ESTADÍSTICA

Tablas de frecuencia. Distribución de frecuencias

La distribución de frecuencias o tabla de frecuencias es una ordenación en forma de tabla de los datos estadísticos, asignando a cada dato su frecuencia correspondiente.

Tipos de frecuencias

Frecuencia absoluta

La frecuencia absoluta es el número de veces que aparece un determinado valor en un estudio estadístico.
Se representa por f_i.
La suma de las frecuencias absolutas es igual al número total de datos, que se representa por N.

$$f_1 + f_2 + f_3 + \ldots + f_n = N$$

Para indicar resumidamente estas sumas se utiliza la letra griega Σ (sigma mayúscula) que se lee suma o sumatorio.

$$\sum_{i=1}^{i=n} f_i = N$$

Frecuencia relativa

La frecuencia relativa es el cociente entre la frecuencia absoluta de un determinado valor y el número total de datos.
Se puede expresar en tantos por ciento y se representa por n_i.

$$n_i = \frac{f_i}{N}$$ La suma de las frecuencias relativas es igual a 1.

Ejemplo:

Durante el mes de julio, en una ciudad se han registrado las siguientes temperaturas máximas:
32, 31, 28, 29, 33, 32, 31, 30, 31, 31, 27, 28, 29, 30, 32, 31, 31, 30, 30, 29, 29, 30, 30, 31, 30, 31, 34, 33, 33, 29, 29.
En la primera columna de la tabla colocamos la variable ordenada de menor a mayor, en la segunda hacemos el recuento y en la tercera anotamos la frecuencia absoluta.

x_i	Recuento	f_i	F_i	n_i	N_i
27	I	1	1	0.032	0.032
28	II	2	3	0.065	0.097
29	~~IIII~~ I	6	9	0.194	0.290
30	~~IIII~~ II	7	16	0.226	0.516
31	~~IIII~~ III	8	24	0.258	0.774
32	III	3	27	0.097	0.871
33	III	3	30	0.097	0.968
34	I	1	31	0.032	1
		31		1	

Diagrama de barras y polígonos de frecuencias

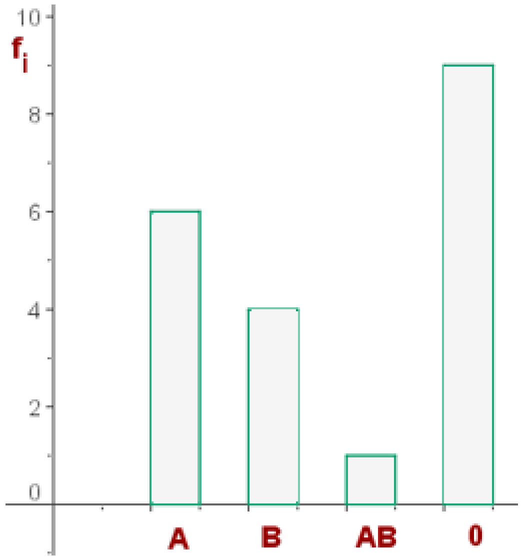
Diagrama de barras

Un diagrama de barras se utiliza para representar datos cualitativos o datos cuantitativos de tipo discreto.

Se representan sobre unos ejes de coordenadas, en el eje de abscisas se colocan los valores de la variable, y sobre el eje de ordenadas las frecuencias absolutas o relativas o acumuladas. Los datos se representan mediante barras de una altura proporcional a la frecuencia.

Ejemplo: _Un estudio hecho al conjunto de los 20 alumnos de una clase para determinar su grupo sanguíneo ha dado el siguiente resultado:

Grupo sanguíneo	f_i
A	6
B	4
AB	1
0	9
	20



Polígonos de frecuencia

Un polígono de frecuencias se forma uniendo los extremos de las barras mediante segmentos. También se puede realizar trazando los puntos que representan las frecuencias y uniéndolos mediante segmentos.

Ejemplo: _Las temperaturas en un día de otoño de una ciudad han sufrido las siguientes variaciones:

Hora	Temperatura
6	7º
9	12°
12	14°
15	11°
18	12°
21	10°
24	8°

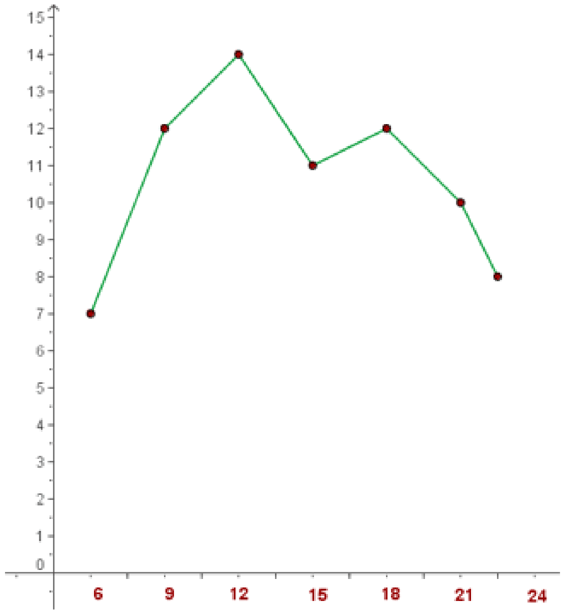


Diagrama de sectores

Un **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.

$$\alpha = \frac{360°}{N} \cdot f_i$$

El diagrama circular se construye con la ayuda de un transportador de ángulos.

Ejemplo: _

En una clase de 30 alumnos, 12 juegan a baloncesto, 3 practican la natación, 9 juegan al fútbol y el resto no practica ningún deporte.

$$\alpha_1 = \frac{360°}{30} \cdot 12 = 144° \qquad \alpha_2 = \frac{360°}{30} \cdot 3 = 36°$$

$$\alpha_3 = \frac{360°}{30} \cdot 9 = 108° \qquad \alpha_4 = \frac{360°}{30} \cdot 6 = 72°$$

	Alumnos	Ángulo
Baloncesto	12	144°
Natación	3	36°

Fútbol	9	108°
Sin deporte	6	72°
Total	30	360°

Histograma

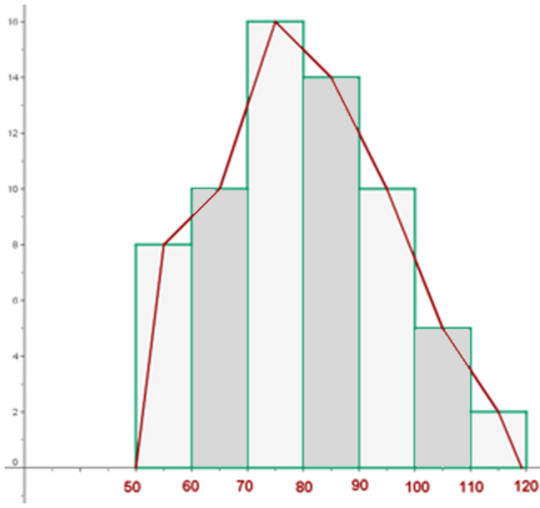
Un **histograma** es una **representación gráfica** de una **variable** en forma de **barras**. En el **eje abscisas** se construyen unos **rectángulos** que tienen por **base la amplitud del intervalo**, y por **altura**, la**frecuencia absoluta** de cada **intervalo**. La **superficie** de cada **barra** es **proporcional** a la **frecuencia** de los **valores** representados.

Polígono de frecuencia

Para construir el **polígono de frecuencia** se toma la **marca de clase** que coincide con el **punto medio** de cada **rectángulo**.

Ejemplo: _El peso de 65 personas adultas viene dado por la siguiente tabla:

	c_i	f_i	F_i
[50, 60)	55	8	8
[60, 70)	65	10	18
[70, 80)	75	16	34
[80, 90)	85	14	48
[90, 100)	95	10	58
[100, 110)	105	5	63
[110, 120)	115	2	65
		65	



Parámetros estadísticos

Un **parámetro estadístico** es un **número** que se obtiene a partir de los **datos** de una **distribución estadística**.

Los **parámetros estadísticos** sirven para sintetizar la información dada por una tabla o por una gráfica.

Tipos de parámetros estadísticos

Hay **tres tipos parámetros estadísticos**: De centralización, de posición y de dispersión.

Medidas de centralización

Nos indican en torno a qué valor (centro) se distribuyen los datos.

La **medidas de centralización** son:

Media aritmética: La **media** es el valor **promedio** de la distribución.

Mediana: La **mediana** es la **puntación** de la escala que **separa la mitad superior** de la distribución y **la inferior**, es decir divide la serie de datos en **dos partes iguales**.

Moda: La **moda** es el **valor** que **más se repite** en una distribución.

Medidas de dispersión

Las **medidas de dispersión** nos informan sobre cuanto se alejan del centro los valores de la distribución.

Las **medidas de dispersión** son:

Rango o recorrido

El **rango** es la **diferencia** entre el **mayor** y el **menor** de los **datos** de una distribución estadística.

Desviación media

La **desviación media** es la **media aritmética** de los **valores absolutos** de las **desviaciones** respecto a la **media**.

Varianza

La **varianza** es la **media aritmética** del **cuadrado de las desviaciones** respecto a la **media**.

Desviación típica

La **desviación típica** es la **raíz cuadrada** de la **varianza**.

Desviación media

Desviación respecto a la media

La **desviación respecto a la media** es la **diferencia** en valor absoluto entre cada **valor** de la variable estadística y la **media aritmética**.

$D_i = |x - \bar{x}|$

Desviación media

La **desviación media** es la **media aritmética** de los **valores absolutos de las desviaciones respecto a la media**.

La **desviación media** se representa por $D_{\bar{x}}$

$$D_{\bar{x}} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \ldots + |x_n - \bar{x}|}{N} \qquad D_{\bar{x}} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{N}$$

Ejemplo: Calcular la **desviación media** de la distribución: 9, 3, 8, 8, 9, 8, 9, 18

$$\bar{x} = \frac{9+3+8+8+9+8+9+18}{8} = 9$$

$$D_{\bar{x}} = \frac{|9-9| + |3-9| + |8-9| + |8-9| + |9-9| + |8-9| + |9-9| + |18-9|}{8} = 2.25$$

Desviación media para datos agrupados

Si los datos vienen agrupados en una **tabla de frecuencias**, la expresión de la **desviación media** es:

$$D_{\bar{x}} = \frac{|x_1 - \bar{x}| f_1 + |x_2 - \bar{x}| f_2 + \ldots + |x_n - \bar{x}| f_n}{N} \qquad D_{\bar{x}} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}| f_i}{N}$$

Ejemplo: Calcular la **desviación media** de la distribución:

	x_i	f_i	$x_i \cdot f_i$	$\lvert x - \bar{x}\rvert$	$\lvert x - \bar{x}\rvert \cdot f_i$
[10, 15)	12.5	3	37.5	9.286	27.858
[15, 20)	17.5	5	87.5	4.286	21.43
[20, 25)	22.5	7	157.5	0.714	4.998
[25, 30)	27.5	4	110	5.714	22.856
[30, 35)	32.5	2	65	10.714	21.428
		21	457.5		98.57

$$\bar{x} = \frac{457.5}{21} = 21.786 \qquad D_{\bar{x}} = \frac{98.57}{21} = 4.69$$

Varianza

La **varianza** es la **media aritmética del cuadrado de las desviaciones respecto a la media** de una distribución estadística. La varianza se representa por σ^2.

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{N} \qquad \sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{N}$$

Varianza para datos agrupados

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + \ldots + (x_n - \bar{x})^2 f_n}{N} \qquad \sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 f_i}{N}$$

Para simplificar el **cálculo de la varianza** vamos o utilizar las siguientes expresiones que son equivalentes a las anteriores.

$$\sigma^2 = \frac{x_1^2 + x_2^2 + \ldots + x_n^2}{N} - \bar{x}^2 \quad \sigma^2 = \sum_{i=1}^{n} \frac{x_i^2}{N} - \bar{x}^2$$

Ejercicios de varianza

Ejercicio 1: **Calcular la varianza** de la distribución: 9, 3, 8, 8, 9, 8, 9, 18

$$\bar{x} = \frac{9+3+8+8+9+8+9+18}{8} = 9$$

$$\sigma^2 = \frac{(9-9)^2 + (3-9)^2 + (8-9)^2 + (8-9)^2 + (9-9)^2 + (8-9)^2 + (9-9)^2 + (18-9)^2}{8} = 15$$

Ejercicio 2: **Calcular la varianza** de la distribución de la tabla:

	x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
[10, 20)	15	1	15	225
[20, 30)	25	8	200	5000
[30,40)	35	10	350	12 250
[40, 50)	45	9	405	18 225
[50, 60	55	8	440	24 200
[60,70)	65	4	260	16 900
[70, 80)	75	2	150	11 250
		42	1 820	88 050

$$\bar{x} = \frac{1820}{42} = 43.33 \qquad \sigma^2 = \frac{88050}{42} - 43.33^2 = 218.94$$

Propiedades de la varianza

1 La **varianza** será siempre un **valor positivo o cero**, en el caso de que las puntuaciones sean iguales.

2 Si a todos los **valores** de la variable se les **suma** un **número** la **varianza no varía**.

3 Si todos los **valores** de la variable se **multiplican** por un **número** la **varianza** queda **multiplicada** por el**cuadrado** de dicho **número**.

4 Si tenemos varias distribuciones con la misma **media** y conocemos sus respectivas **varianzas** se puede calcular la **varianza total**.

Si todas las muestras tienen el mismo tamaño:

$$\sigma^2 = \frac{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2}{n}$$

Si las muestras tienen distinto tamaño:

$$\sigma^2 = \frac{k_1 \cdot \sigma_1^2 + k_2 \cdot \sigma_2^2 + \ldots + k_n \cdot \sigma_n^2}{k_1 + k_2 + \ldots + k_n}$$

Observaciones sobre la varianza

1 La **varianza**, al igual que la media, es un índice muy sensible a las puntuaciones extremas.

2 En los casos que **no se pueda hallar la media** tampoco será posible hallar la **varianza**.

3 La **varianza** no viene expresada en las mismas unidades que los datos, ya que las desviaciones están elevadas al cuadrado.

Desviación típica

La **desviación típica** es la **raíz cuadrada de la varianza**. Es decir, la raíz cuadrada de la media de los cuadrados de las puntuaciones de desviación. La **desviación típica** se representa por **σ**.

$$\sigma = \sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots+(x_n-\bar{x})^2}{N}} \qquad \sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{N}}$$

Desviación típica para datos agrupados

$$\sigma = \sqrt{\frac{(x_1-\bar{x})^2 f_1+(x_2-\bar{x})^2 f_2+\ldots+(x_n-\bar{x})^2 f_n}{N}} \qquad \sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2 f_i}{N}}$$

Para simplificar el cálculo vamos o utilizar las siguientes expresiones que son equivalentes a las anteriores.

$$\sigma = \sqrt{\frac{x_1^2+x_2^2+\ldots+x_n^2}{N}-\bar{x}^2} \qquad \sigma = \sqrt{\sum_{i=1}^{n}\frac{x_i^2}{N}-\bar{x}^2}$$

$$\sigma = \sqrt{\frac{x_1^2 f+x_2^2 f_2+\ldots+x_n^2 f_n}{N}-\bar{x}^2} \qquad \sigma = \sqrt{\sum_{i=1}^{n}\frac{x_i^2 f_i}{N}-\bar{x}^2}$$

Ejercicios de desviación típica

Ejercicio 1: Calcular la **desviación típica** de la distribución: 9, 3, 8, 8, 9, 8, 9, 18

$$\bar{x} = \frac{9+3+8+8+9+8+9+18}{8} = 9$$

$$\sigma = \sqrt{\frac{(9-9)^2+(3-9)^2+(8-9)^2+(8-9)^2+(9-9)^2+(8-9)^2+(9-9)^2+(18-9)^2}{8}} = 3.87$$

Ejercicio 2: **Calcular la desviación típica** de la distribución de la tabla:

	x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
[10, 20)	15	1	15	225
[20, 30)	25	8	200	5000
[30,40)	35	10	350	12 250
[40, 50)	45	9	405	18 225
[50, 60)	55	8	440	24 200
[60,70)	65	4	260	16 900
[70, 80)	75	2	150	11 250
		42	1 820	88 050

$$\bar{x} = \frac{1820}{42} = 43.33 \qquad \sigma = \sqrt{\frac{88050}{42}-43.33^2} = 14.797$$

Propiedades de la desviación típica

1 La **desviación típica** será siempre un **valor positivo o cero**, en el caso de que las puntuaciones sean iguales.
2 Si a todos los **valores** de la variable se les **suma** un **número** la **desviación típica no varía**.
3 Si todos los **valores** de la variable se **multiplican** por un **número** la **desviación típica** queda **multiplicada** por dicho **número**.
4 Si tenemos varias distribuciones con la misma **media** y conocemos sus respectivas **desviaciones típicas** se puede calcular la **desviación típica total**.
Si todas las muestras tienen el mismo tamaño:

$$\sigma = \sqrt{\frac{\sigma_1^2+\sigma_2^2+\ldots+\sigma_n^2}{n}}$$

Si las muestras tienen distinto tamaño:

$$\sigma = \sqrt{\frac{k_1 \cdot \sigma_1^2 + k_2 \cdot \sigma_2^2 + \ldots + k_n \cdot \sigma_n^2}{k_1 + k_2 + \ldots + k_n}}$$

Observaciones sobre la desviación típica

1 La **desviación típica**, al igual que la media y la varianza, es un índice muy sensible a las puntuaciones extremas.
2 En los casos que **no se pueda hallar la media** tampoco será posible hallar la **desviación típica**.
3 Cuanta más pequeña sea la **desviación típica** mayor será la **concentración de datos** alrededor de la **media**.

Coeficiente de variación

El **coeficiente de variación** es la relación entre la **desviación típica** de una muestra y su **media**. $C.V = \frac{\sigma}{\overline{X}}$

El **coeficiente de variación** se suele expresar en **porcentajes**: $C.V = \frac{\sigma}{\overline{X}} \cdot 100$

El **coeficiente de variación** permite comparar las **dispersiones** de dos distribuciones distintas, siempre que sus **medias** sean **positivas**. Se calcula para cada una de las distribuciones y los valores que se obtienen se comparan entre sí. La **mayor dispersión** corresponderá al valor del **coeficiente de variación mayor**.

Ejercicio: Una distribución tiene x = 140 y σ = 28.28 y otra x = 150 y σ = 24. ¿Cuál de las dos presenta mayor dispersión?

$C.V_1 = \frac{28.28}{140} \cdot 100 = 20.2\%$ $\qquad C.V_2 = \frac{24}{150} \cdot 100 = 16\%$

La primera distribución presenta mayor dispersión.

Resumen de Estadística descriptiva

La **Estadística** trata del recuento, ordenación y clasificación de los datos obtenidos por las observaciones, para poder hacer comparaciones y sacar conclusiones.

Conceptos de Estadística

Población: Una **población** es el conjunto de todos los elementos a los que se somete a un estudio estadístico.
Individuo: Un **individuo** o **unidad estadística** es cada uno de los elementos que componen la población.
Muestra: Una **muestra** es un conjunto representativo de la población de referencia, el número de individuos de una muestra es menor que el de la población.
Muestreo: El **muestreo** es la reunión de datos que se desea estudiar, obtenidos de una proporción reducida y representativa de la población.
Valor: Un **valor** es cada uno de los distintos resultados que se pueden obtener en un estudio estadístico. Si lanzamos una moneda al aire 5 veces obtenemos dos valores: cara y cruz.
Dato: Un **dato** es cada uno de los valores que se ha obtenido al realizar un estudio estadístico. Si lanzamos una moneda al aire 5 veces obtenemos 5 datos: cara, cara, cruz, cara, cruz.

Variables estadísticas

Variable cualitativa: Las **variables cualitativas** se refieren a **características o cualidades** que **no** pueden ser medidas con **números**. Podemos distinguir dos tipos:
Variable cualitativa nominal: Una **variable cualitativa nominal** presenta **modalidades no numéricas** que **no** admiten un **criterio de orden**.
Variable cualitativa ordinal o variable cuasi cuantitativa: Una **variable cualitativa ordinal** presenta **modalidades no numéricas**, en las que existe un **orden**.
Variable cuantitativa: Una **variable cuantitativa** es la que se expresa mediante un **número**, por tanto se pueden realizar o**peraciones aritméticas** con ella. Podemos distinguir dos tipos:
Variable discreta: Una **variable discreta** es aquella que toma **valores aislados**, es decir **no** admite **valores intermedios** entre dos valores específicos.
Variable continua: Una **variable continua** es aquella que puede tomar **valores comprendidos entre dos números**.
Distribución de frecuencias: La **distribución de frecuencias** o **tabla de frecuencias** es una **ordenación** en forma de **tabla** de los **datos estadísticos**, asignando a cada **dato** su **frecuencia correspondiente**.
Diagrama de barras: Un **diagrama de barras** se utiliza para de presentar **datos cualitativos** o **datos cuantitativos de tipo discreto**. Los **datos** se representan mediante **barras** de una **altura proporcional** a la **frecuencia**.

Polígonos de frecuencias: Un **polígono de frecuencias** se forma uniendo los **extremos** de las **barras** mediante **segmentos.** También se puede realizar trazando los **puntos** que representan las **frecuencias** y uniéndolos mediante **segmentos**.

Diagrama de sectores: Un **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.

$$\alpha = \frac{360°}{N} \cdot f_i$$

Histograma: Un **histograma** es una **representación gráfica** de una **variable** en forma de **barras**. Se utilizan para **variables continuas** o para **variables discretas**, con un gran número de datos, y que se han agrupado en **clases**. En el **eje abscisas** se construyen unos **rectángulos** que tienen por **base la amplitud del intervalo**, y por **altura**, la**frecuencia absoluta** de cada **intervalo**.

Medidas de centralización

Moda: La **moda** es el **valor** que tiene **mayor frecuencia absoluta**. Se representa por $\mathbf{M_o}$. Se puede hallar la **moda** para **variables cualitativas** y **cuantitativas**.

Cálculo de la moda para datos agrupados

1º Todos los intervalos tienen la misma amplitud.

$$Mo = L_i + \frac{f_{i+1}}{f_{i-1} + f_{i+1}} \cdot a_i$$

2º Los intervalos tienen amplitudes distintas.

En primer lugar tenemos que hallar las alturas.

$$h_i = \frac{f_i}{a_i}$$

La clase modal es la que tiene mayor altura.

$$Mo = L_i + \frac{h_{i+1}}{h_{i-1} + h_{i+1}} \cdot a_i$$

Mediana: Es el **valor** que ocupa el **lugar central** de todos los **datos** cuando éstos están **ordenados de menor a mayor**. La **mediana** se representa por $\mathbf{M_e}$. La **mediana** se puede **hallar** sólo para **variables cuantitativas**.

Cálculo de la mediana

1 Ordenamos los **datos** de **menor a mayor**.

2 Si la serie tiene un **número impar de medidas** la **mediana** es la **puntuación central** de la misma.

3 Si la serie tiene un **número par** de puntuaciones la **mediana** es la **media** entre las dos **puntuaciones centrales**.

Cálculo de la mediana para datos agrupados: La **mediana** se encuentra en el **intervalo** donde la **frecuencia acumulada** llega hasta la **mitad de la suma de las frecuencias absolutas**.

Es decir tenemos que buscar el intervalo en el que se encuentre $\frac{N}{2}$.

$$Me = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} \cdot a_i$$

Media aritmética: La **media aritmética** es el **valor** obtenido al **sumar** todos los **datos** y **dividir** el resultado entre el **número** total de**datos**.

$\bar{x}$ es el símbolo de la **media aritmética**.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{N}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{N}$$

Media aritmética para datos agrupados

Si los **datos** vienen **agrupados** en una tabla de frecuencias, la expresión de la **media** es:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \ldots + x_n f_n}{N}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{N}$$

Medidas de dispersión

Desviación media: La **desviación media** es la **media aritmética** de los **valores absolutos de las desviaciones respecto a la media**. La **desviación media** se representa por $D_{\bar{x}}$

$$D_{\bar{x}} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \ldots + |x_n - \bar{x}|}{N}$$

$$D_{\bar{x}} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{N}$$

Desviación media para datos agrupados: Si los datos vienen agrupados en una **tabla de frecuencias**, la expresión de la **desviación media** es:

$$D_{\bar{x}} = \frac{|x_1 - \bar{x}| f_1 + |x_2 - \bar{x}| f_2 + \ldots + |x_n - \bar{x}| f_n}{N}$$

$$D_{\bar{x}} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}| f_i}{N}$$

Varianza: La **varianza** es la **media aritmética del cuadrado de las desviaciones respecto a la media** de una distribución estadística. La varianza se representa por σ^2.

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{N}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{N}$$

Varianza para datos agrupados

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + \ldots + (x_n - \bar{x})^2 f_n}{N} \quad \sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 f_i}{N}$$

Para simplificar el **cálculo de la varianza** vamos o utilizar las siguientes expresiones que son equivalentes a las anteriores.

$$\sigma^2 = \frac{x_1^2 + x_2^2 + \ldots + x_n^2}{N} - \bar{x}^2 \quad \sigma^2 = \sum_{i=1}^{n} \frac{x_i^2}{N} - \bar{x}^2$$

Desviación típica: La **desviación típica** es la **raíz cuadrada de la varianza**.

La **desviación típica** se representa por **σ**.

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{N}} \quad \sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{N}}$$

Desviación típica para datos agrupados

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + \ldots + (x_n - \bar{x})^2 f_n}{N}} \quad \sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 f_i}{N}}$$

Para simplificar el cálculo vamos o utilizar las siguientes expresiones que son equivalentes a las anteriores.

$$\sigma = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_n^2}{N} - \bar{x}^2} \quad \sigma = \sqrt{\sum_{i=1}^{n} \frac{x_i^2 f_i}{N} - \bar{x}^2}$$

Desviación típica para datos agrupados

$$\sigma = \sqrt{\frac{x_1^2 f + x_2^2 f_2 + \ldots + x_n^2 f_n}{N} - \bar{x}^2} \qquad \sigma = \sqrt{\sum_{i=1}^{n} \frac{x_i^2}{N} - \bar{x}^2}$$

Coeficiente de variación: El **coeficiente de variación** es la relación entre la **desviación típica** de una muestra y su **media**.

$$C.V = \frac{\sigma}{\bar{x}}$$

$$C.V = \frac{\sigma}{\bar{x}} \cdot 100$$

Ejercicio de estadística

Las temperaturas máximas en una ciudad durante el mes de junio fueron: 28 ºC, 29 ºC, 28 ºC, 30 ºC, 30 ºC, 29 ºC, 30 ºC, 31 ºC, 29 ºC, 29 ºC, 30 ºC, 31 ºC, 31 ºC, 31 ºC, 32 ºC, 33 ºC, 34 ºC, 34 ºC, 35 ºC, 31 ºC, 31 ºC, 32 ºC, 32 ºC, 33 ºC, 33 ºC, 31 ºC, 32 ºC, 32 ºC, 33 ºC, 33 ºC, 34 ºC.

Calcula la moda:

Calcula la mediana:

Calcula la media:

Calcula el rango:

Calcula la desviación media:

Calcula la varianza:

Calcula la desviación típica:

Ejercicios resueltos de la desviación típica

1. Hallar la **desviación media, la varianza y la desviación típica** de la series de números siguientes:

2, 3, 6, 8, 11.

12, 6, 7, 3, 15, 10, 18, 5.

2, 3, 6, 8, 11.

Media

$$\bar{x} = \frac{2+3+6+8+11}{5} = 6$$

Desviación típica

$$\sigma = \sqrt{\frac{2^2+3^2+6^2+8^2+11^2}{5} - 6^2} = 2.61$$

12, 6, 7, 3, 15, 10, 18, 5.

Media

$$\bar{x} = \frac{12+6+7+3+15+10+18+5}{8} = \frac{76}{8} = 9.5$$

Desviación típica

$$\sigma = \sqrt{\frac{12^2+6^2+7^2+3^2+15^2+10^2+18^2+5^2}{8} - 9.5^2} = 3.35$$

2. Un pediatra obtuvo la siguiente tabla sobre los meses de edad de 50 niños de su consulta en el momento de andar por primera vez:

Meses	Niños
9	1
10	4
11	9
12	16

13	11
14	8
15	1

Calcular la **desviación típica**.

x_i	f_i	N_i	$x_i \cdot f_i$	$x^2_i \cdot f_i$
9	1	1	9	81
10	4	5	40	400
11	9	14	99	1089
12	16	30	192	2304
13	11	41	143	1859
14	8	49	112	1568
15	1	50	15	225
	50		610	7526

$$\sigma = \sqrt{\frac{7526}{50} - 12.2^2} = 1.30$$

3. El resultado de lanzar dos dados 120 veces viene dado por la tabla**:**

Sumas	Veces
2	3
3	8
4	9
5	11
6	20
7	19
8	16
9	13
10	11
11	6
12	4

Calcular la **desviación típica**.

x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
2	3	6	12
3	8	24	72
4	9	36	144
5	11	55	275
6	20	120	720
7	19	133	931
8	16	128	1024
9	13	117	1053
10	11	110	1100

11	6	66	726
12	4	48	576
	120	843	6633

$$\bar{x} = \frac{843}{120} = 7.025 \qquad \sigma = \sqrt{\frac{6633}{120} - 7.025^2} = 2.434$$

EJEMPLOS CON DATOS AGRUPADOS EN INTERVALOS DE VALORES

Ejemplo 3

Intervalos	x_i	n_i	$x_i \cdot n_i$
(18 - 22]	20	2	40
(22 - 26]	24	14	336
(26 - 30]	28	12	336
(30 - 34]	32	12	384
(34 - 38]	36	14	504
(38 - 42]	40	6	240
		60	1840

$$\bar{x} = \frac{1840}{60} = 30{,}67 \text{ años}$$

Intervalos	x_i	n_i	$\lvert x_i - \bar{x} \rvert$	$\lvert x_i - \bar{x} \rvert \cdot n_i$
(18 - 22]	20	2	10,67	21,33
(22 - 26]	24	14	6,67	93,33
(26 - 30]	28	12	2,67	32,00
(30 - 34]	32	12	1,33	16,00
(34 - 38]	36	14	5,33	74,67
(38 - 42]	40	6	9,33	56,00
		60		293,33

$$\text{D.M.} = \frac{293{,}33}{60} = 4{,}89 \text{ años}$$

Intervalos	x_i	n_i	$x_i^2 n_i$
(18 - 22]	20	2	800
(22 - 26]	24	14	8064
(26 - 30]	28	12	9408
(30 - 34]	32	12	12288
(34 - 38]	36	14	18144
(38 - 42]	40	6	9600
		60	58304

$$s^2 = \frac{58304}{60} - 30{,}67^2 = 971{,}73 - 940{,}65 = 31{,}08 \text{ años}^2$$

$$s = \sqrt{31{,}08} = 5{,}57 \text{ años}$$

EXÁMENES

1. Valencia-2011-RESUELTO

Las frecuencias del número de asignaturas suspendidas en una clase de 20 alumnos es:

X_i (número de asignaturas suspendidas)	**0**	**1**	**2**	**3**	**4**
F_i (frecuencias)	7	3	4	1	5

Calcula: la media, la mediana y la moda de la distribución.

Xi	fi	xi*fi
0	7	0
1	3	3
2	4	8
3	1	3
4	5	20

$$Media: X_i = \frac{\sum_{i=1}^{n} f_i \cdot x_i}{N} = \frac{0+3+8+3+20}{20} = \frac{34}{20} = 1.7$$

$$Moda = 0 \Rightarrow se \quad repite \quad 7 \quad veces \quad y \qquad Mediana: \frac{X_{10}+X_{11}}{2} = \frac{1+2}{2} = 1.5$$

Si la serie tiene un número par de puntuaciones la mediana es la media entre las dos puntuaciones centrales

---

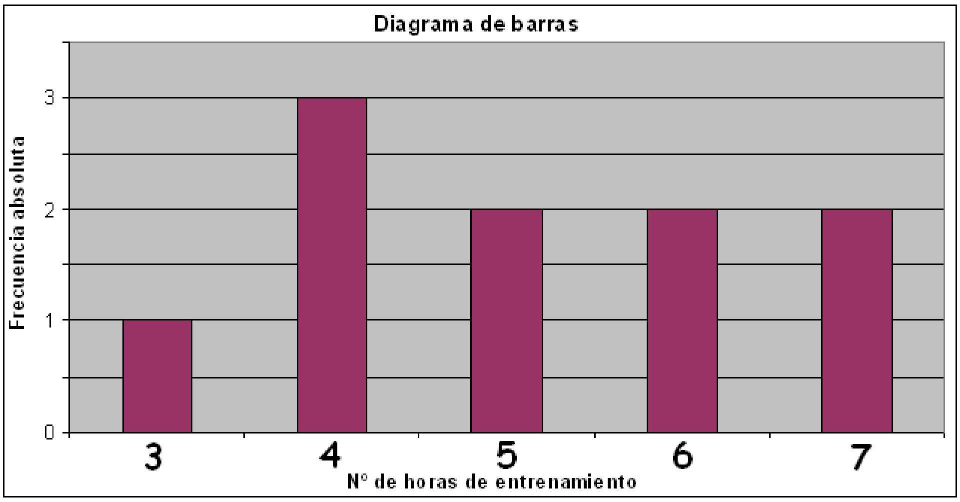
2. Andalucía-2013-RESUELTO

4. El número de horas diarias que entrena un grupo de 10 ciclistas es: 5, 6, 4, 7, 5, 4, 7, 6, 4, 3.. (2,5 puntos, 0,5 por apartado A, B y 1,5 por apartado C)

A. Organiza la información en la siguiente tabla:

Nº de horas de entrenamiento	Frecuencia absoluta
3	1
4	3
5	2
6	2
7	2

B. Representa esta información en un diagrama de barras.



C. Calcula la media y la desviación típica del número de horas de entrenamiento.

Media $\bar{x} = \frac{\Sigma(f_i \cdot x_i)}{N} = \frac{1\cdot3+3\cdot4+2\cdot5+2\cdot6+2\cdot7}{10} = \frac{51}{10} = 5{,}1$ horas de entrenamiento

Desviación típica:

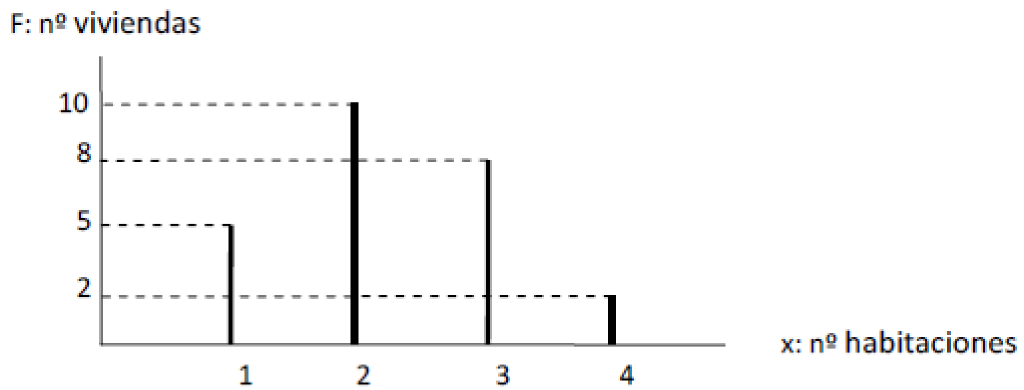
$$\sigma^2 = \sqrt{\frac{\Sigma f_i \cdot (x_i-\bar{x})^2}{N}} = \sqrt{\frac{1\cdot(3-5{,}1)^2+3\cdot(4-5{,}1)^2+2\cdot(5-5{,}1)^2+2\cdot(6-5{,}1)^2+2\cdot(7-5{,}1)^2}{10}} =$$

$$= \sqrt{\frac{4{,}41+3{,}63+0{,}02+1{,}62+7{,}22}{10}} = \sqrt{\frac{18{,}7}{10}} = \sqrt{1{,}87} = 1{,}37$$

---

3. Valencia-2013

En un estudio sobre determinadas características sociológicas de un barrio, elegimos aleatoriamente 25 viviendas del mismo y computamos el número de habitaciones de cada una de ellas. El resultado viene representado en el siguiente diagrama de barras:



Calcular: La media, la mediana y la moda del número de habitaciones de la muestra.

---

4. Valencia-2012

La distribución de las multas por infracciones de tráfico en una ciudad A a lo largo de un determinado período de tiempo viene dada por la relación:

400 multas de 50 €
250 multas de 120 €
150 multas de 200 €.
Calcular:
El valor medio de las multas por dichas infracciones.
La varianza y la desviación típica de dicha distribución de sanciones

---

5. Valencia-2010

Se ha realizado un estudio estadístico en un gran centro comercial sobre el dinero que un/a cliente/a gasta al realiza sus compras en un día cualquiera de la semana. Este estudio nos aporta la siguiente información:

Dinero (€)	[0-100[	[100-200[	[200-300[	[300-400[	[400-500[
Nº personas	1000	1100	1600	1000	300

a) Halla el gasto medio realizado por los clientes ese día.
b) Si a todas las personas que gastan más de 300 euros se les obsequia con un regalo ¿cuál es el porcentaje de clientes que reciben dicho regalo?

---

6. Castilla-La mancha 2012

Lanzamos dos dados, sumamos las puntuaciones y anotamos los resultados. Repetimos la experiencia 30 veces:
11, 8, 9, 9, 3, 4, 11, 7, 7, 8, 7, 5, 6, 4, 4, 7, 10, 2, 6, 10, 7, 7, 6, 2, 8, 7, 5, 8, 6, 9
a) Confecciona una tabla de frecuencias.
b) Calcula los siguientes parámetros estadísticos:

1. Media aritmética.
2. Moda.
3. Mediana.
4. Varianza.
5. Desviación típica.

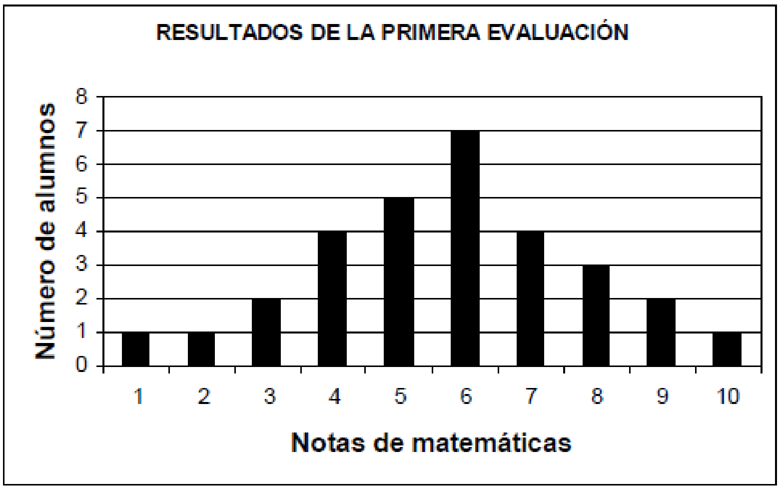
---

7. El porcentaje de población activa dedicada a la agricultura en 30 países africanos es:
47 24 70 63 91 61 63 75 56 57 68 74 77 69 68 70 75 64 37 36 65 91 62 14 66 81 24 66 63 43
a) Agrupa estos datos en cinco intervalos de igual amplitud.
b) Calcula la media, moda y mediana.
c) Calcula la varianza, la desviación típica y el coeficiente de variación.

---

8. Extremadura-2009

Las calificaciones que obtuvieron los alumnos de una clase en matemáticas, en la primera evaluación, se encuentran recogidas en el siguiente diagrama de barras:



a) Elaborar la tabla de frecuencias correspondiente.
b) Calcular la media.
c) Indicar cuál es la moda.
d) Calcular la desviación típica.
e) Determinar el porcentaje de aprobados en matemáticas (notas mayores o iguales a 5).

9. Extremadura-2010
Las estaturas de cinco personas en cm han sido: **160 170 165 155 160**
a) Elaborar una tabla de frecuencias.
b) Calcular la estatura media.
c) Calcular la desviación típica.
d) ¿Cuál es la moda?
e) ¿Qué porcentaje de personas están por debajo de la estatura media?

10. Extremadura-2008
El gasto mensual en Telefonía Móvil de 10 alumnos de 4º de ESO es el siguiente:
8, 8, 10, 10,10, 14, 18, 18, 22, 22 (En euros)
a) Elabora una tabla de frecuencias absolutas y relativas.
b) Halla la media y la desviación típica de la distribución.
c) Dibuja un gráfico estadístico adecuado.

11. Cataluña-2014
Fem una enquesta sobre el nombre de llibres llegits per cada persona entrevistada durant l'últim any, i obtenim els resultats següents: 0, 2, 1, 1, 1, 2, 3, 0, 0, 1, 2, 1, 4, 5, 0, 3, 2, 4, 3, 2, 0, 0, 2, 2, 2, 1, 3, 0, 1, 4, 2, 3, 1, 1, 1, 2, 2, 3, 1, 3
***a*)** Feu una taula amb la freqüència absoluta, la freqüència relativa i el tant per cent.
***b*)** Calculeu-ne la mitjana aritmètica i la desviació típica.
$\bar{x}$ = **1,775** σ = **1,274**
***c*)** Quin tant per cent de persones llegeixen tres o més llibres anualment?
17,5 % + 7,5 % + 2,5 % = **27,5 %**

12. Castilla-La mancha-2014
Durante el mes de Julio, en una determinada ciudad se han registrado las siguientes temperaturas máximas: 32, 31, 28, 29, 29, 33, 32, 31, 30, 31, 31, 27, 29, 29, 30, 32, 31, 31, 30, 30, 29, 29, 30, 30, 31, 30, 31, 34, 33, 33, 29.
a) Elabora una tabla de frecuencias y representa la distribución mediante un diagrama de barras.
b) Halla la moda, media y mediana.
c) Halla la varianza y la desviación típica

13. Madrid-2014
En un hospital se quiere estimar el peso de las niñas recién nacidas. Para ello se seleccionan, de forma aleatoria, cien de estas, obteniéndose los siguientes resultados

Intervalos (Kg)	[1;1,5)	[1,5;2)	[2;2,5)	[2,5;3)	[3;3,5)	[3,5;4)	[4;4,5)	[4,5;5)
Nº de niños	1	2	5	20	40	26	5	1

Calcule

a) La media, la moda, la mediana y la desviación típica.

b) El porcentaje de niñas con un peso superior a 3 Kg.

---

14. Castilla-La Mancha-2013

Los siguientes valores representan los pesos de una serie de personas:

63 75 80 89 65 74 72 69 82 91 96 105 67 82 86 87 78 65 94 93 94 78 76 106 100 70 84 82 76 84 94 102 68 64 82

a) Agrupa los datos en intervalos de amplitud 10, halla las marcas de clase y realiza una tabla estadística con los datos.

b) Calcula la media, mediana, moda, varianza y desviación típica

c) Realiza el diagrama de barras de los datos y el polígono de frecuencias

---

15. Castilla-La Mancha-2013

Estudiamos el número de televisores que hay en cada vivienda y obtenemos los siguientes datos: 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 3, 3, 2, 2, 1, 1, 1, 2, 2, 3, 3, 3, 4, 4, 3, 3, 2, 2, 2, 1, 4, 1, 1, 1, 2, 3, 1, 1, 2

a. Construye la tabla de frecuencias

b. Calcula la media, moda y mediana

c. Calcula la varianza, desviación típica y coeficiente de variación.
