

ESTADÍSTICA

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

Un estudio estadístico consta de las siguientes fases:

- ✓ Recogida de datos.
- ✓ Organización y representación de datos.
- ✓ Análisis de datos.
- ✓ Obtención de 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

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 operaciones 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 PARA DATOS AGRUPADOS EN INTERVALOS

Las distribuciones de frecuencias son tablas que resumen los datos originales en frecuencias.

Los tipos de frecuencia pueden ser:

- Frecuencia Absoluta (f).-

Es el número de veces que se repite el valor de cada variable. La suma de frecuencias absolutas es siempre al total de datos observados.

- Frecuencia Relativa (fr).-

Indica la proporción con que se repite un valor. Es el cociente entre la frecuencia absoluta y el número total de datos. La suma de las frecuencias relativas es siempre 1

- Frecuencia Acumulada (fa).-

Indica el número de valores que son menores o iguales que el valor dado. Es la suma de la frecuencia absoluta primera con la segunda, este valor con la tercera, y así sucesivamente.

- Frecuencia Porcentual (f%).-

Llamada también frecuencia relativa porcentual. Se obtiene multiplicando la frecuencia relativa por 100. La suma de las frecuencias porcentuales es siempre 100%. Se calcula así:

$$f\% = fr \cdot 100$$

- Frecuencia Relativa Acumulada (fra).-

Es la suma de la frecuencia relativa primera con la segunda, este valor con la tercera, y así sucesivamente.

- Frecuencia Relativa Acumulada Porcentual (fra%).-

Indica el número de valores que son menores o iguales que el valor dado. Se obtiene multiplicando la frecuencia relativa acumulada por 100.

REGLAS GENERALES PARA FORMAS DISTRIBUCIONES DE FRECUENCIAS PARA DATOS AGRUPADOS EN INTERVALOS

Cuando los datos contienen una gran cantidad de elementos, para facilitar los cálculos es necesario agruparlos, a estos grupos se los llama intervalos o clases. Un intervalo es una serie de números incluidos entre dos extremos, así por ejemplo, el intervalo 40 – 45 está formado por 40, 41, 42, 43, 44 y 45, siendo 40 el límite inferior, 45 el límite superior, 39,5 límite real inferior (límite inferior disminuido en 5 décimas) y 40,5 el límite real superior (límite superior aumentado en 5 décimas).

Las reglas generales para formas distribuciones de frecuencias para datos agrupados en intervalos son:

1) Calcule el Rango (R).- También se llama recorrido o amplitud total. Es la diferencia entre el valor mayor y el menor de los datos.

$$R = x_{m\acute{a}x} - x_{m\acute{i}n}$$

2) Seleccione el Número de Intervalos de Clase (n_i).- No debe ser menor de 5 y mayor de 12, ya que un número mayor o menor de clases podría oscurecer el comportamiento de los datos. Para calcular el número de intervalos se aplica la regla de Sturges:

$$n_i = 1 + 3,32 \cdot \log(n)$$

Siendo n el tamaño de la muestra.

3) Calcule el Ancho del Intervalo (i).- Se obtiene dividiendo el Rango para el número de intervalos

$$i = \frac{R}{n_i}$$

Cuando el valor de i no es exacto, se debe redondear al valor superior más cercano. Esto altera el valor de rango por lo que es necesario efectuar un ajuste así:

$$\text{Nuevo } R = n_i \cdot i$$

Por ejemplo:

Si una distribución de 40 datos el valor mayor es 41 y el menor es 20 se tiene:

Calculando el Rango se obtiene:

$$R = x_{m\acute{a}x} - x_{m\acute{i}n} = 41 - 20 = 21$$

Calculando el número de intervalos se obtiene:

$$n_i = 1 + 3,32 \cdot \log(n) = 1 + 3,32 \cdot \log 40 = 6,32 = 6$$

Calculando el ancho se obtiene:

$$i = \frac{R}{ni} = \frac{21}{6} = 3,5$$

Redondeando se obtiene: $i = 4$

Calculando el nuevo rango se obtiene:

$$\text{Nuevo } R = ni \cdot i = 6 \cdot 4 = 24$$

El exceso de 3 que se tiene en este caso se distribuye entre x_{\max} y x_{\min} . Por lo general se agrega al mayor y se quita al menor. Como por ejemplo, se podría agregar 2 al valor mayor y quitar 1 al valor menor, obteniéndose los siguientes nuevos valores:

$$\begin{aligned} x_{\max} &= 41 + 2 = 43 \\ x_{\min} &= 20 - 1 = 19 \end{aligned}$$

O también se podría agregar 1 al valor mayor y quitar 2 al valor menor, obteniéndose los siguientes nuevos valores:

$$\begin{aligned} x_{\max} &= 41 + 1 = 42 \\ x_{\min} &= 20 - 2 = 18 \end{aligned}$$

4) Forme los Intervalos de Clase agregando $i-1$ al límite inferior de cada clase, comenzando por el X_{\min} del rango.

5) Se realiza el Conteo de Datos que cae dentro de cada clase (frecuencia absoluta)

6) Calcule la Marca de Clase (x_m).- Es el valor medio de cada clase, se obtiene sumando los límites superior (L_s) e inferior (L_i) del intervalo y dividiendo ésta suma entre 2

$$x_m = \frac{L_s + L_i}{2}$$

7) Calcule las Frecuencias

EJEMPLO ILUSTRATIVO

A 40 estudiantes se les pidió que estimen el número de horas que habrían dedicado a estudiar la semana pasada (tanto en clase como fuera de ella), obteniéndose los siguientes resultados:

36	30	47	60	32	35	40	50
54	35	45	52	48	58	60	38
32	35	56	48	30	55	49	39
58	50	65	35	56	47	37	56
58	50	47	58	55	39	58	45

Solución:

1) Calculando el Rango se obtiene:

$$R = x_{\max} - x_{\min} = 65 - 30 = 35$$

2) Calculando el número de intervalos se obtiene:

$$n_i = 1 + 3,32 \cdot \log(n) = 1 + 3,32 \cdot \log 40 = 6,32 = 6$$

3) Calculando el ancho se obtiene:

$$i = \frac{R}{ni} = \frac{35}{6} = 5,83$$

Redondeando se obtiene: $i = 6$, por lo que es necesario realizar un ajuste al rango.

4) Calculando el nuevo rango se obtiene:

$$\text{Nuevo } R = ni \cdot i = 6 \cdot 6 = 36$$

El exceso de 1 que se tiene en este caso se distribuye entre x_{\max} y x_{\min} . En este ejemplo, se podría agregar 1 al valor mayor y no quitar nada al valor menor, o no agregar nada al mayor y quitar 1 al menor. Al elegir la primera opción se obtiene:

$$\begin{aligned} x_{\max} &= 65 + 1 = 66 \\ x_{\min} &= 30 - 0 = 30 \end{aligned}$$

5) Formando los intervalos de clase agregando $i-1$ ($6-1=5$) al límite inferior de cada clase, comenzando por el X_{\min} del rango se obtiene:

$30+5 = 35$; $36+5 = 41$; $42+5 = 47$; $48+5 = 53$; $54+5 = 59$; $60+5 = 65$

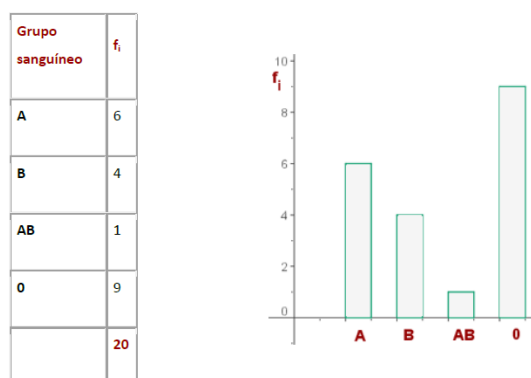
DIAGRAMA DE BARRAS

Un diagrama de barras se utiliza para de presentar 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:



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:

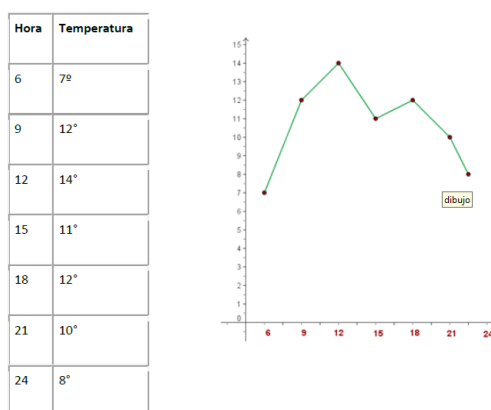


DIAGRAMA DE SECTORES

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.

$$\alpha = \frac{360^\circ}{N} \cdot f_i$$

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, 4 juegan al fútbol y el resto no practica ningún deporte.

$$\alpha_1 = \frac{360^\circ}{30} \cdot 12 = 144^\circ$$

$$\alpha_2 = \frac{360^\circ}{30} \cdot 3 = 36^\circ$$

$$\alpha_3 = \frac{360^\circ}{30} \cdot 9 = 108^\circ$$

$$\alpha_4 = \frac{360^\circ}{30} \cdot 6 = 72^\circ$$

	Alumnos	Ángulo
Baloncesto	12	144°
Natación	3	36°
Fútbol	9	108°
Sin deporte	6	72°
Total	30	360°



5

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.

La superficie de cada barra es proporcional a la frecuencia de los valores representados.

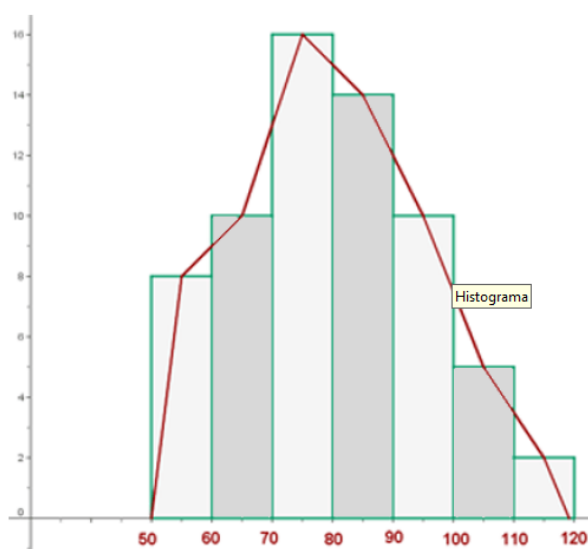
POLÍGONO DE FRECUENCIA

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

Ejemplo

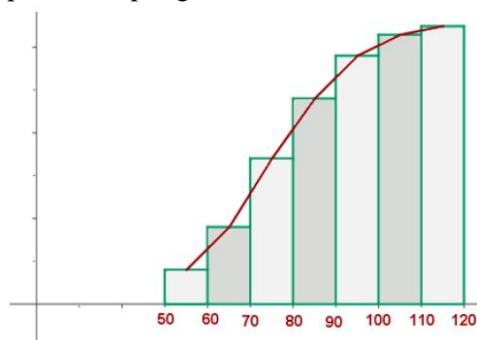
El peso de 65 personas adultas viene dado por la siguiente tabla:

	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)	110	5	63
[110, 120)	115	2	65



HISTOGRAMA Y POLÍGONO DE FRECUENCIAS ACUMULADAS

Si se representan las frecuencias acumuladas de una tabla de datos agrupados se obtiene el histograma de frecuencias acumuladas o su correspondiente polígono.



PARÁMETRO ESTADÍSTICO

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
- De dispersión.

MEDIDAS DE CENTRALIZACIÓN (Medidas De Tendencia Central)

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

La medidas de centralización son:

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**.

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

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

Ejemplo

Los pesos de seis amigos son: 84, 91, 72, 68, 87 y 78 kg. Hallar el peso medio.

$$\bar{x} = \frac{84 + 91 + 72 + 68 + 87 + 78}{6} = 80 \text{ Kg}$$

Media aritmética para datos agrupados

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

$$\bar{x} = \frac{x_1f_1 + x_2f_2 + x_3f_3 + \dots + x_nf_n}{N}$$

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

Ejercicio de media aritmética

En un test realizado a un grupo de 42 personas se han obtenido las puntuaciones que muestra la tabla. Calcula la puntuación media.

Intervalo	x_i	f_i	$x_i \cdot f_i$
[10, 20)	15	1	15
[20, 30)	25	8	200
[30,40)	35	10	350
[40, 50)	45	9	405
[50, 60)	55	8	440
[60,70)	65	4	260
[70, 80)	75	2	150
		42	1 820

$$\bar{x} = \frac{1820}{42} = 43.33$$

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 Me.

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.

2, 3, 4, 4, 5, 5, 5, 6, 6

Me= 5

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

7, 8, 9, 10, 11, 12

Me= 9.5

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 $\frac{N}{2}$.

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

L_i es el límite inferior de la clase donde se encuentra la mediana.

$\frac{N}{2}$ es la semisuma de las frecuencias absolutas.

F_{i-1} es la frecuencia acumulada anterior a la clase mediana.

a_i es la amplitud de la clase.

La mediana es independiente de las amplitudes de los intervalos.

Ejemplo

Calcular la mediana de una distribución estadística que viene dada por la siguiente tabla:

	f_i	F_i
[60, 63)	5	5
[63, 66)	18	23
[66, 69)	42	65
[69, 72)	27	92
[72, 75)	8	100
	100	

$$100/2 = 50$$

Clase de la mediana: [66, 69)

$$Me = 66 + \frac{50 - 23}{42} \cdot 3 = 67.93$$

MODA

La moda es el valor que tiene mayor frecuencia absoluta.

Se representa por M_o .

Se puede hallar la moda para variables cualitativas y cuantitativas.

Hallar la moda de la distribución:

2, 3, 3, 4, 4, 4, 5, 5

$$M_o = 4$$

Si en un grupo hay dos o varias puntuaciones con la misma frecuencia y esa frecuencia es la máxima, la distribución es bimodal o multimodal, es decir, tiene varias modas.

1, 1, 1, 4, 4, 5, 5, 5, 7, 8, 9, 9, 9

$$M_o = 1, 5, 9$$

Cuando todas las puntuaciones de un grupo tienen la misma frecuencia, no hay moda.

2, 2, 3, 3, 6, 6, 9, 9

Si dos puntuaciones adyacentes tienen la frecuencia máxima, la moda es el promedio de las dos puntuaciones adyacentes.

0, 1, 3, 3, 5, 5, 7, 8 $Mo = 4$

Cálculo de la moda para datos agrupados

1ª Todos los intervalos tienen la misma amplitud.

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

L_i es el límite inferior de la clase modal.

f_i es la frecuencia absoluta de la clase modal.

f_{i-1} es la frecuencia absoluta inmediatamente inferior a la clase modal.

f_{i+1} es la frecuencia absoluta inmediatamente posterior a la clase modal.

a_i es la amplitud de la clase.

Ejemplo

Calcular la moda de una distribución estadística que viene dada por la siguiente tabla:

	f_i
[60, 63)	5
[63, 66)	18
[66, 69)	42
[69, 72)	27
[72, 75)	8
	100

$$Mo = 66 + \frac{(42 - 18)}{(42 - 18) + (42 - 27)} \cdot 3 = 67.846$$

$$Mo = 66 + \frac{27}{18 + 27} \cdot 3 = 67.8$$

CUARTILES

Los cuartiles son los tres valores de la variable que dividen a un conjunto de datos ordenados en cuatro partes iguales.

Q_1 , Q_2 y Q_3 determinan los valores correspondientes al 25%, al 50% y al 75% de los datos.

Q_2 coincide con la mediana.

Cálculo de los cuartiles

1 Ordenamos los datos de menor a mayor.

2 Buscamos el lugar que ocupa cada cuartil mediante la expresión $\frac{k \cdot N}{4}$, $k = 1, 2, 3$.

Número impar de datos

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

2, 3, 4, 5, 6, 7, 9
 $\downarrow \quad \downarrow \quad \downarrow$
 $Q_1 \quad Q_2 \quad Q_3$

Número par de datos

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

1, $\frac{2+3}{2}$, $\frac{4+5}{2}$, $\frac{6+7}{2}$, 9
 $\downarrow \quad \downarrow \quad \downarrow$
 $Q_1 \quad Q_2 \quad Q_3$

Cálculo de los cuartiles para datos agrupados

En primer lugar buscamos la clase donde se encuentra $\frac{k \cdot N}{4}$, $k = 1, 2, 3$, en la tabla de las frecuencias acumuladas.

L_i es el límite inferior de la clase donde se encuentra el cuartil.

N es la suma de las frecuencias absolutas.

F_{i-1} es la frecuencia acumulada anterior a la clase del cuartil.

a_i es la amplitud de la clase.

$$Q_k = L_i + \frac{\frac{k \cdot N}{4} - F_{i-1}}{f_i} \cdot a_i$$

Ejercicio de cuartiles

Calcular los cuartiles de la distribución de la tabla:

	f_i	F_i
[50, 60)	8	8
[60, 70)	10	18
[70, 80)	16	34
[80, 90)	14	48
[90, 100)	10	58
[100, 110)	5	63
[110, 120)	2	65
	65	

Cálculo del primer cuartil

$$\frac{65 \cdot 1}{4} = 16.25$$

$$Q_1 = 60 + \frac{16.25 - 8}{10} \cdot 10 = 68.25$$

Cálculo del segundo cuartil

$$\frac{65 \cdot 2}{4} = 32.5$$

$$Q_2 = 70 + \frac{32.5 - 18}{16} \cdot 10 = 79.0625$$

Cálculo del tercer cuartil

$$\frac{65 \cdot 3}{4} = 48.75$$

$$Q_3 = 90 + \frac{48.75 - 48}{10} \cdot 10 = 90.75$$

DECILES

Los deciles son los nueve valores que dividen la serie de datos en diez partes iguales.

Los deciles dan los valores correspondientes al 10%, al 20%... y al 90% de los datos.

D_5 coincide con la mediana.

Cálculo de los deciles

En primer lugar buscamos la clase donde se encuentra $\frac{k \cdot N}{10}$, $k = 1, 2, \dots, 9$, en la tabla de las frecuencias acumuladas.

L_i es el límite inferior de la clase donde se encuentra el decil.

N es la suma de las frecuencias absolutas.

$$D_k = L_i + \frac{\frac{k \cdot N}{10} - F_{i-1}}{f_i} \cdot a_i$$

F_{i-1} es la frecuencia acumulada anterior a la clase el decil.

a_i es la amplitud de la clase.

Ejercicio de deciles

Calcular los deciles de la distribución de la tabla:

	f_i	F_i
[50, 60)	8	8
[60, 70)	10	18
[70, 80)	16	34
[80, 90)	14	48
[90, 100)	10	58
[100, 110)	5	63
[110, 120)	2	65
	65	

Cálculo del segundo decil

$$\frac{65 \cdot 2}{10} = 13$$

$$D_2 = 60 + \frac{13 - 8}{10} \cdot 10 = 65$$

Cálculo del tercer decil

$$\frac{65 \cdot 3}{10} = 19.5$$

$$D_3 = 70 + \frac{19.5 - 18}{16} \cdot 10 = 70.94$$

Cálculo del cuarto decil

$$\frac{65 \cdot 4}{10} = 26$$

$$D_4 = 70 + \frac{26 - 18}{16} \cdot 10 = 75$$

PERCENTILES

Los percentiles son los 99 valores que dividen la serie de datos en 100 partes iguales.
Los percentiles dan los valores correspondientes al 1%, al 2%... y al 99% de los datos.
P₅₀ coincide con la mediana.

Cálculo de los percentiles

En primer lugar buscamos la clase donde se encuentra $\frac{k \cdot N}{100}$, $k = 1, 2, \dots, 99$, en la tabla de las frecuencias acumuladas.

L_i es el límite inferior de la clase donde se encuentra el percentil.

N es la suma de las frecuencias absolutas.

F_{i-1} es la frecuencia acumulada anterior a la clase del percentil.

$$P_k = L_i + \frac{\frac{k \cdot N}{100} - F_{i-1}}{f_i} \cdot a_i$$

a_i es la amplitud de la clase.

Ejercicio de percentiles

Calcular el percentil 35 y 60 de la distribución de la tabla:

	f _i	F _i
[50, 60)	8	8
[60, 70)	10	18
[70, 80)	16	34
[80, 90)	14	48
[90, 100)	10	58
[100, 110)	5	63
[110, 120)	2	65
	65	

Percentil 35

$$\frac{65 \cdot 35}{100} = 22.75$$

$$P_{35} = 70 + \frac{22.75 - 18}{16} \cdot 10 = 72.97$$

Percentil 60

$$\frac{65 \cdot 60}{100} = 39$$

$$P_{60} = 80 + \frac{39 - 34}{14} \cdot 10 = 83.57$$

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 $D_{\bar{x}}$

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

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

13

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:

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

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

Ejemplo

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

	x_i	f_i	$x_i \cdot f_i$	$ x - x_i $	$ x - x_i \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	10.174	21.428
		21	457.5		98.57

$$\bar{x} = \frac{457.5}{21} = 21.786$$

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

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 .

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

Varianza para datos agrupados

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

15

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

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

Varianza para datos agrupados

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

Ejercicios de varianza

Calcular la varianza de la distribución:

9, 3, 8, 8, 9, 8, 9, 18

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

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

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

$$\bar{x} = \frac{1820}{42} = 43.33$$

$$\sigma^2 = \frac{88050}{42} - 43.33^2 = 218.94$$

Propiedades de la varianza

- La varianza será siempre un valor positivo o cero, en el caso de que las puntuaciones sean iguales.
- Si a todos los valores de la variable se les suma un número la varianza no varía.
- Si todos los valores de la variable se multiplican por un número la varianza queda multiplicada por el cuadrado de dicho número.
- Si tenemos varias distribuciones con la misma media y conocemos sus respectivas varianzas se puede calcular la varianza total.

Observaciones sobre la varianza

- La varianza, al igual que la media, es un índice muy sensible a las puntuaciones extremas.
- En los casos que no se pueda hallar la media tampoco será posible hallar la varianza.
- 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 .

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

17

Desviación típica para datos agrupados

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

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

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

Desviación típica para datos agrupados

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

Ejercicios de desviación típica

Calcular la **desviación típica** de la distribución:

9, 3, 8, 8, 9, 8, 9, 18

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

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

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

$$\bar{x} = \frac{1820}{42} = 43.33$$

$$\sigma = \sqrt{\frac{88050}{42} - 43.33^2} = 14.797$$

Propiedades de la desviación típica

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

Observaciones sobre la desviación típica

- La desviación típica, al igual que la media y la varianza, es un índice muy sensible a las puntuaciones extremas.
- En los casos que no se pueda hallar la media tampoco será posible hallar la desviación típica.
- 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.

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

El coeficiente de variación se suele expresar en porcentajes:

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

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 $\sigma = 28.28$ y otra $x = 150$ y $\sigma = 25$. ¿Cuál de las dos presenta mayor dispersión?

$$C.V_1 = \frac{28.28}{140} \cdot 100 = 20.2\%$$

$$C.V_2 = \frac{24}{150} \cdot 100 = 16\%$$

La primera distribución presenta mayor dispersión.

PUNTUACIONES TÍPICAS

Puntuaciones diferenciales

Las puntuaciones diferenciales resultan de restarles a las puntuaciones directas la media aritmética.

$$x_i = X_i - \bar{X}$$

Puntuaciones típicas

Las puntuaciones típicas son el resultado de dividir las puntuaciones diferenciales entre la desviación típica. Este proceso se llama tipificación.

Las puntuaciones típicas se representan por z .

$$z = \frac{X_i - \bar{X}}{\sigma}$$

Observaciones sobre puntuaciones típicas

- La media aritmética de las puntuaciones típicas es 0.
- La desviación típica de las puntuaciones típicas es 1.
- Las puntuaciones típicas son adimensionales, es decir, son independientes de las unidades utilizadas.

- Las puntuaciones típicas se utilizan para comparar las puntuaciones obtenidas en distintas distribuciones.

Ejemplo

En una clase hay 15 alumnos y 20 alumnas. El peso medio de los alumnos es 58.2 kg y el de las alumnas y 54.4 kg. Las desviaciones típicas de los dos grupos son, respectivamente, 3.1 kg y 5.1 kg. El peso de José es de 70 kg y el de Ana es 65 kg. ¿Cuál de ellos puede, dentro del grupo de alumnos de su sexo, considerarse más grueso?

$$z_1 = \frac{70 - 58.2}{3.1} = 3.81$$

$$z_2 = \frac{65 - 52.4}{5.1} = 2.47$$

José es más grueso respecto de su grupo que Ana respecto al suyo.