

Tema 15. Estadística.

15.1. Caracteres y variables estadísticas.

Se llama carácter estadístico a cada uno de los aspectos observables de un grupo. Los caracteres estadísticos pueden ser:

- Cualitativos. Son aquellos que no pueden ser ni medidos ni contados. Por ejemplo el color de los ojos, el nombre de las personas, etc...
- Cuantitativos. Son aquellos que se pueden medir o contar. Por ejemplo la edad o el peso de una persona.

En este tema nos centraremos en los caracteres cuantitativos, que también reciben el nombre de **variables estadísticas**, y pueden ser:

- Discretas. Si solo pueden tomar valores numéricos separados entre sí. Por ejemplo el número de hijos de una familia.
- Continuas. Cuando pueden tomar todos los valores posibles dentro de un intervalo. Por ejemplo el peso de una persona.

15.2. Población y Muestra.

Una población es el conjunto de todos los elementos sometidos al estudio, p.e. alumnos de 4º de E.S.O.

Una muestra es el subconjunto de la población formado por los elementos elegidos para realizar el estudio, p.e. los 30 alumnos de 4º de E.S.O del Colegio el Taller.

15.3. Tabla de Frecuencias.

Los datos estadísticos se organizan en tablas en las que aparecen:

- Los valores de la variable. Que pueden ser:
 - Discreta. X_i que son los distintos valores de la variable.
 - Continua. Clases: los intervalos en los que se agrupa la variable.
Marca de clases (x_i): Pto. medio de cada intervalo.

- Frecuencia Absoluta (f_i). Es el número de veces que se repite un valor de x_i en la muestra. La suma de todas las frecuencias absolutas coincide con el tamaño de la muestra (que se representa por N)
- Frecuencia Relativa (h_i). Es el cociente entre la frecuencia absoluta y el tamaño de la muestra.
 e.d. $h_i = f_i/N$
- Porcentaje (P_i). Se obtiene a partir de $P_i = 100 \times h_i$
- Frecuencia Absoluta Acumulada (F_i). Es la suma de todas las frecuencias absolutas de todos los valores menores o iguales que él, es decir

$$F_i = f_1 + f_2 + \dots + f_i$$
- Frecuencia Relativa Acumulada (H_i). Se obtiene a partir de

$$H_i = F_i/N$$
- Porcentaje Acumulado (P_i). Se obtiene a partir de la fórmula

$$P_i = 100 \times H_i$$

15.4. Tabla de frecuencias de una variable discreta.

Las calificaciones de los 30 alumnos de la clase en un examen de matemáticas son las siguientes:

4, 7, 5, 3, 5, 4, 4, 5, 6, 8,
 7, 5, 5, 8, 6, 5, 6, 7, 6, 4,
 5, 8, 5, 5, 3, 4, 5, 6, 5, 5,

Construye la tabla de frecuencias.

Variable x_i	Frec. Abs. f_i	Frec. Rel. h_i	Porcent. P_i	Fr. Abs. Ac. F_i	Fr. Rel. Ac. H_i	Porc. Acum. P_i
3	2	$\frac{2}{30} = 0'06$	6%	2	$\frac{2}{30} = 0'06$	6%
4	5	$\frac{5}{30} = 0'16$	16%	$5+2 = 7$	$\frac{7}{30} = 0'23$	23%
5	12	$\frac{12}{30} = 0'4$	40%	$12+5+2 = 19$	$\frac{19}{30} = 0'63$	63%
6	5	$\frac{5}{30} = 0'16$	16%	$5+12+5+2 = 24$	$\frac{24}{30} = 0'8$	80%
7	3	$\frac{3}{30} = 0'1$	10%	$3+5+12+5+2 = 27$	$\frac{27}{30} = 0'9$	90%
8	3	$\frac{3}{30} = 0'1$	10%	30	$\frac{30}{30} = 1$	100%

$N = 30$

15.5. Tabla de frecuencias de una variable continua

El peso de 20 alumnos de la clase es el siguiente

53, 58, 62, 65, 71, 73, 60, 68, 72, 54

55, 63, 64, 67, 68, 64, 70, 73, 64, 52

Construye la tabla de frecuencias agrupando los datos por intervalos de 5 en 5 comenzando por 50 kg.

Clases	Marca de Clase x_i	Frec. Abs. f_i	Frec. Rel. h_i	Porcent. P_i	Fr. Abs. Ac. F_i	Fr. Rel. Ac. H_i	Porc. Acum. P_i
[50,55)	52'5	3	$\frac{3}{20} = 0'15$	15%	3	$\frac{3}{20} = 0'15$	15%
[55,60)	57'5	2	$\frac{2}{20} = 0'1$	10%	2+3=5	$\frac{5}{20} = 0'25$	25%
[60,66)	62'5	5	$\frac{5}{20} = 0'25$	25%	5+2+3=10	$\frac{10}{20} = 0'5$	50%
[66,70)	67'5	5	$\frac{5}{20} = 0'25$	25%	5+5+2+3=15	$\frac{15}{20} = 0'75$	75%
[70,75)	72'5	5	$\frac{5}{20} = 0'25$	25%	20	$\frac{20}{20} = 1$	100%

$N = 20$

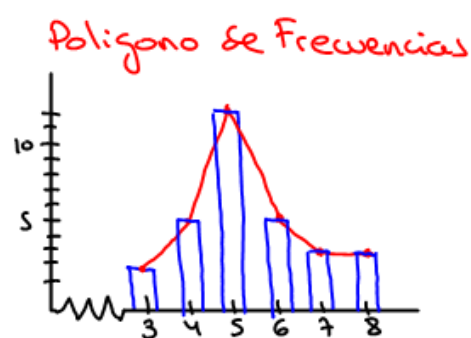
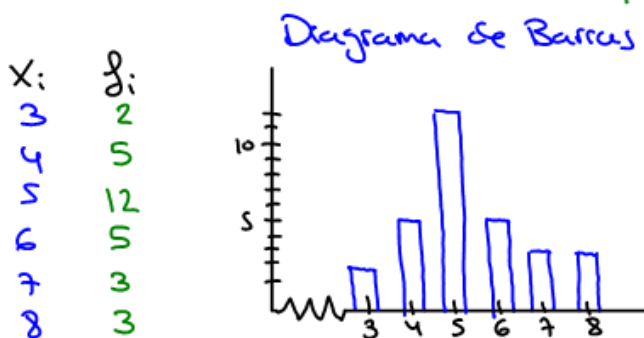
15.6. Gráficos Estadísticos.

• Variable Discreta.

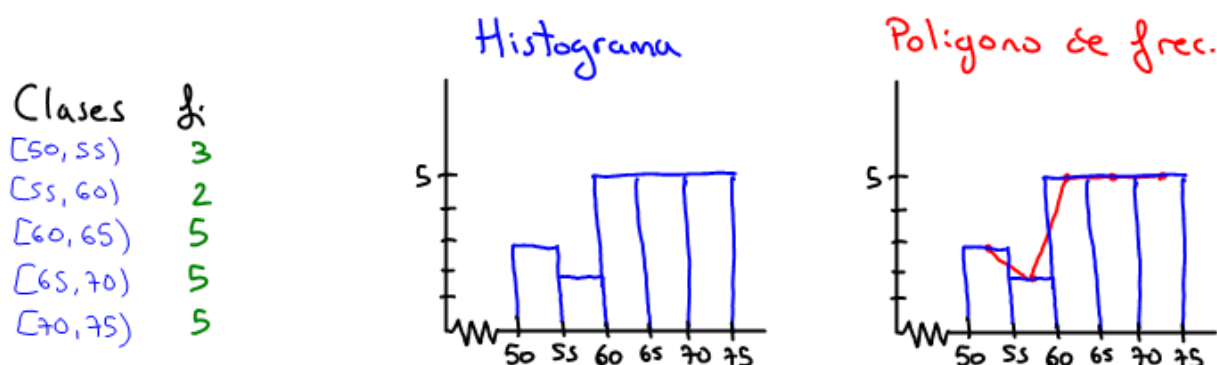
- Diagrama de barras. Es un gráfico formado por barras separadas. El eje OX representa los valores de x_i y el eje OY nos da sus frecuencias absolutas (f_i).

- Polígono de frecuencias. Es la grafica que se crea uniendo los pto. medios de cada barra del diagrama de barras.

Ejemplo. Representa gráficamente el ejemplo de variable discreta visto en el pto. 14.4.



- Variable continua.
 - Histograma. Es el gráfico formado por rectángulos unidos, que tienen de base cada uno de los intervalos y de altura su frecuencia absoluta correspondiente.
 - Polígono de frecuencias. Es la gráfica que se crea uniendo los pts. medios de cada barra del diagrama de barras.
- Ejemplo. Representa gráficamente el ejemplo de variable continua visto en el pto. 14.5.



15.7. Parámetros de Centralización.

- Media Aritmética (\bar{x}). Es el resultado que se obtiene al sumar todos los datos de la muestra y dividirlo entre el tamaño de la muestra.
 Para calcular la media aritmética se crea una nueva columna en la tabla de frecuencias llamada $x_i \cdot f_i$. La media aritmética es el cociente de la suma de los números de esta nueva columna entre el tamaño de la muestra.

$$e.d \quad \bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_n \cdot f_n}{N} = \frac{\sum x_i \cdot f_i}{N}$$

- Moda. Es el valor de x_i con mayor frecuencia absoluta.
- Mediana. Es el valor de x_i que cumple que su frecuencia absoluta acumulada (F_i) es el primero que es mayor que $N/2$.

15.8. Parámetros de Dispersión.

- Recorrido. Es la diferencia entre el mayor y el menor de los datos de la muestra.
- Varianza. (V). Se obtiene creando una nueva columna en la tabla de frecuencias llamada $x_i^2 \cdot f_i$.
La varianza se calcula a partir de la fórmula.

$$V = \frac{\sum x_i^2 \cdot f_i}{N} - (\bar{x})^2$$

- Desviación típica. (σ). Es la raíz cuadrada del valor absoluto de la varianza.

$$\text{e.d. } \sigma = \sqrt{|V|}$$

- Coeficiente de Variación (CV). Es el cociente entre la desviación típica y la media aritmética.

$$\text{e.d. } CV = \frac{\sigma}{\bar{x}}$$

15.9. Ejemplos.

• Variable Discreta.

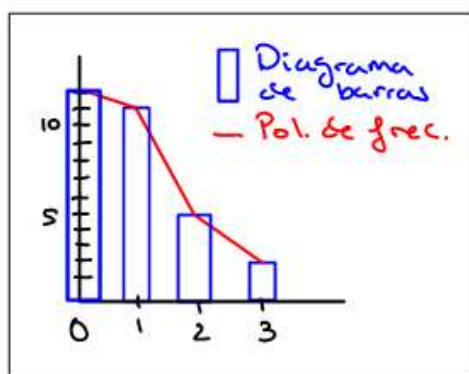
Se ha preguntado a 30 personas el número de libros que leen mensualmente. Sus respuestas han sido:

2, 1, 0, 0, 0, 1, 2, 1, 3, 0, 0, 1, 3, 1, 0,
 0, 1, 1, 0, 0, 2, 0, 1, 0, 0, 1, 2, 2, 1, 1

- Construye la tabla de frecuencias
- Representa el diagrama de barras y el polígono de frecuencias.
- Calcula la media aritmética, la moda y la mediana.
- Calcula el recorrido, la varianza, la desviación típica y el coeficiente de variación.

X_i	f_i	h_i	P_i	F_i	H_i	P_i	$X_i \cdot f_i$	$X_i^2 \cdot f_i$
0	12	$\frac{12}{30} = 0'4$	40%	12	$\frac{12}{30} = 0'4$	40%	$0 \cdot 12 = 0$	$0 \cdot 0 = 0$
1	11	$\frac{11}{30} = 0'36$	36%	$11+12=23$	$\frac{23}{30} = 0'76$	76%	$1 \cdot 11 = 11$	$1 \cdot 11 = 11$
2	5	$\frac{5}{30} = 0'16$	16%	$5+11+12=28$	$\frac{28}{30} = 0'93$	93%	$2 \cdot 5 = 10$	$2 \cdot 10 = 20$
3	2	$\frac{2}{30} = 0'06$	6%	30	$\frac{30}{30} = 1$	100%	$3 \cdot 2 = 6$	$3 \cdot 6 = 18$
							27	49

$N=30$



Media aritm. $\bar{X} = \frac{27}{30} = 0'9$

Moda = 0

Mediana = 1

Recorrido = $3 - 0 = 3$

Varianza $V = \frac{49}{30} - (0'9)^2 = 1'63 - 0'81 = 0'82$

Desv. típica $\sigma = \sqrt{0'82} = 0'9$

Coef. de variación $CV = \frac{0'9}{0'9} = 1$

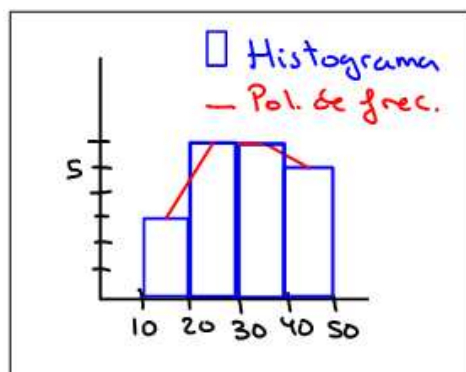
• Variable Continua.

Se ha pasado un test de conocimientos básicos de matemáticas a 20 personas. De las 50 preguntas las respuestas acertadas han sido:

23, 25, 17, 31, 43, 28, 30, 33, 47, 12,
 41, 40, 35, 31, 29, 30, 23, 42, 28, 15

- Construye la tabla de frecuencias agrupando los datos por intervalos de 10 en 10.
- Representa el histograma y el polígono de frecuencias.
- Calcula la media aritmética, la moda y la mediana.
- Calcula el recorrido, la varianza, la desviación típica y el coeficiente de variación.

Clases	X_i	f_i	h_i	P_i	F_i	H_i	P_i	$X_i \cdot f_i$	$X_i^2 \cdot f_i$
[10,20)	15	3	$\frac{3}{20} = 0'15$	15%	3	$\frac{3}{20} = 0'15$	15%	$15 \cdot 3 = 45$	$15 \cdot 45 = 675$
[20,30)	25	6	$\frac{6}{20} = 0'3$	30%	$6+3=9$	$\frac{9}{20} = 0'45$	45%	$25 \cdot 6 = 150$	$25 \cdot 150 = 3750$
[30,40)	35	6	$\frac{6}{20} = 0'3$	30%	$6+6+3=15$	$\frac{15}{20} = 0'75$	75%	$35 \cdot 6 = 210$	$35 \cdot 210 = 7350$
[40,50)	45	5	$\frac{5}{20} = 0'25$	25%	20	$\frac{20}{20} = 1$	100%	$45 \cdot 5 = 225$	$45 \cdot 225 = 10125$
		20						630	21900



Media aritm. $\bar{x} = \frac{630}{20} = 31'5$

Moda = 25 y 35

Mediana = 35

Recorrido = $47 - 12 = 35$

Varianza $V = \frac{21900}{20} - (31'5)^2$
 $= 1095 - 992'25 = 102'75$

Desv. típica $\sigma = \sqrt{102'75} = 10'13$

Coef. de variación $CV = \frac{10'13}{31'5} = 0'32$