

ESTADÍSTICAS

INDICE

Contenido

1	DATOS ESTADÍSTICOS	2
2	TABLAS DE FRECUENCIAS (VARIABLES DISCRETAS)	2
3	TABLAS DE FRECUENCIAS (VARIABLES CONTINUA)	3
4	GRÁFICOS ESTADÍSTICOS	3
4.1	PICTOGRAMAS	3
4.2	DIAGRAMA DE BARRAS	3
4.3	HISTOGRAMA	4
4.4	DIAGRAMA DE SECTORES	5
5	MEDIDAS DE CENTRALIZACIÓN	5
6	MEDIDAS DE POSICIÓN	6
7	MEDIDAS DE DISPERSIÓN:	8

1 Datos estadísticos

Población: Es el conjunto de elementos sobre el que se realiza el estudio

Muestra: Como no podemos hacer el estudio sobre toda la población, la muestra es la parte de la población sobre la que hacemos el estudio

Individuo: Es cada uno de los elementos que forman la muestra o población

Para hacer un estudio sobre los pasajeros de los viajeros de los vehículos privados en España, preguntamos a 1000 conductores.

Muestra: los 1000 conductores a los que preguntamos.

Población: todos los conductores de España.

Individuos: cada uno de los conductores

Variable estadística: Cada una de las características que podemos estudiar

Tipos de variables:

- Cualitativas los valores son cualidades (son rubios, adultos, niños..).
- Cuantitativas los valores son numéricos.
 - Discretas: solo puede tomar algunos valores (números de pasajeros 1,2,3..).
 - Continuas: tiene tantos valores como queramos en un intervalo (altura de los pasajeros entre 1,20-1,50..)

2 Tablas de frecuencias (variables discretas)

Ejemplo:

Analizamos las notas obtenidas por los 30 alumnos de una clase en el último examen:

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

N : Tamaño de la población o muestra (30)

X_i variable o dato estadístico (variando i desde 1 hasta n , obtenemos todos los datos).

Frecuencia absoluta f_i : Es el número de veces que ese dato aparece en la muestra

Frecuencia absoluta acumulada F_i : Para cada dato es igual a la suma de las f_i de todos los datos inferiores a él.

Frecuencia Relativa h_i : La frecuencia relativa de un dato es igual al cociente de la frecuencia absoluta de ese dato dividido por el número total de datos N

Frecuencia relativa acumulada H_i : Para cada dato es igual a la suma de las h_i de todos los datos inferiores a él.

*Las frecuencias acumuladas no se pueden calcular con variable cualitativas

X_i	f_i	F_i	h_i	h_i %	H_i	H_i %
1	3	3	0,1	10%	0,1	10%
2	3	6	0,1	10%	0,2	20%
3	3	9	0,1	10%	0,3	30%
4	4	13	0,13333	13%	0,43333	43%
5	6	19	0,2	20%	0,63333	63%
6	6	25	0,2	20%	0,83333	83%
7	3	28	0,1	10%	0,93333	93%
8	2	30	0,06667	7%	1	100%
	N=30		1			

3 Tablas de frecuencias (variables continua)

Analizamos la altura, en cm, de los 30 alumnos de la clase:

134 125 133 136 138 141 139 150 155 123
 137 141 144 126 152 129 132 136 138 139
 140 120 126 123 121 129 145 152 157 120

Agrupamos los datos en intervalos. La amplitud de cada intervalo se calcula

$$\text{Amplitud del intervalo} = \frac{\text{Valor maximo} - \text{valor minimo}}{\sqrt{N}};$$

Siendo N el numero de datos

$$\text{Amplitud del intervalo} = \frac{157 - 120}{\sqrt{30}} = 6,755; \text{ tomamos } 6$$

Marca de clase Xi es el punto medio de una clase. Se obtiene hallando la media entre el límite superior y el inferior de cada clase.

Altura	Xi (marca de clase)	fi	Fi	hi	hi %	Hi	Hi %
[120-126)	123	6	6	0,2	20%	0,2	20%
[126-132)	129	4	10	0,133333	13%	0,333333	33%
[132-138)	135	6	16	0,2	20%	0,533333	53%
[138-144)	141	7	23	0,233333	23%	0,766667	77%
[144-150)	147	2	25	0,066667	7%	0,833333	83%
[150-156)	153	4	29	0,133333	13%	0,966667	97%
[156-162)	159	1	30	0,033333	3%	1	100%

4 Gráficos estadísticos

4.1 Pictogramas

Son representaciones gráficas, en las que se utiliza figuras relacionadas con el tema que se trate. No son muy rigurosos.

El tamaño de las figuras mantiene una proporción con los datos representados.

4.2 Diagrama de barras

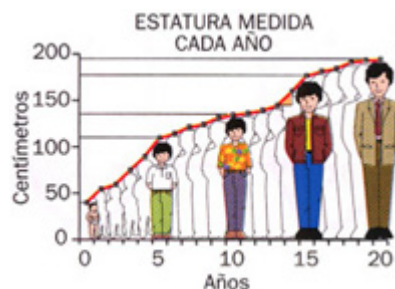
Sirve para representar las variables cualitativas y las cuantitativas discretas.

Es una representación gráfica formada por barras finas separadas y de forma que la altura de las barras es proporcional a las frecuencias.

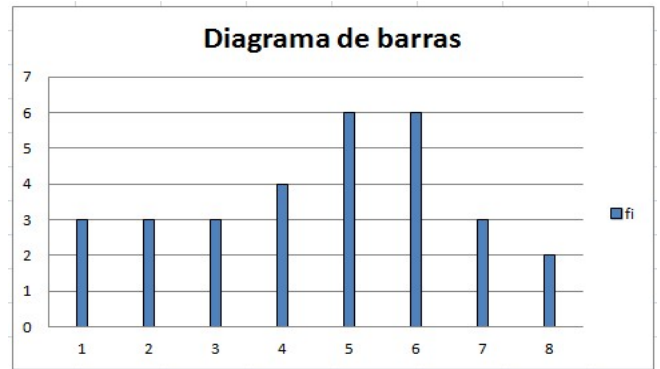
Representamos los datos del ejemplo anterior

Analizamos las notas obtenidas por los 30 alumnos de una clase en el último examen:

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



	fi	Fi	hi	hi %	Hi	Hi %
1	3	3	0,1	10%	0,1	10%
2	3	6	0,1	10%	0,2	20%
3	3	9	0,1	10%	0,3	30%
4	4	13	0,133333	13%	0,433333	43%
5	6	19	0,2	20%	0,633333	63%
6	6	25	0,2	20%	0,833333	83%
7	3	28	0,1	10%	0,933333	93%
8	2	30	0,066667	7%	1	100%
	N=30		1	100%		



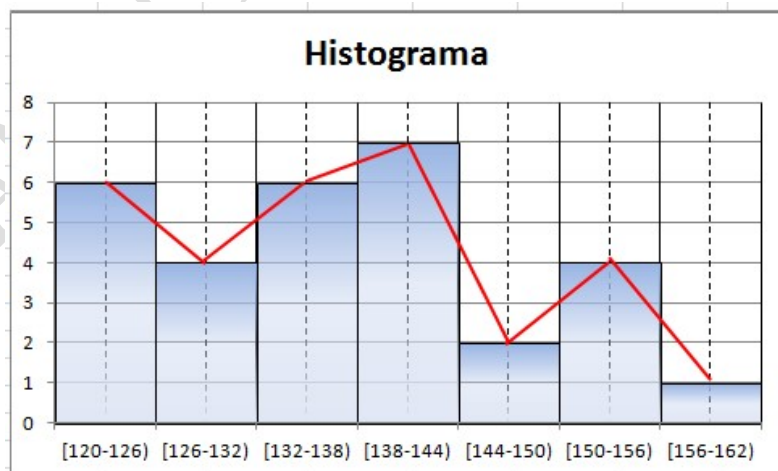
4.3 Histograma

Se utiliza para representar estadísticas con variables cuantitativas continuas.

Representamos los datos del ejemplo anterior, (altura, en cm, de los 30 alumnos de la clase).

Altura	X_i (marca de clase)	fi	Fi	hi	hi %	Hi
[120-126)	123	6	6	0,2	20%	0,2
[126-132)	129	4	10	0,133333	13%	0,333333
[132-138)	135	6	16	0,2	20%	0,533333
[138-144)	141	7	23	0,233333	23%	0,766667
[144-150)	147	2	25	0,066667	7%	0,833333
[150-156)	153	4	29	0,133333	13%	0,966667
[156-162)	159	1	30	0,033333	3%	1

Se divide el eje de abscisas en intervalos y se levanta un rectángulo en cada intervalo de forma que la altura sea igual a la frecuencia.



El Polígono de frecuencias se halla uniendo los puntos medios de los lados superiores del rectángulo del histograma

4.4 Diagrama de sectores

Analizamos las notas obtenidas por los 30 alumnos de una clase en el último examen:

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

Xi	fi	Fi	hi	hi %	Hi	Hi %
1	3	3	0,1	10%	0,1	10%
2	3	6	0,1	10%	0,2	20%
3	3	9	0,1	10%	0,3	30%
4	4	13	0,133333	13%	0,433333	43%
5	6	19	0,2	20%	0,633333	63%
6	6	25	0,2	20%	0,833333	83%
7	3	28	0,1	10%	0,933333	93%
8	2	30	0,066667	7%	1	100%
	N=30		1	100%		

El círculo se divide en tantos sectores como datos tenga la variable y el sector circular es proporcional a la frecuencia del dato.

Diagrama de sectores:
(Variable Xi; frecuencia fi)

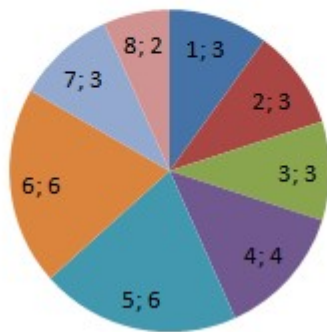
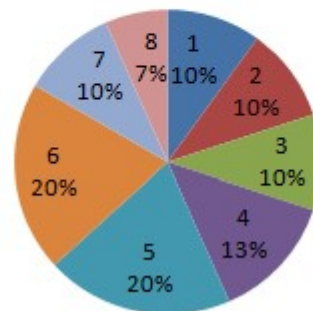


Diagrama de sectores:
(Variable Xi; frecuencia fi%)



5 Medidas de centralización

Media Aritmetica $\bar{X}$:

Se halla multiplicando todos los datos por su frecuencia y dividiéndolo por el número total de datos.

$$\bar{X} = \frac{\sum_{i=1}^n fi * Xi}{N}$$

Moda, Mo Es el dato que tiene mayor frecuencia. Si la variable es continua le llamamos intervalo modal

Mediana Me: Es el dato que ocupa el valor central, después de ordenar los datos de menor a mayor. Si el número de datos es par se halla la media de los dos datos centrales

Ejemplo:

En una clase de 31 alumnos, preguntamos a cada alumno las horas que ve la televisión a la semana

Horas de TV (Xi)	1	2	3	4	5	6	7
Nº de alumnos(fi)	1	2	1	5	6	10	6

Media aritmética:

$$\bar{X} = \frac{1 \cdot 1 + 2 \cdot 2 + 3 \cdot 1 + 4 \cdot 5 + 6 \cdot 10 + 7 \cdot 6}{31} = \frac{1 + 4 + 3 + 20 + 60 + 42}{31}$$

$$\bar{X} = \frac{130}{31} = 4,19$$

Modas: Es la frecuencia que más se repite *6 horas*.

Mediana : El dato que ocupa el punto medio **6**

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

6 Medidas de posición

Las medidas de posición nos indican el lugar que ocupa un dato dentro del conjunto ordenado de de Valores

Cuartiles: Q_1, Q_2, Q_3 Son los tres valores que dividen al conjunto de datos ordenados en cuatro partes porcentualmente iguales. En cada tramo están el 25% de los datos del estudio

Percentiles: Son puntos que dividen el conjunto de datos en 100 partes iguales.

Para hacer estas medidas la variable debe ser cuantitativa y si la variable es continua tomaremos como valor la marca de clase.

1. Ejercicios estadísticas

Analizar las notas obtenidas por los 30 alumnos de una clase en el último examen:

Y calcular los cuartiles Q_1, Q_2, Q_3 y el percentil P_{85}

Xi	fi	Fi	hi	hi %	Hi	Hi %			
1	3	3	0,1	10%	0,1	10%			
2	3	6	0,1	10%	0,2	20%			
3	3	9	0,1	10%	0,3	30%	Q1	7,5	25%
4	4	13	0,133333	13%	0,433333	43%			
5	6	19	0,2	20%	0,633333	63%	Q2	15	50%
6	6	25	0,2	20%	0,833333	83%	Q3	22,5	75%
7	3	28	0,1	10%	0,933333	93%	P85	25,5	85%
8	2	30	0,066667	7%	1	100%			
	N=30		1	100%					

Q1: 25% de los valores están por debajo de este cuartil y el 75% por encima

Para calcularlo dividimos el numero de datos (30) por 4; $30/4=7,5 \Rightarrow 25\%$

En la columna de las frecuencias acumuladas (Fi) el primer dato cuyo valor sea mayor o igual que 7,5 es el 9 luego Q_1 corresponde al dato 3 (la cuarta parte de las notas son menores o iguales a 3).

Q2: 50% de los valores están por debajo de este cuartil y el 50% por encima

Para calcularlo dividimos el numero de datos (30) por 2; $30/2=15 \Rightarrow 50\%$

En la columna de las frecuencias acumuladas (Fi) el primer dato cuyo valor sea mayor o igual que 15 es el 19 luego Q2 corresponde al dato 5 (la mitad de las notas son menores o iguales a 5).

Q3: 75% de los valores están por debajo de este cuartil y el 25% por encima

Para calcularlo multiplicamos el numero de datos (30) por $3/4$; $30 \cdot 3/4 = 22,5 \Rightarrow 75\%$

En la columna de las frecuencias acumuladas (Fi) el primer dato cuyo valor sea mayor o igual que 22,5 es el 25 luego Q3 corresponde al dato 6 (El 75% de las notas son menores o iguales a 6).

Percentil 85: 85% de los valores están por debajo de este percentil

Para calcularlo multiplicamos el numero de datos (30) por $85/100$; $30 \cdot 85/100 = 25,5 \Rightarrow 85\%$

En la columna de las frecuencias acumuladas (Fi) el primer dato cuyo valor sea mayor o igual que 25,5 es el 28 luego P85 corresponde al dato 7 (El 85% de las notas son menores o iguales a 7).

2. Ejercicios estadísticas

¿Qué valor podemos añadir a este conjunto de datos 18, 7, 7, 9, 11, 10, 12, 15, 21 para que la mediana siga siendo la misma ?

Solución

Ordenamos los datos de menor a mayor

7, 7, 9, 10, **11**, 12, 15, 18, 21

La mediana es el 11 pues al ser el número de datos impar, es el dato que ocupa la posición central

Para que al añadir un dato tenga la misma median añadiremos el **11**

Los ordenamos: 7, 7, 9, 10, **11, 11**, 12, 15, 18, 21

Al ser el número de datos par los datos que ocupan la posición central son 11, 11 luego la mediana será $(11+11)/2=11$

3. Ejercicios estadísticas

Salen 30 plazas a concurso oposicion y se presentan 300 personas.

Hallar la nota de corte.

Se obtienen las siguientes notas

Notas	3	4	5	6	7	8	9
fi	14	66	73	50	64	28	5

Solución

Hallamos la Fi (frecuencia absoluta acumulada)

Notas	3	4	5	6	7	8	9
fi	14	66	73	50	64	28	5
Fi	14	80	153	203	267	295	300

Como de 300 solo pasan 30 el valor de corte es $300-30=270$ que corresponde al percentil

$$P = \frac{270 * 100}{300} = 90\% = \text{Percentil 90} \text{ corresponde a la nota 8}$$

La nota mínima para coger plaza es el 8

7 Medidas de dispersión:

Las medidas de dispersión, también llamadas medidas de variabilidad, muestran la variabilidad de una distribución, indicando por medio de un número si las diferentes puntuaciones de una variable están muy alejadas de la media. Cuanto mayor sea ese valor, mayor será la variabilidad, y cuanto menor sea, más homogénea será a la media. Así se sabe si todos los casos son parecidos o varían mucho entre ellos.

Rango o recorrido:

$R = \text{Max valor de la variable} - \text{Min valor de la variable}$

Desviación Media:

Desviación media o promedio de un conjunto de datos es la media de las desviaciones absolutas y es un resumen de la dispersión estadística.

Se calcula en valor absoluto pues el sumatorio siempre daría 0

$$\sum_{i=1}^n fi(Xi - \bar{X})$$

La desviación media se expresa, de acuerdo a esta fórmula:

$$DM = \frac{\sum_{i=1}^n fi|Xi - \bar{X}|}{N};$$

Varianza (σ^2):

Es la media de los cuadrados de las desviaciones.

$$\sigma^2 = \frac{\sum_{i=1}^n fi(Xi - \bar{X})^2}{N}$$

Proceso para desarrollar la otra fórmula de la varianza:

Dado que es una identidad notable, cuadrado de una diferencia lo desarrollamos $(a - b)^2 = a^2 + b^2 - 2ab$

$$\begin{aligned} \sigma^2 &= \frac{\sum_{i=1}^n fi(Xi - \bar{X})^2}{N} = \frac{\sum_{i=1}^n fiXi^2}{N} + \frac{\sum_{i=1}^n fi\bar{X}^2}{N} - \frac{2\sum_{i=1}^n fiXi\bar{X}}{N} \\ &= \frac{\sum_{i=1}^n fiXi^2}{N} + \frac{N\bar{X}^2}{N} - \frac{2\bar{X}\sum_{i=1}^n fiXi}{N} = \frac{\sum_{i=1}^n fiXi^2}{N} + \bar{X}^2 - 2\bar{X} * X = \\ &= \frac{\sum_{i=1}^n fiXi^2}{N} + \bar{X}^2 - 2\bar{X}^2 = \frac{\sum_{i=1}^n fiXi^2}{N} - \bar{X}^2 \end{aligned}$$

Notas

$$\sum_{i=1}^n fi = N; \text{ y por otro lado } \frac{\sum_{i=1}^n fiXi}{N} = \bar{X}$$

$$\sigma^2 = \frac{\sum_{i=1}^n fi * Xi^2}{N} - \bar{X}^2$$

Desviación Típica:

Es la raíz cuadrada positiva de la varianza

$$\sigma = \sqrt{\frac{\sum_{i=1}^n f_i * X_i^2}{N} - \bar{X}^2}; \quad \sigma = \sqrt{\frac{\sum_{i=1}^n f_i (X_i - \bar{X})^2}{N}}$$

Coefficiente de variación:

Es el cociente entre la desviación Típica y la media

En estadística, cuando se desea hacer referencia a la relación entre el tamaño de la media y la variabilidad de la variable, se utiliza el coeficiente de variación

Por otro lado presenta problemas ya que a diferencia de la desviación típica este coeficiente es variable ante cambios de origen. Por ello es importante que todos los valores sean positivos y su media dé, por tanto, un valor positivo. *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.*

$$CV = \frac{\sigma}{\bar{X}}$$

4. Ejercicios estadísticas

Juan ha obtenido las siguientes notas en Ingles: 4, 6, 3, 5, 8, 7, 5. Halla las medidas de dispersion.

Solución:

X_i	f_i	$X_i * f_i$	$ X_i - \bar{X} $	$f_i X_i - \bar{X} $	$(X_i - \bar{X})^2$	$f_i (X_i - \bar{X})^2$
3	1	3	2,428571429	2,428571429	5,897959184	5,897959184
4	1	4	1,428571429	1,428571429	2,040816327	2,040816327
5	2	10	0,428571429	0,857142857	0,183673469	0,367346939
6	1	6	0,571428571	0,571428571	0,326530612	0,326530612
7	1	7	1,571428571	1,571428571	2,469387755	2,469387755
8	1	8	2,571428571	2,571428571	6,612244898	6,612244898
	7	38		9,428571429		17,71428571
	N	$\sum_{i=1}^n f_i * X_i$		$\sum_{i=1}^n f_i X_i - \bar{X} $		$\sum_{i=1}^n f_i (X_i - \bar{X})^2$

Media aritmética

$$\bar{X} = \frac{\sum_{i=1}^n f_i * X_i}{N} = \frac{38}{7}; \quad \bar{X} = 5,42857142;$$

Desviación media

$$DM = \frac{\sum_{i=1}^n f_i |X_i - \bar{X}|}{N} = \frac{9,428571429}{7}; \quad DM = 1,346938776;$$

Varianza (σ^2):

$$\sigma^2 = \frac{\sum_{i=1}^n f_i (X_i - \bar{X})^2}{N} = \frac{17,71428571}{7}; \quad \sigma^2 = 2,530612$$

Desviación típica (σ):

$$\sigma = \sqrt{\frac{\sum_{i=1}^n f_i (X_i - \bar{X})^2}{N}} = \sqrt{2,530612}; \quad \sigma = 1,590789818$$

Coefficiente de variación

$$CV = \frac{\sigma}{\bar{X}} = \frac{1,590789818}{1,590789818}; CV = 0,29304023; 29,30\%$$

5. Ejercicios estadísticas

Calcular y comparar las medidas de dispersion de estos datos

Nº de Hijos	0	1	2	3
Nº de alumnos por centro	115	410	220	142

Solución:

x_i	$ x_i - \bar{X} $	$(x_i - \bar{X})^2$
0	1,5	2,25
1	0,5	0,25
2	0,5	0,25
3	1,5	2,25
6	4	5

x_i	$ x_i - \bar{X} $	$(x_i - \bar{X})^2$
115	106,75	11395,5625
142	79,75	6360,0625
220	1,75	3,0625
410	188,25	35438,0625
887	376,5	53196,75

Media aritmética $\bar{X} = 6/4 = 1,5$

Desviación media $= 4/4 = 1$

Varianza $(\sigma^2) = 5/4 = 1,25$

Desviación típica $(\sigma): \sqrt{1,25} = 1,12$

CV $= 1,12/1,5 = 0,75 = 75\%$

Media aritmética $= 887/4 = 221,75$

Desviación media $= 376,5/4 = 94,125$

Varianza $(\sigma^2) = 53196,75/4 = 13299,19$

Desviación típica $(\sigma): \sqrt{13299,19} = 115,32$

CV $= 115,32/221,75 = 0,52 = 52\%$

6. Ejercicios estadísticas

Comparar la desviación de las siguientes variable:

Peso de los bagones de tren $\bar{X} = 4000 \text{ Kg}$ y $\sigma = 200 \text{ Kg}$ y peso de los pollos $\bar{X} = 3 \text{ Kg}$ y $\sigma = 0,02 \text{ Kg}$;

Solución:

Trenes:

CV $= 200/4000 = 0,05 = 5\%$

Pollos:

CV $= 0,02/3 = 0,0066 = 0,6\%$

7. Ejercicios estadísticas

En una prueba de Baloncesto en la que participan 100 alumnos el numero de aciertos es el que se muestra en la tabla. Hallar las medidas de centralización y dispersion

Aciertos	f _i
(0-5)	3
(5-10)	10
(10-15)	25
(15-20)	38
(20-25)	16
(25-30)	8
	100

Solución:

Hallamos la marca X_i

Aciertos	X_i	f_i	$X_i \cdot f_i$	$ X_i - \bar{X} $	$f_i X_i - \bar{X} $	$(X_i - \bar{X})^2$	$f_i (X_i - \bar{X})^2$
(0-5)	2,5	3	7,5	12,85	38,55	165,1225	495,3675
(5-10)	7,5	7	52,5	7,85	54,95	61,6225	431,3575
(10-15)	12,5	42	525	2,85	119,7	8,1225	341,145
(15-20)	17,5	32	560	2,15	68,8	4,6225	147,92
(20-25)	22,5	10	225	7,15	71,5	51,1225	511,225
(25-30)	27,5	6	165	12,15	72,9	147,6225	885,735
		100	1535		426,4		2812,75
		N	$\sum_{i=1}^n f_i \cdot X_i$		$\sum_{i=1}^n f_i X_i - \bar{X} $		$\sum_{i=1}^n f_i (X_i - \bar{X})^2$

$$\text{Rango} = 30 - 0 = 30$$

$$\text{Media Aritmetica } \bar{X} = \frac{1535}{100} = 15,35;$$

$$\text{Desviacion media} = \frac{426,4}{100} = 4,26;$$

$$\text{Varianza } (\sigma^2) = \frac{2812,75}{100} = 28,13;$$

$$\text{Desviacion tipica } (\sigma) = \sqrt{28,13} = 5,30;$$

$$\text{Coeficiente de Variación } CV = \frac{5,30}{15,35} = 0,34; \Rightarrow 34\%;$$

8. Ejercicios estadísticas

Un corredor entrena de Lunes a Viernes recorriendo las siguientes distancias 5,4,4,3,7 km. Si el sabado tambien entrena,

a) ¿Cuantos Km ha de recorrer para que la media se la misma?

b) Para que la mediana no varie

c) Para que la moda no varie.

Solución

$$\text{La Media aritmetica } \bar{X} = \frac{5 + 4 + 4 + 3 + 7}{5} = 4,6$$

Mediana 3,4,4,5,7, es el 4

Moda es el 4;

$$\text{a) Media aritmetica } \bar{X} = \frac{5 + 4 + 4 + 3 + 7 + a}{6} = 4,6;$$

$$a = 4,6 \cdot 6 - (5 + 4 + 4 + 3 + 7) = 27,6 - 23; a = 4,6$$

b) Mediana 3,4,4,5,7; debe ser 4 pues al ser el numero de datos par, se halla

$$\text{la media de los dos centiles } \frac{4 + 4}{2} = 4;$$

c) Moda, se debe añadir un 4 para que sea el valor que mas se repite

9. Ejercicios estadísticas

Una persona ingresa 5000 euros en un fondo de inversión el 1 de enero del 2011. Si la rentabilidad del fondo a lo largo de 4 años ha sido 2011 un 4%; 2012 un 6%; 2013 un -5% y en el 2014 un 3%. Calcular la rentabilidad media de dicho fondo a lo largo de estos años si no se ha retirado nada de dinero..

Solución

Año	2011	2012	2013	2014
Rentabilidad %	4	6	-5	3
Capital	5200	5512	5236,4	5393,492

$$1^{\circ} \text{ año } 5000 * 1,04 = 5200;$$

$$2^{\circ} \text{ año } 5200 * 1,06 = 5512;$$

$$3^{\circ} \text{ año } 5512 * 0,95 = 5236,4;$$

$$4^{\circ} \text{ año } 5236,4 * 1,03 = 5393,5$$

$$\text{Rentabilidad media} = \frac{5393,5}{5000} = 1,079 \Rightarrow 7,9\%$$

$$\text{La media de las rentabilidades } \bar{X} = \frac{4 + 6 - 5 + 3}{4} = 2\%$$

10. Ejercicios estadísticas

Tomamos una muestra de los salarios mensuales en € de 30 personas, que trabajan en una empresa:

Mujeres: 1200, 1300, 700, 900, 900, 1100, 1200, 1100, 1400, 1200, 1000, 1300, 1200, 1100, 1000;

Hombres: 1200, 1300, 1500, 1300, 1400, 1000, 1600, 1500, 1600, 1400, 900, 1000, 1500, 1600, 1700,

- Calcular la distribución de frecuencias, media, mediana y moda de cada grupo.
- Calcular las medidas de dispersión
- Comparar los dos grupos.

Solución

a)

Mujeres: 700, 900, 900, 1000, 1000, 1100, 1100, 1100, 1200, 1200, 1200, 1200, 1300, 1300, 1400

$$\bar{X} = \frac{700 + 900 * 2 + 1000 * 2 + 1100 * 3 + 1200 * 4 + 1300 * 2 + 1400}{15} = 1106,67$$

Mediana: 1100€

Moda: 1200€

b) Medidas de dispersion

Salario	fi	xi*fi	Xi - X̄	fi(Xi - X̄)	(Xi - X̄) ²	fi(Xi - X̄) ²
700	1	700	406,6666667	406,6666667	165377,7778	165377,7778
900	2	1800	206,6666667	413,3333333	42711,11111	85422,22222
1000	2	2000	106,6666667	213,3333333	11377,77778	22755,55556
1100	3	3300	6,666666667	20	44,44444444	133,3333333
1200	4	4800	93,33333333	373,3333333	8711,11111	34844,44444
1300	2	2600	193,3333333	386,6666667	37377,77778	74755,55556
1400	1	1400	293,3333333	293,3333333	86044,44444	86044,44444
	15	16600		2106,666667		469333,3333
	N	$\sum_{i=1}^n fi * Xi$		$\sum_{i=1}^n fi Xi - \bar{X} $		$\sum_{i=1}^n fi (Xi - \bar{X})^2$

Desviación media = $2106,7/15 = 140,4$

Varianza (σ^2) = $469333,3/15 = 31288,9$

Desviación típica (σ): $\sqrt{31288,9} = 176,9$

Coefficiente de Variación CV = $176,9/1106,67 = 0,1598 = 15,98\%$

Hombres: 900, 1000, 1000, 1200, 1300, 1300, 1400, 1400, 1500, 1500, 1500, 1600, 1600, 1600, 1700

$$\bar{X} = \frac{900 + 1000 * 2 + 1200 + 1300 * 2 + 1400 * 2 + 1500 * 3 + 1600 * 3 + 1700}{15} = 1366,67$$

Mediana: 1400€

Moda: 1300 y 1500€

b)

Salario	fi	xi*fi	Xi - X̄	fi(Xi - X̄)	(Xi - X̄) ²	fi(Xi - X̄) ²
900	1	900	466,6666667	466,6666667	217777,7778	217777,7778
1000	2	2000	366,6666667	733,3333333	134444,4444	268888,8889
1200	1	1200	166,6666667	166,6666667	27777,77778	27777,77778
1300	2	2600	66,66666667	133,3333333	4444,444444	8888,88889
1400	2	2800	33,33333333	66,66666667	1111,111111	2222,222222
1500	3	4500	133,3333333	400	17777,77778	53333,33333
1600	3	4800	233,3333333	700	54444,44444	163333,3333
1700	1	1700	333,3333333	333,3333333	111111,1111	111111,1111
	15	20500		3000		853333,3333
	N	$\sum_{i=1}^n fi * Xi$		$\sum_{i=1}^n fi Xi - \bar{X} $		$\sum_{i=1}^n fi (Xi - \bar{X})^2$

Desviación media = $3000/15 = 200$

Varianza (σ^2) = $853333,3/15 = 56888,9$

Desviación típica (σ): $\sqrt{56888,9} = 238,5$

Coefficiente de Variación CV = $238,5/1366,7 = 0,1745 = 17,45\%$

C) Comparativa

- El salario de los hombres es mas alto
- La desviacion media es mas alta en los hombres
- El coeficiente de variacion es mayor en los hombres