

UNIDAD 7: ESTADÍSTICA

1.- CONCEPTOS ELEMENTALES

Cuando se hace un estudio estadístico el investigador decide si analizará toda la población o una muestra elegida previamente.

- **Población:** es el conjunto formado por todos los elementos de los que se quiere estudiar alguna característica. Cada elemento de la población se llama **individuo**.
- **Muestra:** es un subconjunto de la población, se toma cuando la población es muy numerosa. Debe elegirse de forma que sea representativa de toda la población estudiada. El número de elementos de la muestra se llama tamaño y se representa por la letra **N**.

Interés en el deporte Tenis, entonces esa será la población. A veces no tenemos claro cuánta cantidad de personas conforman la población y esta limitación es importante conocerla.



¿Qué elementos influyen a la hora de seleccionar la muestra?

Hay dos casos y es cuando se conoce el tamaño de la población y cuando no, por ejemplo,

Una vez tengamos elegida la población y la muestra en caso necesario, tendremos que elegir las características o propiedades que queremos estudiar. Cada una de las características o propiedades es una **variable estadística**. Dependiendo de los posibles valores que puedan tomar se clasifican en variables cualitativas o cuantitativas.

- **Variables cualitativas** : si los valores que toma son cualidades, no se pueden expresar mediante un número.
Por ejemplo: nacionalidad, deporte, partido político, color de los ojos, ...
- **Variables cuantitativas**: si los valores que toma son numéricos.
 - **Discretas**: cuando solo puede tomar valores enteros.
Por ejemplo: número de hermanos, edad,...
 - **Continuas**: cuando puede tomar cualquier valor de un intervalo dado. Estos valores los agruparemos en intervalos llamados **clases**.
Por ejemplo: peso, temperatura,...

2.- TABLAS DE FRECUENCIAS

Cuando las variables del estudio están escogidas, tenemos que recoger los datos. Después de recopilar los datos, para organizarlos y analizarlos se utiliza una tabla, llamada **tabla de frecuencias**.

- **x_i** : valores que toma la variable (cualitativa o cuantitativa discreta) o punto medio del intervalo (cuantitativa continua), llamado marca de la clase.
- **f_i** se llama **frecuencia absoluta** y es el número de veces que la variable toma el valor x_i . La suma de las frecuencias absolutas es igual al $n.^{\circ}$ total de datos.
- **F_i** se llama **frecuencia absoluta acumulada** y se calcula sumando los valores de las frecuencias absolutas de los valores que son menores o iguales que él. F_i representa el número de veces que la variable toma un valor menor o igual que x_i .
- **h_i** se llama **frecuencia relativa** y se calcula dividiendo cada f_i entre el número total de datos, N . Esta frecuencia nos informa acerca de la proporción o el peso que tiene algún valor en la población o en la muestra.

- H_i se llama **frecuencia relativa acumulada** y se calcula sumando los valores de las frecuencias relativas de los valores que son menores o iguales que él. H_i representa la proporción de datos que hay menores o iguales al valor x_i correspondiente.

Ejemplo variable cuantitativa discreta: Estas son las edades de los integrantes de un equipo deportivo:

14 15 13 13 14 15 15 18 14 13 15 13 14 15 16 14 15 13 13 15

a) Elabora la tabla de frecuencias.

x_i	f_i	F_i	h_i	H_i	%
13	6	6	0,3	0,3	30%
14	5	11	0,25	0,55	25%
15	7	18	0,35	0,9	35%
16	1	19	0,05	0,95	5%
18	1	20	0,05	1	5%
Total	20		1		100%

- b) ¿Cuántos chicos y chicas tienen 15 años?
Hay 7 integrantes de 15 años.
- c) ¿Cuántos chicos y chicas tienen 14 años o menos?
Hay 11 integrantes con 14 años o menos.
- d) ¿Qué porcentaje de chicos y chicas tiene 13 años?
El 30% de los chicos y chicas tiene 13 años.

Ejemplo variable cuantitativa continua : Las notas en la materia de matemáticas de un grupo de alumnos y alumnas han sido:

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

a) Elabora la tabla de frecuencias

La nota mínima es 2 y la máxima es 7,8. Tomamos, por ejemplo, intervalos de amplitud 1:

Clase	x_i	f_i	F_i	h_i	H_i	%
[2,3)	2,5	3	3	0,15	0,15	15%
[3,4)	3,5	2	5	0,1	0,25	10%
[4,5)	4,5	3	8	0,15	0,4	15%
[5,6)	5,5	5	13	0,25	0,65	25%
[6,7)	6,5	3	16	0,15	0,8	15%
[7,8)	7,5	4	20	0,2	1	20%
Total		20		1		100

- b) ¿Cuántos alumnos y alumnas tienen una nota superior o igual a 5 e inferior a 6? Hay 5 alumnos y alumnas que tienen una nota igual o superior a 5 e inferior a 6.
- c) ¿Cuántos alumnos y alumnas tienen una nota inferior a 7?
Hay 16 alumnos y alumnas con nota inferior a 7.
- d) ¿Qué porcentaje de alumnado tiene una nota superior o igual a 3 e inferior a 4? El 10% del alumnado tiene una nota igual o superior a 3 e inferior a 4.
- e) ¿Cuál es el porcentaje de aprobados?
El porcentaje de aprobados es del 60%

3.- GRÁFICOS ESTADÍSTICOS

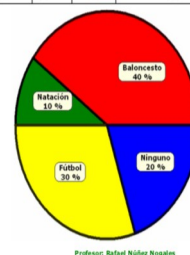
Los datos obtenidos en un estudio estadístico los podemos representar con diferentes gráficos. Los gráficos nos ayudan a analizar los datos a simple vista.

Diagrama de sectores: Puede aplicarse a cualquier tipo de variable, aunque es el más adecuado para variables cuantitativas. Es un círculo dividido en sectores de ángulo proporcional a la frecuencia de cada valor. La amplitud de cada sector se obtiene multiplicando la frecuencia relativa por 360°.

Ejemplo:

Deporte	f _i	h _i	Ángulo del sector
Baloncesto	12	0,4	0,4 · 360° = 144°
Natación	3	0,1	0,1 · 360° = 36°
Fútbol	9	0,3	0,3 · 360° = 108°
Ninguno	6	0,2	0,2 · 360° = 72°
Total	N=30	1	

Fútbol	9	30	30% de 360° = 108°
Ninguno	6	20	20% de 360° = 72°
Total	30 = n	100	360°

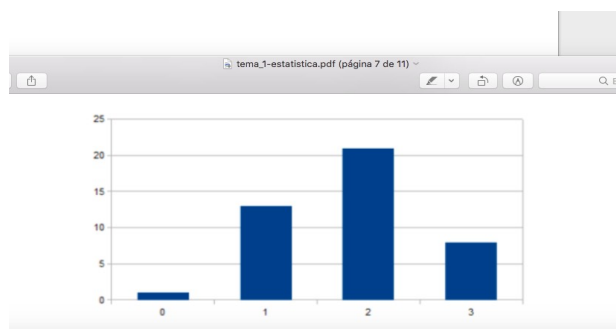


Ejercicios:
 Pág. 216:
 Pág. 217:
 Pág. 224:

Diagrama de barras: Puede aplicarse a variables cualitativas o cuantitativas discretas, aunque se considera el idóneo para las variables cuantitativas discretas. Cada valor se corresponde con una barra de longitud proporcional a su frecuencia. Las barras deben ser de la misma anchura y debemos dibujarlas separadas.

Ejemplo:

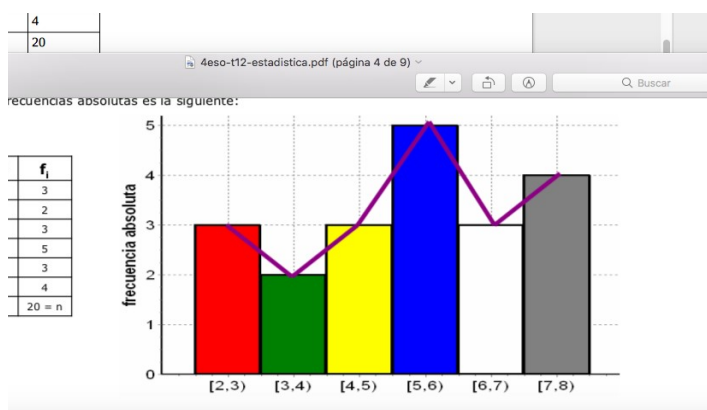
N.º coches en casa	f _i
0	1
1	13
2	21
3	8
Total	N=43



Histograma: se utilizan cuando los datos vienen agrupados en intervalos o clases. Cada valor se representa con un rectángulo de anchura el intervalo correspondiente y con la altura proporcional a su frecuencia. Las barras debemos dibujarlas juntas.

Ejemplo:

Clases	f _i
[2,3)	3
[3,4)	2
[4,5)	3
[5,4)	5
[6,7)	3
[7,8)	4
Total	20



4.- MEDIDAS DE CENTRALIZACIÓN Y POSICIÓN

Las **medidas de centralización** son la **media, la mediana y la moda**. Nos indican en torno a que valor distribuyen los datos.

Los **cuartiles** (deciles y percentiles) son **medidas de posición**. Dividen un conjunto de datos en grupos con el mismo número de individuos.

La media, mediana y los cuartiles solo se pueden calcular cuando la variable es cuantitativa. La moda se puede calcular para cualquier tipo de variable.

Media (\bar{x}): es la suma de todos los datos dividida entre el n.º total de datos. Si trabajamos con pocos datos los sumamos y dividimos entre el n.º de datos.

En caso contrario aplicamos la fórmula: $\bar{x} = \frac{\sum x_i \cdot f_i}{n}$

Si los datos están agrupados en intervalos, se toma como x_i la marca de la clase de cada intervalo.

Moda (Mo): es el valor que más se repite en los datos, es decir, el que tiene mayor frecuencia absoluta.

Si los datos están agrupados en intervalos, el intervalo modal es el que presenta mayor frecuencia.

Puede haber más de una moda o puede que no haya moda porque todos los valores tengan la misma frecuencia absoluta.

Mediana (Me): es el valor que ocupa la posición central una vez ordenados los datos en orden creciente, es decir, el valor que es mayor que el 50% y menor que el otro 50%.

Si el número de datos es impar: la mediana es el dato que ocupa la posición central.

Si el número de datos es par: la mediana es la media aritmética de los valores centrales.

Si los datos están agrupados en intervalos, el intervalo mediano es el que contiene a la mediana.

5.- MEDIDAS DE DISPERSIÓN

Para conocer la distribución de los datos respecto de los valores centrales estudiamos las medidas de dispersión: recorrido o rango, varianza y desviación típica.

- **Recorrido o rango** (R): es la diferencia entre el mayor y el menor valor de x_i . Si los datos están agrupados en intervalos es la diferencia entre el extremo superior del intervalo mayor y el extremo inferior del menor.

- **Varianza** (s^2 o σ^2): se calcula mediante la fórmula: $\sigma^2 = \frac{\sum x_i^2 \cdot f_i}{N} - \bar{x}^2$

Cuanto más pequeña sea más concentrados estarán los valores alrededor de la media.

- **Desviación típica** (s o σ): se calcula $\sigma = \sqrt{\sigma^2}$

- **Coefficiente de variación**: se calcula $CV = \frac{\sigma}{\bar{x}}$

El resultado de suele expresar en %, y cuanto menor es, más representativa de la función es la media.

Ejemplo: La edad de los 50 socios de un club deportivo juvenil viene dado por :

Edad	16	17	18	19	20	21
N.º socios	5	8	10	12	9	6

- a) Calcula media, moda, mediana y cuartiles.
 b) Calcula el rango, la varianza y la desviación típica.

x_i	f_i	F_i	h_i	H_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
16	5	5	0,1	0,1	80	1280
17	8	13	0,16	0,26	136	2312
18	10	23	0,2	0,46	180	3240
19	12	35	0,24	0,7	228	4332
20	9	44	0,18	0,88	180	3600
21	6	50	0,12	1	126	2646
TOT	50				930	17410

a)

$$\bar{x} = \frac{\sum x_i \cdot f_i}{N} = \frac{930}{50} = 18,6$$

$$Mo = 19$$

$$\frac{N}{4} = \frac{50}{4} = 12,5 \quad Q_1 = 17$$

$$\frac{2N}{4} = \frac{100}{4} = 25 \quad Q_2 = Me = 20$$

$$\frac{3N}{4} = \frac{150}{4} = 37,5 \quad Q_3 = 20$$

$$R = 21 - 16 = 5$$

b)

$$\frac{N}{4} = \frac{50}{4} = 12,5 \quad Q_1 = 17$$

$$\frac{2N}{4} = \frac{100}{4} = 25 \quad Q_2 = Me = 20$$

$$\frac{3N}{4} = \frac{150}{4} = 37,5 \quad Q_3 = 20$$

$$R = 21 - 16 = 5$$

$$\sigma^2 = \frac{\sum x_i^2 \cdot f_i}{N} - \bar{x}^2 = \frac{17410}{50} - 18,6^2 = 2,24$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{2,24} = 1,5$$