



1 Introducción

La **Estadística** es la ciencia que se encarga de recoger, organizar, describir e interpretar datos referidos a distintos fenómenos para, posteriormente, analizarlos e interpretarlos.

La Estadística tiene mucha importancia en la actualidad, fundamentalmente por la presencia en diversos ámbitos de la vida cotidiana y por la contribución que ejerce en otras materias, como por ejemplo la Medicina, la Biología o la Economía.

La gran utilización del lenguaje y técnicas estadísticas en medios de comunicación, política, sanidad, economía, etc, han convertido en básica la necesidad de interpretar y analizar los datos para comprender la realidad que nos rodea. Es importante ser críticos con las informaciones estadísticas, detectar muestras sesgadas y resultados no representativos que se pueden encontrar en cualquier sector de la sociedad actual, por ejemplo las encuestas y sondeos electorales.

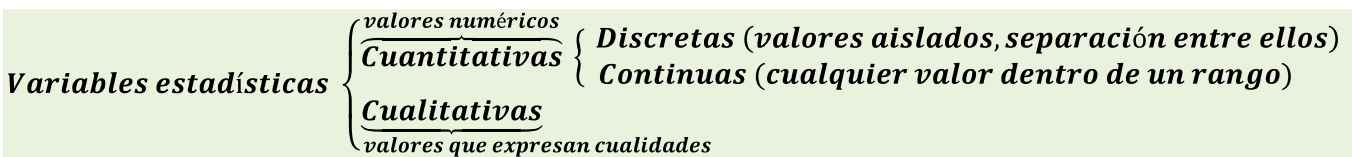
La presencia de la Estadística en multitud de ámbitos en la vida cotidiana en los países desarrollados, ha hecho que se considere como parte de la cultura matemática general que debería tener cualquier persona, como ciudadano y miembro de la sociedad. Es decir, los conocimientos estadísticos contribuyen en el desarrollo personal del individuo, no sólo le aporta conocimientos sino que la gran vinculación con la realidad le servirá a formarse como persona. Estos conocimientos le ayudarán a tomar decisiones en situaciones de incertidumbre y a ser crítico con los resultados que obtenga y encuentre en cualquier sector o situación a lo largo de su vida. Por ello, considero necesaria la educación estadística de forma efectiva para llegar a lograr desarrollar todas estas capacidades.

2 Conceptos básicos

- **Población:** es el conjunto de elementos que son objeto del estudio estadístico.
- **Muestra:** es la parte de la población que se estudia cuando ésta es muy grande; a partir de ella se deducen características de la población.
- **Individuo:** es cada uno de los elementos de la población o de la muestra.
- **Tamaño muestral:** es el número de individuos de la muestra.

3 Variable estadística

Una **variable estadística** es el conjunto de valores que puede tomar cada uno de los fenómenos estudiados a través de la Estadística. Las variables con las que trabajamos pueden ser cualitativas o cuantitativas y estas últimas, a su vez, discretas o continuas:



Ejercicio 1

Clasifica en variables cualitativas o cuantitativas discretas o continuas cada una de las características siguientes y decide si es más conveniente estudiar la población o la muestra:

- La marca de teléfono móvil que utilizan los alumnos de tu instituto.
- El número de hermanos que tiene cada uno de los alumnos de tu clase.
- El peso de los habitantes empadronados en el distrito de Retiro en Madrid.
- El tiempo dedicado al sueño diario de los adolescentes entre 14 y 16 años en una ciudad.
- El deporte preferido por los profesores de tu instituto.
- El número de ascensores de los hoteles de Madrid.

4 Tabla de frecuencias

Consideremos una **variable estadística** X con resultados posibles x_1, x_2, \dots, x_k y número de datos N .

• **Frecuencia absoluta** f_i de un dato x_i es el número de veces que se repite.

Es claro que la suma de las frecuencias absolutas coincide con el número total de datos N :

$$N = f_1 + f_2 + \dots + f_k = \sum_{i=1}^k f_i$$

• **Frecuencia relativa** h_i de un dato x_i es el cociente entre su frecuencia absoluta y el total de datos: $h_i = \frac{f_i}{N}$

La suma de las frecuencias relativas vale 1:

$$\sum_{i=1}^k h_i = h_1 + h_2 + \dots + h_k = \frac{f_1 + f_2 + \dots + f_k}{N} = \frac{N}{N} = 1$$

• El **porcentaje** $p_i\%$ de un dato x_i se calcula multiplicando su frecuencia relativa por 100: $p_i\% = h_i \cdot 100$

La suma de los porcentajes de todos los datos x_i vale 100:

$$\sum_{i=1}^k r_i = r_1 + r_2 + \dots + r_k = (h_1 + h_2 + \dots + h_k) \cdot 100 = 1 \cdot 100 = 100$$

• **Frecuencia absoluta acumulada** F_i de un dato x_i es la suma de las frecuencias absolutas de los valores que son menores o iguales que el dato x_i :

$$\text{Así, } \begin{cases} F_1 = f_1 \\ F_2 = f_1 + f_2 \\ \dots \dots \dots \dots \dots \\ F_i = f_1 + f_2 + \dots + f_i \end{cases}$$

• **Frecuencia relativa acumulada** H_i de un dato x_i es la suma de las frecuencias relativas de los valores que son menores o iguales que el dato x_i :

$$\text{Así, } \begin{cases} H_1 = h_1 \\ H_2 = h_1 + h_2 \\ \dots \dots \dots \dots \dots \\ H_i = h_1 + h_2 + \dots + h_i \end{cases}$$

Cuando el **número de datos** de una variable estadística X es **grande**, es conveniente organizarlos en una **tabla de distribución de frecuencias** donde los datos figuran en la primera columna y su frecuencia absoluta en la segunda.

Ejercicio 2

Considera la variable estadística X : 1, 3, 2, 1, 4, 5, 6, 4, 7, 6, 5, 4, 8, 6, 9, 6, 5, 4, 5, 1

Construye la tabla de frecuencias absolutas y relativas mediante el recuento de datos.

| Valores de la variable X: x_i | Frecuencia absoluta: f_i | Frecuencia relativa: h_i |
|------------------------------------|-------------------------------|-------------------------------|
| $x_1 = 1$ | $f_1 = 3$ | $h_1 = 3/20 = 0,15$ |
| 2 | | |
| 3 | | |
| 4 | | |
| 5 | | |
| 6 | | |
| 7 | | |
| 8 | | |
| 9 | | |
| Totales | $N = \sum f_i = 20$ | $\sum h_i = 1$ |

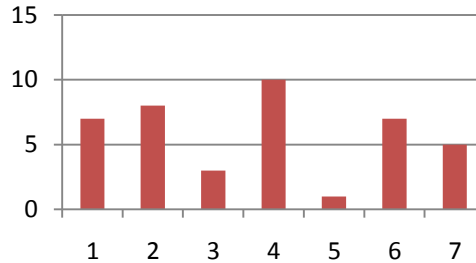
5 Gráficos estadísticos

A través de los **gráficos estadísticos** podemos observar de forma inmediata características importantes de los datos que se estudian.

Nosotros vamos a ver los siguientes:

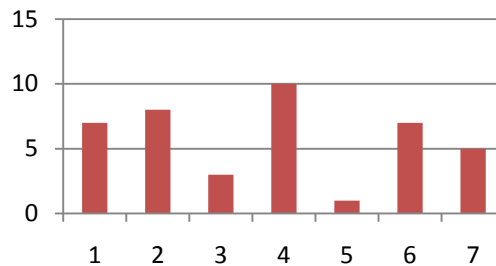
5.1 Diagrama de barras

Utiliza un sistema de ejes; en el horizontal se representan los valores de la variable estadística y en el vertical su frecuencia mediante una barra de altura proporcional a la misma.



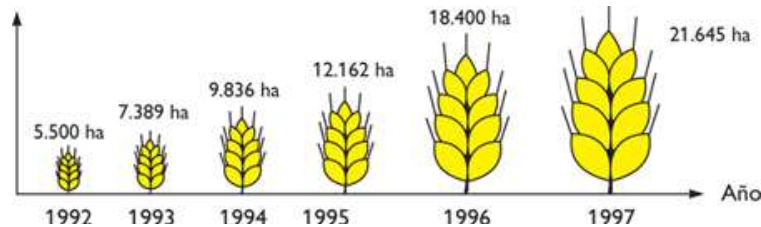
5.2 Polígono de frecuencias

Se forma a partir del diagrama de barras, uniendo mediante líneas los extremos superiores de las barras. **Constrúyelo en el siguiente diagrama:**



5.3 Pictograma

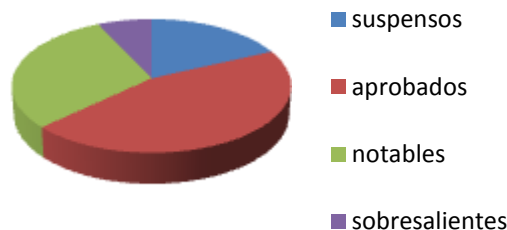
Utiliza dibujos representativos de los datos y tamaño proporcional a la frecuencia de éstos.



5.4 Diagrama de sectores

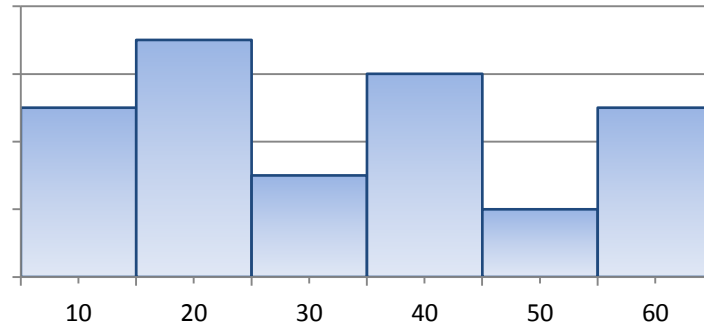
Se puede utilizar para todo tipo de variables, pero se usa frecuentemente para las variables cualitativas. Los datos se representan en un círculo, de modo que el ángulo de cada sector es proporcional a la frecuencia absoluta correspondiente.

El diagrama circular se construye con la ayuda de un transportador de ángulos. La amplitud α° , en grados sexagesimales, de cada sector viene dada por $\alpha^\circ = \frac{f_i}{N} \cdot 360^\circ = h_i \cdot 360^\circ$



5.5 Histograma

Es una representación gráfica de una variable en forma de barras, donde la superficie de cada barra es proporcional a la frecuencia de los valores representados. Sirven para obtener una "primera vista" general, o panorama, de la distribución de la población, o la muestra, respecto a una característica, cuantitativa y generalmente continua.



6 Medidas de centralización

Para hacer un estudio estadístico se utilizan las llamadas medidas estadísticas, de centralización y dispersión. En este curso, vamos a estudiar las medidas de centralización.

Las **medidas de centralización** son ciertos valores numéricos o parámetros que tienen a situarse, en general, hacia el centro del conjunto de datos ordenados, y dan una idea del comportamiento de la distribución. Tratan de resumir la información de la muestra de datos.

Los valores más utilizados son la **media**, la **moda** y la **mediana**.

6.1 Media aritmética

Si los posibles resultados de una variable estadística X de N datos son:

$x_i: x_1, x_2, x_3, \dots, x_k$ con frecuencias absolutas $f_i: f_1, f_2, f_3, \dots, f_k$ respectivamente, llamaremos **media aritmética**, y la representaremos como \bar{x} , al cociente:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_k f_k}{f_1 + f_2 + \dots + f_k} = \frac{\sum_{i=1}^k x_i f_i}{\sum_{i=1}^k f_i} = \frac{\sum_{i=1}^k x_i f_i}{N}$$

es decir, la media aritmética es el cociente entre la suma de todos los valores multiplicados por su frecuencia absoluta y el número total de datos.

Ejercicio 5

Calcula la media en la serie estadística 2, 2, 3, 5, 6, 7, 10, 12

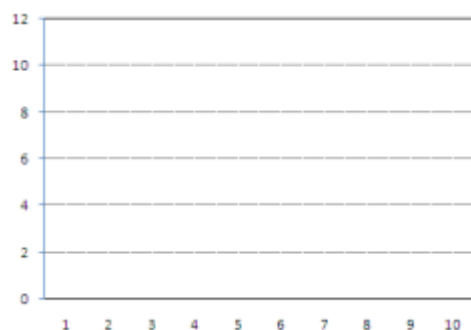
Al tratarse de pocos datos la media se obtiene dividiendo por el número de datos la suma de los valores de la serie, esto es:

$$\text{media} = \frac{47}{8} = 5,875$$

Ejercicio 6

Considera la variable estadística X : 1, 3, 2, 1, 4, 5, 6, 4, 7, 6, 5, 4, 8, 6, 9, 6, 5, 4, 5, 1

a) Dibuja el diagrama de barras que representa los datos.



b) Calcula la media formando primero la tabla de frecuencias ampliada con la columna del producto $x_i f_i$ que aparece en el numerador de la media:

Media: $\bar{x} = \frac{\sum x_i \cdot f_i}{N} = \frac{\quad}{\quad} = 4,6$

| Datos: x_i | f_i | $x_i f_i$ |
|------------------|-------|------------------|
| 1 | | |
| 2 | | |
| 3 | | |
| 4 | | |
| 5 | 4 | 20 |
| 6 | | |
| 7 | | |
| 8 | | |
| 9 | | |
| $N = \sum f_i =$ | | $\sum x_i f_i =$ |

Ejercicio 7: Datos agrupados en intervalos

Las edades de los trabajadores de una fábrica se muestran en la siguiente tabla, donde X son las edades y F las frecuencias absolutas:

| X | [19,25) | [25, 31) | [31, 37) | [37, 43) | 43, 49) |
|---|---------|----------|----------|----------|---------|
| F | 30 | 40 | 50 | 30 | 10 |

- a) Calcula la edad media de los trabajadores.
- b) Calcula la columna de porcentajes y di que porcentaje de trabajadores tienen más de 37 años.
- c) Dibuja el histograma de los datos.

Cuando los datos se presentan **agrupados por intervalos**, como en este ejemplo, procederemos como mostramos a continuación:

Sabemos que hay 30 individuos en el intervalo [19,25); no conocemos la edad de cada uno de ellos; así que convenimos en que las 30 personas tiene la edad que resulte de hallar la **marca de clase** del intervalo.

La **marca de clase** de un intervalo $[a, b]$ viene dada por la media de sus extremos:

| | |
|--|---------------------|
| Marca de clase del intervalo $[a, b]$: | $x = \frac{a+b}{2}$ |
|--|---------------------|

Así, en nuestro ejemplo, la **marca de clase** del primer intervalo **[19,25)** es $x_1 = \frac{19+25}{2} = \frac{44}{2} = 22$ años

A partir de ahora convenimos en que los 30 individuos, con edades en ese intervalo, tienen todos 22 años. Vamos a calcular la marca de clase de los otros intervalos y a completar la tabla de frecuencias correspondiente para responder a las preguntas:

La **marca de clase** del intervalo **[19,25)** es $x_1 = \frac{19+25}{2} = \frac{44}{2} = 22$ años

La **marca de clase** del intervalo **[25,31)** es $x_2 = \frac{25+31}{2} = \frac{56}{2} = 28$ años

La **marca de clase** del intervalo **[31,37)** es $x_3 = \frac{\quad}{2} = \frac{\quad}{2} = 34$ años

La **marca de clase** del intervalo **[37,43)** es $x_4 = \frac{\quad}{2} = \frac{\quad}{2} = 40$ años

La **marca de clase** del intervalo **[43,49)** es $x_5 = \frac{\quad}{2} = \frac{\quad}{2} = 46$ años

Construimos la tabla de frecuencias ampliada con la columna $x_i \cdot f_i$ y la de porcentajes:

| Intervalo | Marca de clase x_i | Frecuencia absoluta f_i | Frecuencia relativa h_i | Porcentaje $p_i\%$ | $x_i f_i$ |
|----------------|----------------------|---------------------------|---------------------------|--------------------|------------------|
| [19, 25] | 22 | 30 | 0,1875 | 18,75% | 660 |
| [25,31] | 28 | 40 | 0,25 | 25% | 1120 |
| [31, 37] | 34 | 50 | | | |
| [37, 43] | 40 | 30 | | | |
| [43, 49] | 46 | 10 | | | |
| Totales | | $N = \sum f_i =$ | 1 | 100% | $\sum x_i f_i =$ |

Cálculos

Número total de datos: $N = \sum f_i = 30 + 40 + 50 + 30 + 10 = 160$

$$h_1 = \frac{f_1}{N} = \frac{30}{160} = 0,1875 \quad ; \quad p_1 = h_1 \cdot 100 = 18,75\% \quad ; \quad x_1 \cdot f_1 = 22 \cdot 30 = 660$$

$$h_2 = \frac{f_2}{N} = \frac{40}{160} = 0,25 \quad ; \quad p_2 = h_2 \cdot 100 = 25\% \quad ; \quad x_2 \cdot f_2 = 28 \cdot 40 = 1120$$

$$h_3 = \frac{f_3}{N} = \text{---} = \boxed{} \quad ; \quad p_3 = h_3 \cdot 100 = \boxed{} \quad ; \quad x_3 \cdot f_3 = 34 \cdot 50 = \boxed{}$$

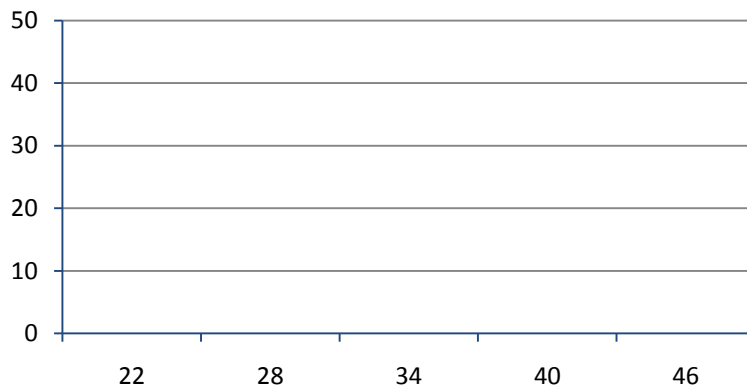
$$h_4 = \frac{f_4}{N} = \text{---} = \boxed{} \quad ; \quad p_4 = h_4 \cdot 100 = \boxed{} \quad ; \quad x_4 \cdot f_4 = 40 \cdot 30 = \boxed{}$$

$$h_5 = \frac{f_5}{N} = \text{---} = \boxed{} \quad ; \quad p_5 = h_5 \cdot 100 = \boxed{} \quad ; \quad x_5 \cdot f_5 = 46 \cdot 10 = \boxed{}$$

a) La edad media es $\bar{x} = \frac{\sum x_i f_i}{N} = \frac{5140}{160} = 32,125$ años

b) Porcentaje de trabajadores con más de 37 años:

c) Histograma



6.2 Moda

La **moda** de una variable estadística es el valor de la variable que presenta mayor frecuencia absoluta. Se representa por M_o . No es tan sensible como la media aritmética a valores extremos.

Puede ocurrir que la moda no sea única.

En el caso de una variable continua hablaremos de intervalo modal.

Ejercicio 8

En las siguientes calificaciones de Matemáticas de 40 alumnos determina la moda.

| | | | | | | | | | |
|-----------------------|---|---|---|---|---|---|---|---|---|
| Calificaciones: x_i | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Nº de alumnos: f_i | 2 | 2 | 4 | 5 | 8 | 9 | 3 | 4 | 3 |

La moda es $M_o = \boxed{}$ ya que es la calificación que ha obtenido un mayor número de alumnos.

La distribución es unimodal.

6.3 Mediana

Llamaremos **mediana** al número que, una vez ordenada la variable estadística de menor a mayor con datos repetidos incluidos, está en la posición central. Es decir, es el valor de la variable tal que el número de observaciones menores que él es igual al número de observaciones mayores que él.

Se representa por **Me**.

Si el número de datos fuera par, tomamos como mediana la media aritmética de los valores centrales y si es impar, el valor central.

Cuando el número de datos es muy grande no es fácil ordenar los datos; en este caso se usan las frecuencias acumuladas F_i .

Ejercicio 9

En la siguiente serie estadística 3, 6, 2, 9, 5, 12, 11, calcula la mediana

Ordenamos los datos de menor a mayor: 2, 3, 5, 6, 9, 11, 12

Como el número de datos es impar la mediana es $Me = \boxed{}$

Ejercicio 10

En la siguiente serie estadística 6, 5, 9, 3, 2, 3, 2, 13, 11, 12, calcula la mediana

Ordenamos los datos de menor a mayor: $\overbrace{2, 2, 3, 3}^4, 5, 6, \overbrace{9, 11, 12, 13}^4$

Como el número de datos es par la mediana es la media de los valores centrales: $Me = \frac{}{2} = \frac{}{2} = 4,5$

Ejercicio 11: Cálculo de la mediana con frecuencias acumuladas

El número de horas diarias que dedican al estudio un grupo de 30 alumnos es

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

Calcula la mediana

Organizamos los datos en una tabla de frecuencias con las columnas que se muestran y en la que ya aparecen algunos resultados. Complétala haciendo los cálculos necesarios.

Cálculos

| Nº de horas. x_i | f_i | h_i | F_i | H_i |
|--------------------|-------|-------|-------|-------|
| 0 | | | | |
| 1 | | 0.267 | | 0.367 |
| 2 | | 0.233 | | 0.6 |
| 3 | | | 24 | |
| 4 | | | | |
| 5 | | | 30 | 1 |

1. ¿Qué quiere decir $F_4 = 24$?

2. ¿Qué significa $H_3 = 0.6$

Tenemos un número grande de datos que es pesado ordenar de menor a mayor, vamos a utilizar las frecuencias acumuladas para hallar la mediana.

La mitad del número de datos es 15. Como teníamos un número par de datos, no existe un valor central, tenemos que buscar los que ocupan los lugares centrales, 15 y 16. Observando las frecuencias acumuladas de la tabla vemos que $F_2 = 11$ y $F_3 = 18$.

Es decir, ordenados los datos de menor a mayor, el último 1 ocupa la posición 11 y de la posición 12 a la 18 hay doses (2). De manera que las posiciones centrales están ocupadas por 2. Así la mediana será la media de ambos, es decir $Me = \frac{2+2}{2} = 2$