info@tuacademiafacil.com

## 1. Variables estadísticas. Definiciones y clasificación

**Población:** Conjunto de individuos que son objeto de un estudio
**Muestra:** Subconjunto de la población
**Unidad estadística:** Cada uno de los elementos de la población: (Individuo)
**Caracteres:** Cada individuo podemos estudiarlo según uno o varios caracteres.
**Modalidades:** las respuestas posibles de estos caracteres.

**Variables estadísticas y clasificación:**

QUANTITATIVA
podemos medir
→ DISCRETAS
nº finito o numerable de valores.
→ CONTINUAS
"muchos valores"
agrupamos en intervalos.

QUALITATIVA
no se puede medir

## 2. Distribuciones de frecuencias. Representaciones gráficas

**Frecuencia absoluta:** el nº de individuos que pertenecen a dicha modalidad $(A_i)$: $n_i$ $i=1 \div n$
**Frecuencia relativa:** cociente entre $\frac{n_i \to \text{frec. absoluta.}}{n \to \text{nº total de datos.}}$ $\quad f_i = \frac{n_i}{n}$

$\sum_i n_i = n \qquad \sum_{i=1}^{n} f_i = 1$

**Ejemplo carácter cuantitativo discreto:**

$\sum n_i = 10+9+6 = 25 = n$
$\sum f_i = 0{,}4 + 0{,}36 + 0{,}24 = 1$

| | $n_i$ | $f_i$ | $\xrightarrow{\times 100}$ $P_i$ | f.abs.acum | f.rel acum |
|---|---|---|---|---|---|
| 0 | 10 | $10/25 = 0{,}4$ | 40% | 10 | 0,4 |
| 1 | 9 | $9/25 = 0{,}36$ | 36% | 19 | 0,76 |
| 2 | 6 | $6/25 = 0{,}24$ | 24% | 25 | 1 |
| | 25 | 1 | 100% | | |

**Ejemplo carácter cuantitativo continuo:**

| $x_i$ | marca clas | $n_i$ | $f_i$ | $\xrightarrow{\times 100}$ $P_i$ | fabs acum | f.relat acum |
|---|---|---|---|---|---|---|
| [0,2) | 1 | 3 | $3/25 = 0{,}12$ | 12% | 3 | 0,12 |
| [2,4) | 3 | 4 | $4/25 = 0{,}16$ | 16% | 7 | 0,28 |
| [4,6) | 5 | 10 | $10/25 = 0{,}4$ | 40% | 17 | 0,68 |
| [6,8) | 7 | 7 | $7/25 = 0{,}28$ | 28% | 24 | 0,96 |
| [8,10] | 9 | 1 | $1/25 = 0{,}04$ | 4% | 25 | 1 |
| | | 25 | 1 | 100% | | |

**Representaciones gráficas:**

**Gráfico circular o de sectores:** ⇒ frec relativa × 360° // frelat × 180° (semicírculo)

| | $n_i$ | $f_i$ | $\xrightarrow{\times 100}$ $P_i$ | f.abs.acum | f.rel acum | |
|---|---|---|---|---|---|---|
| 0 | 10 | $10/25 = 0{,}4$ | 40% | 10 | 0,4 | 144° |
| 1 | 9 | $9/25 = 0{,}36$ | 36% | 19 | 0,76 | 129,6° |
| 2 | 6 | $6/25 = 0{,}24$ | 24% | 25 | 1 | 86,4° |
| | 25 | 1 $\xrightarrow{\times 360}$ | 100% | | | 360° |

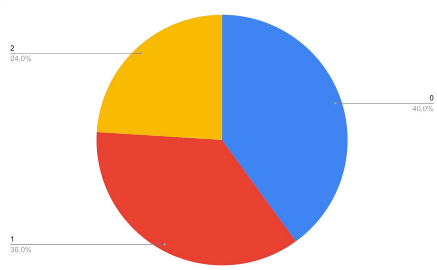
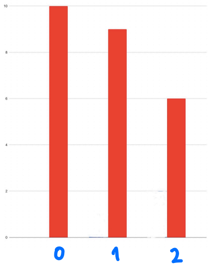
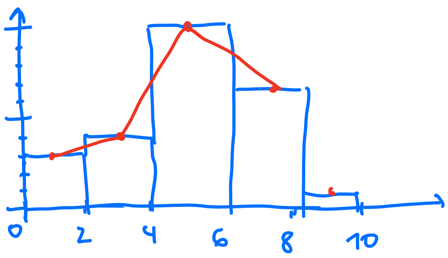


**Diagrama de barras o gráfico de Pareto:**




**Histograma y polígono de frecuencias:**




## 3. Parámetros estadísticos:

**Parámetros estadísticos de centralización:**

**Moda:** el valor de la variable que presenta la mayor frecuencia absoluta.

**Mediana:** datos ordenados de menor a mayor. es el valor que divide la muestra en dos partes con el mismo número de elementos.

n impar: Me; $\frac{n}{2}+0{,}5$ posición de la mediana

n par: $Me = \frac{\bullet + \bullet}{2}$ media aritm de los dos valores; $\frac{n}{2}, \frac{n}{2}+1$ las posiciones de los dos valores

**Media o valor medio:**

$$\bar{x} = \frac{\sum_i x_i}{n} = \frac{\sum_i x_i n_i}{n}$$

$n_i$: f. abs

$$\bar{x} = \sum_i f_i \cdot x_i$$

$f_i$: f. relat.

Propiedades:

$1, 3, 6, 8 \xrightarrow{+2} 3, 5, 8, 10$

$\bar{x} = \frac{1+3+6+8}{4} = 4{,}5 \qquad \bar{x} = \frac{3+5+8+10}{4} = 6{,}5$

1) x y x' relacionadas con una transf. lineal ⇒ sus medias están relacionadas con la misma transf. lin.

2) la media de la diferencia de una variable a su media es cero.

**Momento no centrado de orden r:**

$$\alpha_r = \frac{\sum_i n_i \cdot x_i^r}{n}$$

**Parámetros estadísticos de dispersión:**

**Desviación media:**

$$D.M(x) = \frac{\sum_{i=1}^{k} n_i \cdot |x_i - \bar{x}|}{n}$$

$V(ax) = a^2 V(x)$

$V(x+b) = V(x)$

**Varianza:**

$$V(x) = \frac{\sum_i n_i (x_i - \bar{x})^2}{n}$$

$$x' = \frac{x-b}{a} \qquad V(x') = \frac{1}{a^2} V(x)$$

1 4 7

$$\bar{x} = \frac{1+4+7}{3} = 4 \qquad V(x) = \frac{(1-4)^2 + (4-4)^2 + (7-4)^2}{3} = 6$$

$$x' = \frac{x-1}{2}$$

0 , 1'5 , 3

$$\bar{x} = \frac{1'5+3}{3} = 1'5$$

$$V(x') = \frac{(0-1{,}5)^2 + (1{,}5-1{,}5)^2 + (3-1{,}5)^2}{3} = 1{,}5$$

$$V(x') = \frac{1}{2^2} V(x) = \frac{1}{4} \cdot 6 = 1{,}5$$

**Desviación típica:**

$$S_x = \sqrt{\frac{\sum_i n_i (x_i - \bar{x})^2}{n}}$$

**Momento centrado de orden r:**

$$\mu_r = \frac{\sum n_i (x_i - \bar{x})^r}{n}$$

$\mu_r \quad r = 2 \Rightarrow \mu_2 = \text{Varianza}$.

Cálculo de la varianza:

$$V(x) = \alpha_2 - \alpha_1^2 = \alpha_2 - \bar{x}^2$$

$$V(x) = \frac{1}{n} \sum_i n_i (x_i - \bar{x})^2 = \frac{1}{n} \sum_i n_i (x_i^2 - 2x_i \bar{x} + \bar{x}^2) = \frac{1}{n} \left( \sum_i n_i x_i^2 - 2\bar{x} \sum_i n_i x_i + \bar{x}^2 \underbrace{\sum_i n_i}_{n} \right)$$

$$= \frac{1}{n} \sum_i n_i \cdot x_i^2 - 2\bar{x}^2 + \bar{x}^2 = \frac{1}{n} \sum n_i \cdot x_i^2 - \bar{x}^2 = \alpha_2 - \alpha_1^2$$

**Coeficiente de variación:**

$$C.V = v = \frac{S_x}{|\bar{x}|}$$

Si C.V es pequeño $\Rightarrow$ mejor representará la media al conjunto de datos.

**Características de forma:**

Si distribución simétrica $\Rightarrow$ $\mu_r$ de orden impar son 0.

$$\mu_r = \frac{\sum n_i \cdot (x_i - \bar{x})^r}{n}$$

Fisher $\rightarrow$ coeficiente asimetría: $\gamma_1 = \dfrac{\mu_3}{S^3}$

Si distrib simétrica $\Rightarrow \mu_3 = 0 \Rightarrow \gamma_1 = 0$

$\rightarrow$ Coeficiente de curtosis:

$$\gamma_2 = \frac{\mu_4}{S^4} - 3$$

info@tuacademiafacil.com

## Ejemplos:

**Ejercicio 1.1.** Dada la siguiente distribución de frecuencias de variable discreta, calcúlese: a) mediana, b) moda, c) media, d) varianza y desviación típica y e) porcentaje de datos que pertenecen al intervalo $(\bar{x}-s_x, \bar{x}+s_x)$.

| $x_i$ | 47 | 48 | 49 | 50 | 51 | 52 | 53 | |
|---|---|---|---|---|---|---|---|---|
| $n_i$ | 1 | 3 | 2 | 8 | 3 | 2 | 1 | $n=20$ |

| $X_i$ | $n_i$ | $n_i \cdot x_i$ | $x_i - \bar{x}$ | $(x_i-\bar{x})^2$ | $n_i(x_i-\bar{x})^2$ | $n_i$ acum. |
|---|---|---|---|---|---|---|
| 47 | 1 | 47 | −3 | 9 | 9 | |
| 48 | 3 | 252 | −2 | 4 | 12 | |
| 49 | 2 | 98 | −1 | 1 | 2 | |
| 50 | 8 | 400 | 0 | 0 | 0 | |
| 51 | 3 | 153 | 1 | 1 | 3 | |
| 52 | 2 | 104 | 2 | 4 | 8 | |
| 53 | 1 | 53 | 3 | 9 | 9 | |
| | 20 | 999 | | | 43 | |

c) $\bar{x} = \frac{\sum_i n_i \cdot x_i}{n} = \frac{999}{20} = 49{,}95 \simeq 50$

b) $M_o = 50$

a) $M_e \rightarrow \frac{n}{2}, \frac{n}{2}+1$ $n$ par

posiciones $\underset{50}{10}, \underset{50}{11}$ $\qquad M_e = \frac{50+50}{2} = 50$

d) $s^2 = \frac{\sum_i n_i(x_i-\bar{x})^2}{n} = \frac{43}{20} = 2{,}15$

$s = \sqrt{2{,}15} = 1{,}466$

$\left.C.V = \frac{s}{|\bar{x}|} \simeq \frac{1{,}466}{50} = 0{,}02932\right]$ no preguntan

e) $\bar{x} - s = 49{,}95 - 1{,}466 = 48{,}484$

$\bar{x} + s = 49{,}95 + 1{,}466 = 50{,}966$

los datos que estan $(48{,}48 ; 50{,}966)$

Tenemos $\frac{13}{20} \times 100 = 65\%$

**Ejercicio 1.2.** Sobre la siguiente distribución de frecuencias de variable continua, efectuando un cambio de origen y de escala apropiados, calcúlese: a) media, b) varianza y desviación típica, c) histograma de frecuencias y d) porcentaje de datos en el intervalo $(\bar{x}-s_x, \bar{x}+s_x)$. Interpretación estadística de este resultado.

| Clases | 5-25 | 25-35 | 35-45 | 45-55 | 55-65 | 65-85 | |
|---|---|---|---|---|---|---|---|
| $n_i$ | 2 | 4 | 10 | 10 | 3 | 1 | $n=30$ |

| clases | $n_i$ | M.C | $x'_i = \frac{x_i-40}{5}$ | $n \cdot x'_i$ | $n \cdot x'^2_i$ |
|---|---|---|---|---|---|
| [5,25) | 2 | 15 | −5 | −10 | 50 |
| [25-35) | 4 | 30 | −2 | −8 | 16 |
| [35-45) | 10 | 40 | 0 | 0 | 0 |
| [45-55) | 10 | 50 | 2 | 20 | 40 |
| [55-65) | 3 | 60 | 4 | 12 | 48 |
| [65-85] | 1 | 75 | 7 | 7 | 49 |
| | 30 | | | 21 | 203 |

a) $\bar{x} = \frac{\sum_i n_i \cdot x_i}{n} = \frac{2.15 + 4.30 + 10.40 + 10.50 + 3.60 + 1.75}{30} = 43{,}5$

$\bar{x}' = \frac{21}{30} = 0{,}7$ $\qquad \bar{x}' = \frac{\bar{x} - 40}{5}$

$5\bar{x}' = \bar{x} - 40 \Rightarrow \bar{x} = 5\bar{x}' + 40 = 43{,}5$

$x' = \frac{x-40}{5}$

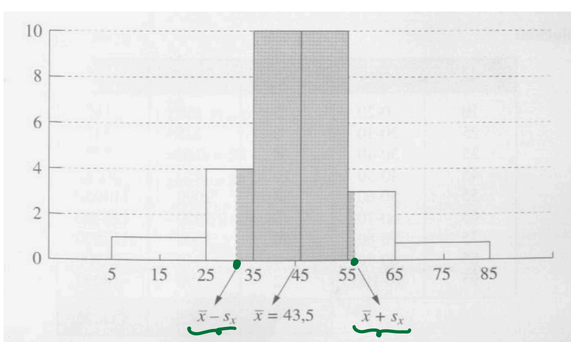
$5x' + 40 = x$

b) $V(x') = \alpha'_2 - \bar{x}'^2 = \frac{203}{30} - (0{,}7)^2 = 6{,}27$

$V(x) = V(5x'+40) = V(5x') = 25\,V(x') = 156{,}75$

$S(x) = \sqrt{156{,}75} \simeq 12{,}52$

c)




d) $\bar{x} - s = 43{,}5 - 12{,}52 = 30{,}98$

$\bar{x} + s = 43{,}5 + 12{,}52 = 56{,}02$

Clase $\left.\begin{array}{l}[45-55] \Rightarrow 10 \\ [35-45] \Rightarrow 10\end{array}\right\}$ enteras

$[25-35] \Rightarrow 4$

$\frac{4}{10} \cdot 4 = 1{,}6$

10 amp $\longrightarrow$ 4

$(35-30{,}98) \longrightarrow x$

$x = \frac{4 \cdot 4}{10}$

$[55-65] \Rightarrow 3$

$\frac{3}{10} \cdot 1 = 0{,}3$

10 ampl $\longrightarrow$ 3

$(56{,}02-55) \longrightarrow x$ $\qquad x = \frac{3 \cdot 1}{10}$

Si sumamos $\Rightarrow$ Total 21,9 sobre 30 $\Rightarrow \frac{21{,}9}{30} \times 100 \simeq \boxed{73\%}$

## 4. Distribución de frecuencias multivariante:

**Definición 10.1.** Se llama *distribución conjunta de frecuencias* de dos variables $X$ e $Y$ a una tabla en donde se representan los valores observados de cada variable y las frecuencias absolutas de cada par.

| X \ Y | 100° | 110° | 120° | |
|---|---|---|---|---|
| 1 | 10 | 6 | 4 | 20 |
| 2 | 8 | 30 | 12 | 50 |
| 3 | 2 | 9 | 19 | 30 |
| | 20 | 45 | 35 | 100 |

$n_{11}$ $n_{12}$ $n_{13}$ $n_{21}$ $n_{22}$ $n_{23}$ $n_{31}$ $n_{32}$ $n_{33}$

$$\sum_i \sum_j n_{ij} = n$$

$$\sum_i n_{ij} = n_{\cdot j}$$

## 5. Distribuciones marginales y condicionadas:

**Distribución marginal:** la que se obtiene al estudiar esa variable con independencia de las demás

$$n_{i\cdot} = \sum_j n_{ij} \qquad n_{\cdot j} = \sum_i n_{ij}$$

| X | 1 | 2 | 3 | |
|---|---|---|---|---|
| | 20 | 50 | 30 | 100 |
| | $n_{1\cdot}$ | $n_{2\cdot}$ | $n_{3\cdot}$ | |

| Y | 100° | 110° | 120° | |
|---|---|---|---|---|
| | 20 | 45 | 35 | 100 |
| | $n_{\cdot 1}$ | $n_{\cdot 2}$ | $n_{\cdot 3}$ | |

**Distribución condicionada:**

Distribución condicionada de Y por X=x a la distribución de frecuencias que se obtiene considerando únicamente las frecuencias $X = x_r$

Y condicionada X=2

| Y | 100° | 110° | 120° | |
|---|---|---|---|---|
| | 8 | 30 | 12 | 50 |

| Y | 100° | 110° | 120° |
|---|---|---|---|
| | 8/50 | 30/50 | 12/50 |

$$f(y_i/x_i) = \frac{n_{ij}}{n_{i\cdot}}$$

$$f(x_i/y_j) = \frac{n_{ij}}{n_{\cdot j}}$$

$$f(y_i/x_i) = \frac{f(x_i,y_i)}{f(x_i)} \qquad f(x_i/y_i) = \frac{f(x_i,y_i)}{f(y_i)}$$

| X \ Y | 100° | 110° | 120° | |
|---|---|---|---|---|
| 1 | 10 | 6 | 4 | 20 |
| 2 | 8 | 30 | 12 | 50 |
| 3 | 2 | 9 | 19 | 30 |
| | 20 | 45 | 35 | 100 |

$f(x_i/y=110°)$

| | |
|---|---|
| 1 | 6/45 |
| 2 | 30/45 |
| 3 | 9/45 |

## 6. Vector de medias. Matriz de varianza y covarianzas:

**Vector de medias:** de la variable k-dimensional $X$, $\bar{X} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_k \end{pmatrix}$

$$\bar{X} = \frac{1\cdot 20 + 2\cdot 50 + 3\cdot 30}{100} = 2{,}1$$

$$\bar{Y} = \frac{20\cdot 100° + 45\cdot 110° + 120°\cdot 35}{100} = 111{,}5°$$

$$\bar{X} = \begin{pmatrix} 2{,}1 \\ 111{,}5 \end{pmatrix}$$

**Matriz de varianza y covarianzas:**

Covarianza:

$$Cov(X,Y) = \frac{\sum (x_i-\bar{x})(y_j-\bar{y})}{n}$$

$$Cov(X,Y) = \frac{\sum_i n_{ij}(x_i-\bar{x})(y_j-\bar{y})}{n} = \sum f_{ij}(x_i-\bar{x})(y_j-\bar{y})$$

Matriz de Covarianzas:

⇓ matriz cuadrada y simétrica

Varianza: $S_x^2$ ; $S_y^2$

$$C = \begin{pmatrix} S_x^2 & Cov(X,Y) \\ Cov(X,Y) & S_y^2 \end{pmatrix}$$

obs: $Cov(X,Y) = Cov(Y,X)$

Variable 2-dimensional

$$C = \begin{pmatrix} S_1^2 & S_{12} & S_{13} & \cdots & S_{1k} \\ S_{21} & S_2^2 & S_{23} & & \vdots \\ S_{31} & S_{32} & S_3^2 & & \\ \vdots & & & \ddots & \\ S_{k1} & \cdots & & & S_k^2 \end{pmatrix}$$

Variable k-dimensional

La matriz de covarianzas es siempre SEMIDEFINIDA POSITIVA

**Coeficiente de correlación :**

$$r = \frac{Cov(X,Y)}{\underbrace{S_x}_{\text{desviación típica } x} \cdot \underbrace{S_y}_{\text{desviación típica } y}}$$

→ es invariante por cambios de origen y de escala.

$r \in [-1,1]$

$$x' = aX + b \qquad y' = cY + d$$

$$r' = \frac{Cov(x',y')}{S_{x'} \cdot S_{y'}} = \frac{\sum (ax + \cancel{b} - a\bar{x} - \cancel{b})(cy' + \cancel{d} - c\bar{y}' - \cancel{d})}{S_{x'} S_{y'}\, n} = \frac{1}{n\, a\, c\, S_x . S_y}\, a \cdot c \sum (x-\bar{x})(y-\bar{y})$$

$$S_{x'} = a S_x \qquad S_{y'} = c S_x$$

$$= r = \frac{Cov(X,Y)}{S_x . S_y}$$

**Ejercicio 1.4.** A partir de los datos de la siguiente tabla, referentes a horas trabajadas en un taller y unidades producidas, calcúlese el vector de medias y el coeficiente de correlación.

| | | | | | | |
|---|---|---|---|---|---|---|
| $x$ horas trabajadas | 88 | 80 | 83 | 84 | 78 | 62 |
| $y$ unidades producidas | 300 | 302 | 316 | 330 | 300 | 250 |
| $x$ horas trabajadas | 82 | 85 | 80 | 84 | 80 | 62 |
| $y$ unidades producidas | 300 | 340 | 315 | 330 | 310 | 243 |

| $x$ | $y$ | $x-\bar{x}$ | $y-\bar{y}$ | $(x-\bar{x})^2$ | $(y-\bar{y})^2$ | $(x-\bar{x})(y-\bar{y})$ |
|---|---|---|---|---|---|---|
| 88 | 300 | 9 | −3 | 81 | 9 | −27 |
| 80 | 302 | 1 | −1 | 1 | 1 | −1 |
| 83 | 316 | 4 | 13 | 16 | 169 | 52 |
| 84 | 330 | 5 | 27 | 25 | 729 | 135 |
| 78 | 300 | −1 | −3 | 1 | 9 | 3 |
| 62 | 250 | −17 | −53 | 289 | 2809 | 901 |
| 82 | 300 | 3 | −3 | 9 | 9 | −9 |
| 85 | 340 | 6 | 37 | 36 | 1369 | 222 |
| 80 | 315 | 1 | 12 | 1 | 144 | 12 |
| 84 | 330 | 5 | 27 | 25 | 729 | 135 |
| 80 | 310 | 1 | 7 | 1 | 49 | 7 |
| 62 | 243 | −17 | −60 | 289 | 3600 | 1020 |
| 948 | 3636 | 0 | 0 | 774 | 9626 | 2450 |

$$\bar{x} = \frac{\sum_i x_i}{n} = \frac{88+80+\cdots+62}{12} = \frac{948}{12} = 79$$

$$\bar{y} = \frac{\sum_i y_i}{n} = \frac{300+302+\cdots+243}{12} = \frac{3636}{12} = 303$$

$$S_x^2 = \frac{\sum_i (x_i-\bar{x})^2}{n} = \frac{774}{12} = 64{,}5 \Rightarrow S_x = \sqrt{64{,}05} = 8{,}03$$

$$S_y^2 = \frac{\sum_i (y_i-\bar{y})^2}{n} = \frac{9626}{12} = 802{,}166 \Rightarrow S_y = \sqrt{802{,}166} = 28{,}3224$$

$$Cov(X,Y) = \frac{\sum_i (x_i-\bar{x})(y_i-\bar{y})}{n} = \frac{2450}{12} = 204{,}166$$

$$r = \frac{Cov(X,Y)}{S_x . S_y} = \frac{204{,}166}{8{,}03 \cdot 28{,}3224} = 0{,}8977$$

matriz covarianzas

$$C = \begin{pmatrix} S_x^2 & Cov(X,Y) \\ Cov(Y,X) & S_y^2 \end{pmatrix} = \begin{pmatrix} 64{,}5 & 204{,}166 \\ 204{,}166 & 802{,}166 \end{pmatrix}$$