístico determina que el cien por cien de los encuestados respiran cada día.
- Todas las opciones anteriores son estudios manipulados o sin sentido.

2. Se quiere conocer la cantidad de CO₂ que hay en el aire en una determinada población. ¿Cuál sería la opción más adecuada para llevar a cabo este estudio?

- Crear un cuestionario abierto preguntando por la cantidad de CO₂ que hay en el aire
- Crear un cuestionario cerrado con las respuestas: 20 mg/m³, 10 mg/m³ y otra cantidad.
- Instalar un aparato medidor en algún punto de la ciudad que registre los datos de cantidad de CO₂ que hay en el aire a lo largo de un periodo determinado de tiempo.

3. Indica si las siguientes variables aleatorias son cualitativas, cuantitativas discretas o cuantitativas continuas.

- Energía aportada por distintas marcas de muesli
- Sistema de calefacción utilizado en el invierno por familias de Madrid
- Volumen de basura generado por las familias de una barriada de Toledo
- Soluciones al problema de la contaminación de las aguas
- Número de coches por familia en la provincia de Málaga

4. Se quiere estudiar el nivel de contaminación del agua de un determinado río. Elige la opción más adecuada para elegir la muestra:

- Se cogería una muestra de agua al azar de cualquier zona del cauce del río.
- Se tomarían varias muestras de agua al azar de distintas zonas a lo largo del cauce del río y en distintos períodos de tiempo.
- Se tomaría una muestra de agua al lado de una fábrica que vierte sus residuos directamente al cauce del río.
- Se tomaría una muestra de agua en el lugar de nacimiento del río.

5. Estás realizando un estudio estadístico para conocer la satisfacción de la gente del barrio con el nuevo polideportivo. ¿Qué forma de elegir la muestra crees que es mejor?
- Preguntar a 50 personas que estén en el polideportivo.
 - Preguntar a 50 personas de tus amistades.
 - Elegir al azar 50 números de teléfono de casas del barrio, llamar y preguntar.
 - Preguntar a 50 personas que estén por la mañana comprando en el mercado.
6. En un determinado paraje se ha medido la altura de 10 olivos, siendo sus alturas 3,5 m; 3,8 m; 3,4 m; 3,1 m; 3,6 m; 3,8 m; 3 m; 3,7 m; 2,8 m; 3,3 m. La altura media de los diez olivos del paraje es de:
- 3 m
 - 3,4 m
 - 4 m
7. Se realiza una encuesta a 100 personas preguntando si separan o no los residuos para reciclarlos, siendo los resultados los recogidos en esta tabla:

|  | Nº de respuestas |
|---|------------------|
| Siempre, clasificando en las categorías: orgánica, vidrio, envases y papel. | 10 |
| Siempre, pero sólo papel y vidrio. | 15 |
| Casi siempre el papel | 23 |
| Casi siempre el vidrio | 18 |
| Normalmente no | 16 |
| Nunca | 10 |
| Otras opciones | 8 |

La Moda es:

- Casi siempre el papel.
- Siempre, clasificando en las categorías: orgánica, vidrio, envases y papel.
- Casi siempre el vidrio.

8. En una recogida de datos sobre los metros cuadrados ocupados por las distintas zonas verdes en dos localidades datos:

Localidad 1

| | m ² zona verde |
|--------|---------------------------|
| Zona1 | 780 |
| Zona2 | 1080 |
| Zona3 | 2200 |
| Zona4 | 2800 |
| Zona5 | 5600 |
| Zona6 | 950 |
| Zona7 | 4200 |
| Zona8 | 2600 |
| Zona9 | 4100 |
| Zona10 | 3500 |

Localidad 2

| | m ² zona verde |
|-------|---------------------------|
| Zona1 | 4500 |
| Zona2 | 600 |
| Zona3 | 1800 |
| Zona4 | 5400 |
| Zona5 | 1000 |
| Zona6 | 700 |
| Zona7 | 1900 |
| Zona8 | 6100 |

¿Cuál de las dos localidades presenta una distribución de zonas verdes más “dispersa”? (Haría falta calcular el coeficiente de variación de los metros cuadrados destinados a zona verde de ambas localidades)

- a. La localidad 1 b. La localidad 2 c. Ambas por igual.
9. Hemos consultado, en diferentes comercios, el precio (en euros) de un determinado modelo de impresora, obteniendo los datos siguientes:

146 - 150 - 141 - 143 - 139 - 144 - 133 - 153

- a) Calcula el precio medio.
 b) ¿Cuál es la mediana?
 c) Halla el recorrido.
 d) Halla la desviación típica.

10. En la familia Fernández, el salario mensual del padre es de 950 €, y el salario de la madre, 1600 €. En la familia Torres, el padre gana 1 800 € al mes, y la madre 750 €. ¿Cuál es el sueldo medio de cada familia? ¿En cuál de ellas es mayor la dispersión? ¿Cuál es el rango en cada familia?

11. En un control de velocidad en carretera se obtuvieron los siguientes datos:

| VELOCIDAD (km/h) | N.º DE COCHES |
|------------------|---------------|
| 60 - 70 | 5 |
| 70 - 80 | 15 |
| 80 - 90 | 27 |
| 90 - 100 | 38 |
| 100 - 110 | 23 |
| 110 - 120 | 17 |

a) Haz una tabla reflejando las marcas de clase y las frecuencias.

b) Calcula la media y la desviación típica.

c) ¿Qué porcentaje circula a más de 90 km/h?

12. Los puntos conseguidos por Teresa y por Rosa en una semana de entrenamiento, jugando al baloncesto, han sido los siguientes:

| | | | | | | | |
|--------|----|----|----|----|----|----|----|
| TERESA | 16 | 25 | 20 | 24 | 22 | 29 | 18 |
| ROSA | 23 | 24 | 22 | 25 | 21 | 20 | 19 |

a) Halla la media de cada una de las dos.

b) Calcula la desviación típica y el coeficiente de variación. ¿Cuál de las dos es más regular?

13. A la pregunta: ¿cuántas personas forman tu hogar familiar?, 40 personas respondieron esto:

4 5 3 6 3 5 4 6 3 2 2 4 6 3 5 3 4 5 3 6
4 5 7 4 6 2 3 4 4 3 4 4 5 3 2 6 3 7 4 3

a) Haz la tabla de frecuencias y el diagrama correspondiente.

b) Calcula la media, la mediana, la moda y la desviación típica.

En un test de inteligencia realizado a una muestra de 200 personas, se han obtenido los resultados siguientes:

a) Dibuja un histograma para representar gráficamente los datos y haz también el polígono de frecuencias.

b) Calcula la media y la desviación típica.

| PUNTUACION | N.º DE PERSONAS |
|------------|-----------------|
| 30 - 40 | 6 |
| 40 - 50 | 18 |
| 50 - 60 | 76 |
| 60 - 70 | 70 |
| 70 - 80 | 22 |
| 80 - 90 | 8 |

14. Al medir el peso al nacer en una determinada especie de animales, hemos obtenido los datos siguientes:

| PESO (kg) | N.º ANIMALES |
|-----------|--------------|
| 3,5 - 4,5 | 1 |
| 4,5 - 5,5 | 8 |
| 5,5 - 6,5 | 28 |
| 6,5 - 7,5 | 26 |
| 7,5 - 8,5 | 16 |
| 8,5 - 9,5 | 1 |

- a) Representa estos datos con el gráfico adecuado.
- b) Calcula la media y la desviación típica.
- c) ¿Qué porcentaje de animales pesó entre 5,5 kg y 6,5 kg? ¿Y entre 4,5 kg y 8,5 kg?

15. Estas son las horas de estudio semanal de un grupo de alumnas y alumnos:

14 9 9 20 18 12 14 6 14 8
15 10 18 20 2 7 18 8 12 10
20 16 18 15 24 10 12 25 24 17
10 4 8 20 10 12 16 5 4 13

a) Reparte estos datos en los intervalos:

1,5-6,5 6,5-11,5 11,5-16,5 16,5-21,5 21,5-26,5

b) Haz la tabla de frecuencias y el histograma.

b) Calcula la media y la desviación típica.

16. Los gastos mensuales de una empresa A tienen una media de 60 000 € y una desviación típica de 7 500 €. En otra empresa más pequeña B, la media es 9 000 €, y la desviación típica, 1 500 €. Calcula, mediante el coeficiente de variación, cuál de las dos tiene más variación relativa.

18. Al observar las notas de un mismo examen en dos grupos de tercero de ESO, se comprueba que en una clase hay siete personas que han tenido un 1 y cinco que han tenido un 10, mientras que en otra hay sólo 2 personas que han sacado un 1 y tres que han tenido un 10. Sabemos, además, las medias y desviación típica en cada una de las clases, que son:

| | NOTA MEDIA | DESVIACIÓN TÍPICA |
|-----------|------------|-------------------|
| Tercero A | 5,43 | 3,01 |
| Tercero B | 5,56 | 1,35 |

¿En qué clase hay mayor número de dieces, en Tercero A o en Tercero B?

19. El servicio de urgencias de un centro de salud ha atendido en los últimos 20 días, en horario de 0:00 horas a 8:00 horas, las siguientes urgencias:

2, 3, 1, 0, 2, 4, 5, 4, 1, 2, 1, 0, 2, 1, 3, 4, 5, 4, 2 y 2

- Construya la tabla de frecuencias de la distribución
- Determine moda, mediana y media aritmética de la distribución.
- Calcula la varianza, la desviación típica.

20. El porcentaje de población activa dedicada a la agricultura en 30 países africanos es:

| | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|
| 47 | 24 | 70 | 63 | 91 | 61 | 63 | 75 | 56 | 57 |
| 68 | 74 | 77 | 69 | 68 | 70 | 75 | 64 | 37 | 36 |
| 65 | 91 | 62 | 14 | 66 | 81 | 24 | 66 | 63 | 43 |

- Agrupar estos datos en cinco intervalos de igual amplitud
- Calcula la media, moda y mediana
- Calcula la varianza, la desviación típica y el coeficiente de variación.

21. La edad de los asistentes a dos congresos se distribuye según esta tabla:

| Edad | [28,34] | (34,40] | (40,46] | (46,52] | (52,58] | (58,64] |
|------------|---------|---------|---------|---------|---------|---------|
| Congreso A | 10 | 20 | 30 | 40 | 30 | 20 |
| Congreso B | 30 | 20 | 30 | 20 | 30 | 20 |

- Calcula la media de asistentes a cada uno de los congresos
- Calcula en cada caso la desviación típica.
- Comenta los resultados obtenidos en los apartados anteriores comparando la distribución de las edades de los asistentes a cada uno de los congresos.

22. Los expertos en baloncesto quieren hacer estudios comparativos sobre las estaturas de los jugadores de 1º división. Las estaturas de los jugadores de dos equipos (A y B) son:

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| EqA | 180 | 186 | 193 | 196 | 202 | 206 | 210 | 184 | 199 | 203 | 207 | 189 | 188 | 183 |
| EqB | 186 | 192 | 198 | 204 | 208 | 188 | 193 | 199 | 209 | 194 | 199 | 194 | 181 | 205 |

Compara, a partir de estos datos, la altura de los dos equipos, llevando a cabo las siguientes cuestiones, para cada equipo:

- Agrupar estos datos en seis intervalos de igual amplitud.
- Calcular la media, la moda y la mediana
- Calcular la varianza, desviación típica y coeficiente de variación

23. Dada la distribución estadística de la siguiente tabla:

| | | | | | | |
|-------|-------|--------|---------|---------|---------|---------|
| X_i | [0,5) | [5,10) | [10,15) | [15,20) | [20,25) | [25,30] |
| n_i | 5 | 7 | 9 | 10 | 4 | 7 |

- Calcular la media, moda, y mediana
- Hallar la varianza y la desviación típica

EJERCICIOS RESUELTOS

19 Estas son las horas de estudio semanal de un grupo de alumnas y alumnos:

14 9 9 20 18 12 14 6 14 8
 15 10 18 20 2 7 18 8 12 10
 20 16 18 15 24 10 12 25 24 17
 10 4 8 20 10 12 16 5 4 13

a) Reparte estos datos en los intervalos:

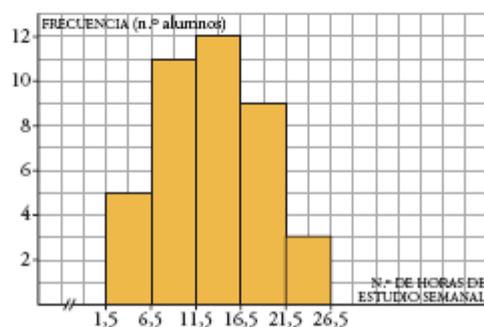
1,5-6,5; 6,5-11,5; 11,5-16,5; 16,5-21,5; 21,5-26,5

Haz la tabla de frecuencias y el histograma.

b) Calcula la media y la desviación típica.

a)

| INTERVALO | FRECUENCIA |
|-----------|------------|
| 1,5-6,5 | 5 |
| 6,5-11,5 | 11 |
| 11,5-16,5 | 12 |
| 16,5-21,5 | 9 |
| 21,5-26,5 | 3 |



b)

| INTERVALO | x_i | f_i | $f_i \cdot x_i$ | $f_i \cdot x_i^2$ |
|-----------|-------|-------|-----------------|-------------------|
| 1,5-6,5 | 4 | 5 | 20 | 80 |
| 6,5-11,5 | 9 | 11 | 99 | 891 |
| 11,5-16,5 | 14 | 12 | 168 | 2 352 |
| 16,5-21,5 | 19 | 9 | 171 | 3 249 |
| 21,5-26,5 | 24 | 3 | 72 | 1 728 |
| | | 40 | 530 | 8 300 |

$$\bar{x} = \frac{530}{40} = 13,25 \text{ h}; \quad \sigma = \sqrt{\frac{8\,300}{40} - (13,25)^2} = 5,6513$$

20 Los gastos mensuales de una empresa A tienen una media de 60 000 € y una desviación típica de 7 500 €. En otra empresa más pequeña B, la media es 9 000 €, y la desviación típica, 1 500 €. Calcula, mediante el coeficiente de variación, cuál de las dos tiene más variación relativa.

$$\text{Empresa A: } CV_A = \frac{\sigma_A}{\bar{x}_A} = \frac{7\,500 \text{ €}}{60\,000 \text{ €}} = 0,125 \rightarrow 12,5\%$$

$$\text{Empresa B: } CV_B = \frac{\sigma_B}{\bar{x}_B} = \frac{1\,500 \text{ €}}{9\,000 \text{ €}} = 0,167 \rightarrow 16,7\%$$

Tiene mayor variación relativa la empresa B.

21 El número de palabras de cada una de las frases de un artículo de economía es:

17 40 22 25 43 21 17 25 37 12
 9 37 32 35 30 21 13 27 41 45
 36 40 30 48 45 41 39 39 40 38
 28 7 33 35 22 34 23

a) Haz una tabla de frecuencias agrupando los datos en los intervalos:

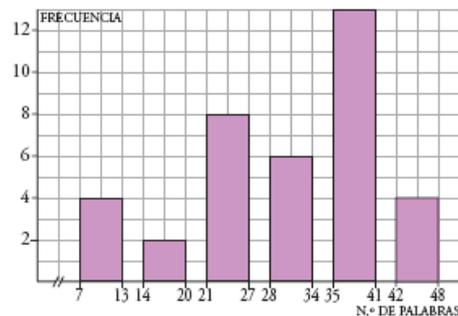
7 - 13, 14 - 20, 21 - 27, 28 - 34, 35 - 41, 42 - 48

Representa estos datos en un histograma.

b) Calcula su media y su desviación típica.

a)

| INTERVALO | FRECUENCIA |
|-----------|------------|
| 7-13 | 4 |
| 14-20 | 2 |
| 21-27 | 8 |
| 28-34 | 6 |
| 35-41 | 13 |
| 42-48 | 4 |



b)

| INTERVALO | x_i | f_i | $f_i \cdot x_i$ | $f_i \cdot x_i^2$ |
|-----------|-------|-------|-----------------|-------------------|
| 7-13 | 10 | 4 | 40 | 400 |
| 14-20 | 17 | 2 | 34 | 578 |
| 21-27 | 24 | 8 | 192 | 4608 |
| 28-34 | 31 | 6 | 186 | 5766 |
| 35-41 | 38 | 13 | 494 | 18772 |
| 42-48 | 45 | 4 | 180 | 8100 |
| | | 37 | 1126 | 38224 |

$$\bar{x} = \frac{1126}{37} = 30,432; \quad \sigma = \sqrt{\frac{38224}{37} - (30,432)^2} = 10,343$$

- 22  Se ha estudiado el grupo sanguíneo de 200 personas, así como el Rh. Algunos de los resultados vienen en la tabla siguiente, aunque nos la han dado incompleta:

| | GRUPO A | GRUPO B | GRUPO AB | GRUPO O | TOTALES |
|---------|---------|---------|----------|---------|---------|
| RH + | 74 | | 6 | 70 | 162 |
| RH - | | 3 | 1 | | |
| TOTALES | | | | 86 | 200 |

- Una tabla de este tipo se llama *tabla de contingencia*. Complétala.
- ¿Qué porcentaje de la población estudiada tiene el grupo B con Rh+?
- ¿Cuál es el porcentaje de la población estudiada que tiene Rh-?
- Se llama *donante universal* al grupo que puede dar sangre a todos los demás; es el O con Rh-. ¿Qué porcentaje de donantes universales hay en la población estudiada?
- Dentro de los del Rh+, ¿qué porcentaje tienen el grupo sanguíneo A?
- Haz un diagrama de sectores para los distintos grupos sanguíneos.

a)

| | GRUPO A | GRUPO B | GRUPO AB | GRUPO O | TOTALES |
|---------|---------|---------|----------|---------|---------|
| RH + | 74 | 12 | 6 | 70 | 162 |
| RH - | 18 | 3 | 1 | 16 | 38 |
| TOTALES | 92 | 15 | 7 | 86 | 200 |

b) $\frac{12 \cdot 100}{200} = 6\%$ tienen grupo B y Rh+.

c) $\frac{38 \cdot 100}{200} = 19\%$ tiene Rh-.

d) $\frac{16 \cdot 100}{200} = 8\%$ de donantes universales.

e) $\frac{74 \cdot 100}{200} = 45,68\%$ tiene grupo A dentro de los de Rh+.

f) Grupo A: $\frac{92 \cdot 360}{200} = 165,6 \rightarrow 165^\circ 36'$

Grupo B: $\frac{15 \cdot 360}{200} = 27 \rightarrow 27^\circ$

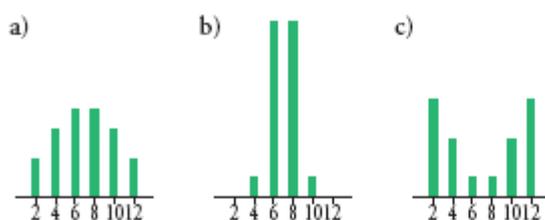
Grupo AB: $\frac{7 \cdot 360}{200} = 12,6 \rightarrow 12^\circ 36'$

Grupo O: $\frac{86 \cdot 360}{200} = 154,8 \rightarrow 154^\circ 48'$



REFLEXIONA SOBRE LA TEORÍA

23 Estas tres distribuciones tienen la misma media. ¿Cuál es?



Sus desviaciones típicas son 3,8; 1,3 y 2,9. Asocia a cada distribución uno de estos valores. (Hazlo observando las gráficas, sin hacer cuentas).

- Media = 7
- Desviaciones típicas:

a) 2,9

b) 1,3

c) 3,8

- 24 ■ ■ ■ Se ha pasado una prueba de 25 preguntas a los 120 estudiantes de un centro escolar. Los resultados obtenidos se recogen en la tabla siguiente:

| N.º DE ACIERTOS | PORCENTAJE |
|-----------------|------------|
| 5 | 10% |
| 15 | 45% |
| 20 | 25% |
| 25 | |

- a) Calcula el número de alumnos que respondió correctamente a todas las preguntas.
 b) Halla la media de aciertos de la población.
 c) Calcula la desviación típica.

- a) El tanto por ciento de alumnos que respondió correctamente a todas las preguntas es: $100 - (10 + 45 + 25) = 20\%$

El número de alumnos que respondieron correctamente a toda la prueba es:

$$\frac{20}{100} \cdot 120 = 24$$

b) $\frac{10}{100} \cdot 120 = 12$ $\frac{45}{100} \cdot 120 = 54$ $\frac{25}{100} \cdot 120 = 30$

| x_i | f_i | $f_i \cdot x_i$ | $f_i \cdot x_i^2$ |
|-------|-------|-----------------|-------------------|
| 5 | 12 | 60 | 300 |
| 15 | 54 | 810 | 12150 |
| 20 | 30 | 600 | 12000 |
| 25 | 24 | 600 | 15000 |
| | 120 | 2070 | 39450 |

$$\bar{x} = \frac{2070}{120} = 17,25$$

c) $\sigma = \sqrt{\frac{39450}{120} - (17,25)^2} = 5,58$

- 12 ■ ■ ■ Contando el número de erratas por página en un libro concreto, David ha obtenido los datos siguientes:

| N.º ERRATAS (x_i) | 0 | 1 | 2 | 3 | 4 | 5 |
|-----------------------|----|----|----|---|---|---|
| N.º PÁGINAS (f_i) | 50 | 40 | 16 | 9 | 3 | 2 |

- a) Halla la media y la desviación típica.
b) ¿Cuál es la moda?

a)

| x_i | f_i | $f_i \cdot x_i$ | $f_i \cdot x_i^2$ |
|-------|-------|-----------------|-------------------|
| 0 | 50 | 0 | 0 |
| 1 | 40 | 40 | 40 |
| 2 | 16 | 32 | 64 |
| 3 | 9 | 27 | 81 |
| 4 | 3 | 12 | 48 |
| 5 | 2 | 10 | 50 |
| | 120 | 121 | 283 |

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{121}{120} = 1,008$$

$$\sigma = \sqrt{\frac{\sum f_i \cdot x_i^2}{\sum f_i} - (\bar{x})^2} = \sqrt{\frac{283}{120} - (1,008)^2} = 1,159$$

- b) $Mo = 0$ erratas (Es el valor con mayor frecuencia)