



ESTADÍSTICA

Es la parte de las matemáticas que se dedica al estudio de cómo recoger, ordenar y analizar datos que han sido obtenidos de fenómenos o situaciones concretas.

¿Qué vamos a estudiar?

Gustos musicales, partido político al que se vota, consumo de alcohol, número de aprobados, peso...

Podemos dividir las variables en dos tipos:

QUALITATIVAS

Elaborado por Armando García de Álvaro

CUANTITATIVAS



VARIABLES ESTADÍSTICAS

➤ CUALITATIVA

- NO SE PUEDE EXPRESAR CON NÚMEROS
- COLOR DE OJOS, GRUPO DE MÚSICA FAVORITO, RELIGIÓN QUE SE PRACTICA, ASIGNATURA FAVORITA

➤ CUANTITATIVA

- SE EXPRESA CON NÚMEROS
- NÚMERO DE HIJOS, EDAD, ALTURA, PESO, NÚMERO ASIGNATURAS SUSPENSAS

¿Dónde lo queremos estudiar?
¿De dónde cogemos los datos?



➤ POBLACIÓN

- Es la totalidad de individuos que queremos estudiar.
- No tienen por qué ser personas
- Ciudadanos de San Martín de Valdeiglesias, bombillas de una fábrica

Elaborado por Armando García de Álvaro

➤ MUESTRA

- Es la parte de la población de la que tomamos los datos.
- Una muestra debe ser **proporcional** y **representativa**

¿Qué hacemos con los datos obtenidos?

- ▶ Los datos obtenidos se organizan en tablas de frecuencias.
- ▶ Se pregunta a los 20 alumnos de una clase a qué edad tuvieron su primer móvil y se obtuvieron las siguientes respuestas:

10	9	12	13	10	11	12	10	10	9
14	12	10	11	7	9	10	12	12	10

10	9	12	13	10	11	12	10	10	9
14	12	10	11	7	9	10	12	12	10

Ordenamos los datos en una tabla de frecuencias.

x_i	f_i	h_i
7		
8		
9		
10		
11		
12		
13		
14		

En la primera columna x_i , colocamos los datos obtenidos de menor a mayor. Si algún número intermedio no aparece, también lo ponemos (aquí ocurre con el 8)

10	9	12	13	10	11	12	10	10	9
14	12	10	11	7	9	10	12	12	10

Ordenamos los datos de menor a mayor y contamos cuántas veces se repite cada dato.



x_i	f_i	h_i
7	1	
8	0	
9	3	
10	7	
11	2	
12	5	
13	1	
14	1	

La segunda columna es la **frecuencia absoluta**, que es el número de veces que se repite cada dato. (Si algún dato no aparecía f_i es cero)

10	9	12	13	10	11	12	10	10	9
14	12	10	11	7	9	10	12	12	10

Cada valor de la frecuencia absoluta lo dividimos por el número total de datos (en este caso 20 datos) y tenemos la frecuencia relativa.

$$1:20=0,05 \quad 0:20=0 \quad 3:20=0,15 \quad 7:20=0,35 \quad 2:20=0,1 \quad 1:20=0,05$$

x_i	f_i	h_i
7	1	0,05
8	0	0
9	3	0,15
10	7	0,35
11	2	0,1
12	5	0,25
13	1	0,05
14	1	0,05

La tercera columna es la **frecuencia relativa (h_i)**, si la multiplicamos por 100 obtenemos el porcentaje de cada dato.

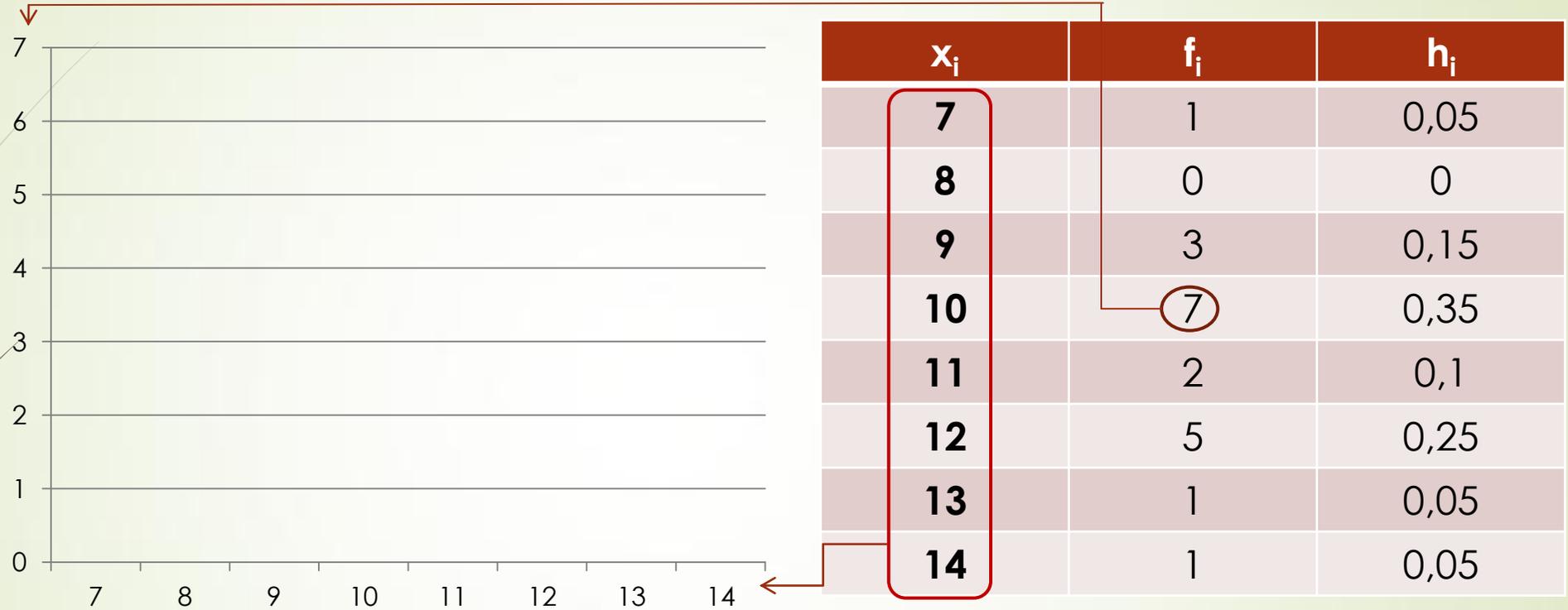
Al 35% de los alumnos de la clase les compraron su primer móvil a los 10 años.



¿Qué hacemos con los datos una vez organizados?

- La mejor forma de presentar o interpretar los datos es a través de un **diagrama de barras**.
- Un diagrama de barras es un gráfico estadístico en el que aparece **una barra vertical por cada dato** obtenido y **la altura** de cada barra **es proporcional a la frecuencia absoluta** de su dato.

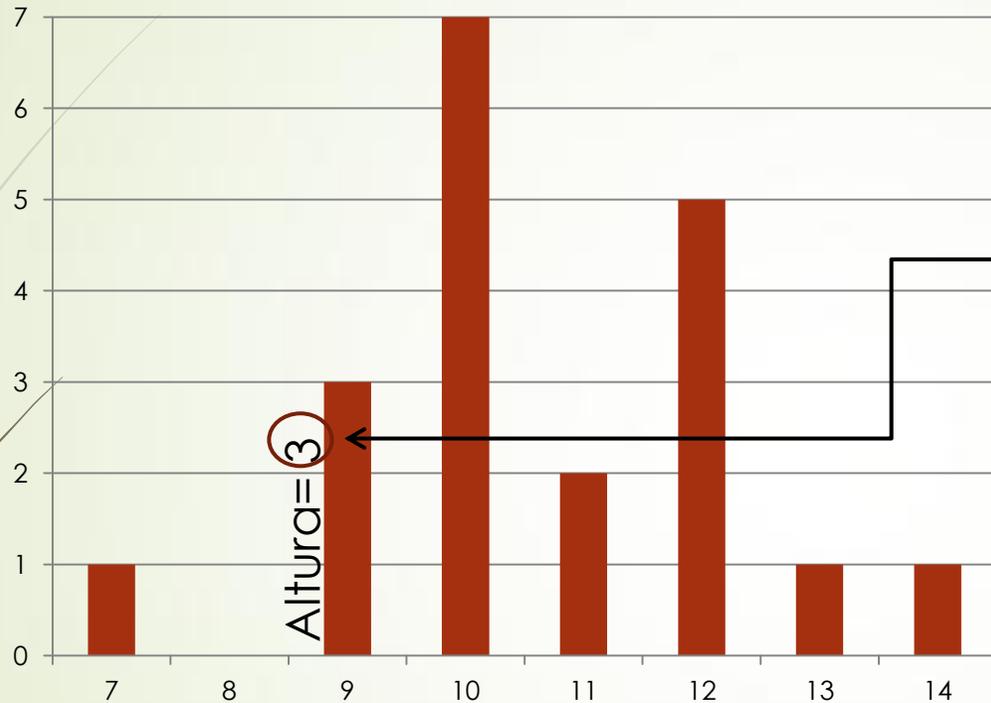
Diagrama de barras



Eje X: Escribimos los datos obtenidos de menor a mayor (columna x_i)

Eje Y: La altura del eje Y debe corresponder al valor más alto de la frecuencia absoluta y dividimos el eje en partes iguales.

Diagrama de barras



x_i	f_i	h_i
7	1	0,05
8	0	0
9	3	0,15
10	7	0,35
11	2	0,1
12	5	0,25
13	1	0,05
14	1	0,05

Para cada dato, hacemos una barra de altura igual a su frecuencia absoluta

Diagrama de barras



Por último siempre se debe acompañar un gráfico de su título y de la información que aparece en cada uno de los ejes.

Parámetros estadísticos

- Son una serie de números que se obtienen a partir de los datos obtenidos en un estudio estadístico.
- Son muy útiles para interpretar los datos.
- Podemos distinguir parámetros de centralización y de dispersión.





Parámetros de centralización

- Nos indican alrededor de qué valor se distribuyen los valores de la variable estadística observada (dónde está el centro de los datos)
- Son parámetros muy fáciles de comprender.
- Se usan para representar de forma global a toda la población. (La edad media de los usuarios de Instagram es 31.19 años)
- Son muy sencillos de calcular

Media aritmética

- ▶ Para obtener la media aritmética de una serie de datos, sumamos todos los datos y dividimos el resultado por el número total de datos
- ▶ Se representa por el símbolo \bar{x}
- ▶ Preguntamos a un alumno por las notas obtenidas en la segunda evaluación y nos dice los siguientes datos
- ▶ Es el parámetro estadístico que se usa más comúnmente

5	4	6	3	7	5	5	6
---	---	---	---	---	---	---	---

Sumamos todos los datos:

$$5+4+6+3+7+5+5+6=41$$

Dividimos el resultado por el número total de datos:

$$41:8=5,125$$

$$\bar{x} = 5,125$$

Mediana

- Es el dato central de todos los que se han obtenido (el que está en el medio).
- Si el número de datos es par no habrá un dato central, así que tomamos la media aritmética de los dos datos centrales.
- Para ver cuál es el dato central, solo hay que ordenar los datos de menor a mayor.
- Se representa por el símbolo Me
- Preguntamos a un alumno por las notas obtenidas en la segunda evaluación y nos dice los siguientes datos

5	4	6	3	7	5	5	6
---	---	---	---	---	---	---	---

Ordenamos todos los datos de menor a mayor:

3 4 5 **5 5** 6 6 7

Como tenemos ocho datos (par) los datos centrales son dos cincos, así que hacemos la media aritmética:

$$5+5=10 \ ; \ 10:2=5 \quad \boxed{Me=5}$$

Moda

- Es el dato que mas se repite de todos los que se han obtenido.
- Si no hay mucha diferencia entre un dato y el resto, no es un parámetro que nos aporte mucha información.
- Se representa por el símbolo M_o
- Preguntamos a un alumno por las notas obtenidas en la segunda evaluación y nos dice los siguientes datos

5	4	6	3	7	5	5	6
---	---	---	---	---	---	---	---

Ordenamos todos los datos de menor a mayor:

3 4 **5 5 5** 6 6 7

Vemos que el número que más se repite es el 5 que se repite tres veces.

La moda es 5 **$M_o=5$**

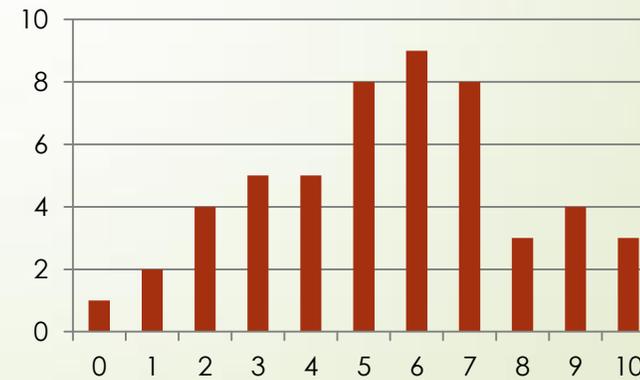
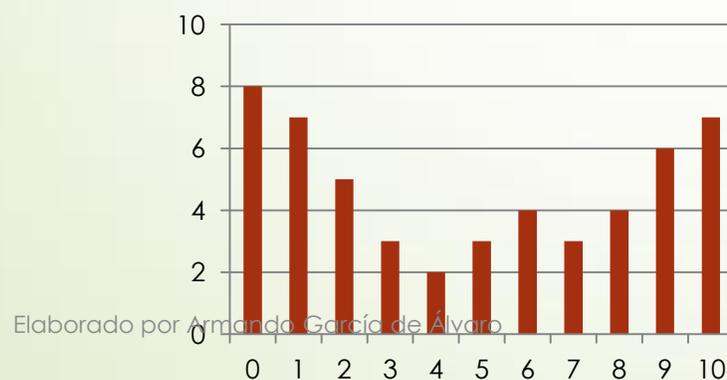


Parámetros de dispersión

- Nos indican si los datos obtenidos son cercanos a un valor central o están muy dispersos.
- Son parámetros más complicados de interpretar y de calcular que los parámetros de centralización.
- Se calculan respecto a la media aritmética (aunque también existen respecto a la mediana)
- Completan la información que nos dan los parámetros de centralización. Con la media y su parámetro de dispersión tenemos bastante información sobre los datos obtenidos.

Varianza

- ▶ Para obtener la varianza de una serie de datos seguimos los siguientes pasos (previamente hemos calculado la media aritmética):
 1. Elevamos cada dato al cuadrado
 2. Sumamos todos los valores obtenidos en 1
 3. Dividimos el valor obtenido en 2 por el número total de datos.
 4. Elevamos la media aritmética al cuadrado.
 5. La varianza es el resultado de restar el resultado de 3 menos el de 4
- ▶ Se representa por el símbolo σ^2
- ▶ No vamos a entrar en la interpretación numérica de la varianza, pero sí podemos comparar varios valores y asociarlos a datos o diagramas.



Varianza

- Preguntamos a un alumno por las notas obtenidas en la segunda evaluación y nos dice los siguientes datos (recordamos que la media vale 5,125)

5	4	6	3	7	5	5	6
---	---	---	---	---	---	---	---

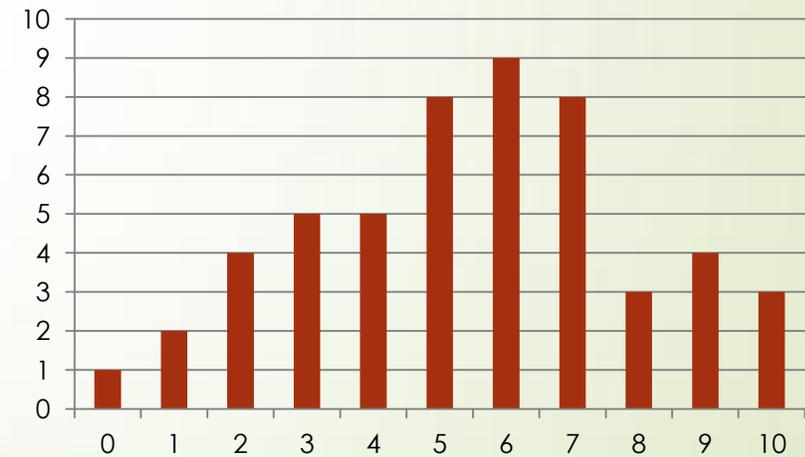
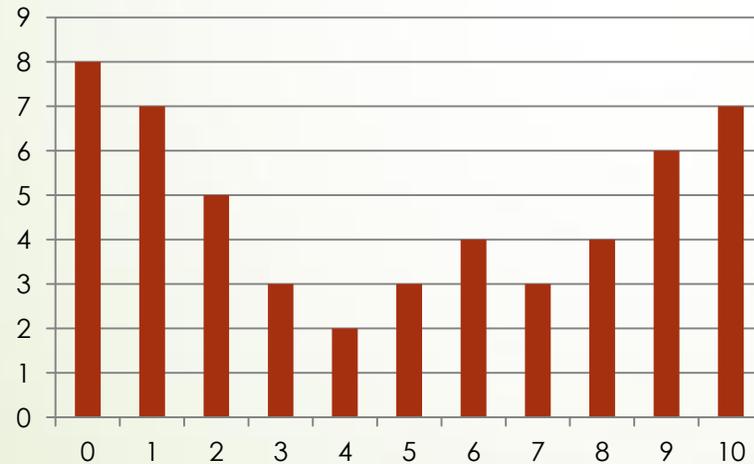
1. Elevamos cada dato al cuadrado
2. Sumamos todos los valores obtenidos en 1
3. Dividimos el valor obtenido en 2 por el número total de datos.
4. Elevamos la media aritmética al cuadrado.
5. La varianza es el resultado de restar el resultado de 3 menos el de 4

1. $5^2=25$ $4^2=16$ $6^2=36$ $3^2=9$ $7^2=49$ $5^2=25$ $5^2=25$ $6^2=36$
2. $25+16+36+9+49+25+25+36=221$
3. $221:8=27,625$
4. $5,125^2=26,266$
5. $27,625-26,266=1,359$

$$\sigma^2=1,359$$

Desviación típica

- Este parámetro es más adecuado que la varianza porque se expresa en las mismas unidades que la media.
- Para calcularlo solo tenemos que hacer al raíz cuadrada de la varianza.
- Se representa por el símbolo σ
- No vamos a entrar en la interpretación numérica de la desviación típica, pero sí podemos comparar varios valores y asociarlos a datos o diagramas.



Desviación típica

- Preguntamos a un alumno por las notas obtenidas en la segunda evaluación y nos dice los siguientes datos (recordamos que la media vale 5,125)

5	4	6	3	7	5	5	6
---	---	---	---	---	---	---	---

Después de haber calculado la varianza: $\sigma^2=1,359$ solo tenemos que hacer la raíz cuadrada del valor obtenido.

$$\sigma = \sqrt{1,359} = 1,166$$

$$\sigma = 1,166$$