

PARTE 1: ESTADÍSTICA UNIDIMENSIONAL Y BIDIMENSIONAL

1. La siguiente tabla recoge las edades de las personas que han subido a un avión.

Edad	[0, 18)	[18, 36)	[36, 54)	[54, 72)	[72, 90)
Nº personas	15	30	20	12	3

- a) Construye la tabla de frecuencias (incluyendo frecuencias relativas y acumuladas).
b) Representa los datos en un histograma.
c) Calcula las medidas de centralización: media, moda y mediana.
d) Calcula las medidas de dispersión: desviación típica, percentil 85 y coeficiente de variación.
2. Se ha realizado una encuesta a 6 propietarios, preguntando el número de personas que habitan el hogar familiar y el número de habitaciones que tiene la casa, obteniendo la tabla dada:

Nº de personas	3	5	4	6	5	4
Nº de habitaciones	2	4	3	5	5	4

- a) Calcula el coeficiente de correlación, indicando todos los cálculos necesarios e interpretando el resultado.
b) Estima el número esperado de personas que viven en el hogar, si éste tiene 6 habitaciones. ¿Es buena la estimación?
3. De una variable bidimensional (X, Y) sabemos que:
- La recta de regresión de X sobre Y es $x = 0,6 - 0,44 \cdot y$
 - $\bar{y} = 1$
- Calcula la media de la variable X.

PARTE 2: PROBABILIDAD

1. Sabiendo que $p(A) = 0,3$; $p(B) = 0,6$; $p(\bar{A} \cap \bar{B}) = 0,8$. Calcula:
 - a) $p(A \cup B)$
 - b) $p(A \cap \bar{B})$

2. Según un estudio, el 75% de los trabajadores de una empresa llegan al trabajo de forma puntual y, de ellos, el 25% va en transporte público. Además, entre los trabajadores que llegan tarde, la mitad va en transporte público. Seleccionado un trabajador al azar, calcular la probabilidad de que:
 - a) Vaya al trabajo en transporte público.
 - b) Llegue tarde al trabajo y vaya en transporte público.
 - c) Si ha ido en transporte público al trabajo, haya sido puntual.

3. El 45% de la población de una ciudad son mujeres:
 - a) Elegidas 5 personas al azar, calcula:
 - i. La probabilidad de que 4 sean mujeres.
 - ii. La probabilidad de que como al menos haya cuatro hombres.
 - b) Si la ciudad tiene 1500 habitantes, ¿cuál es la probabilidad de que haya entre 620 y 700 mujeres?

4. El tiempo de espera de un viajero en una estación de tren, en minutos, sigue una distribución normal de media 8 y desviación típica 2. Sabiendo que cada mañana 450 viajeros acceden a la estación, calcula cuántos de ellos esperan más de 7 minutos y medio.

PARTE 1: ESTADÍSTICA UNIDIMENSIONAL Y BIDIMENSIONAL

1. La siguiente tabla recoge las edades de las personas que han subido a un avión.

Edad	[0, 18)	[18, 36)	[36, 54)	[54, 72)	[72, 90)
Nº personas	15	30	20	12	3

- a) Construye la tabla de frecuencias (incluyendo frecuencias relativas y acumuladas)

Al estar los datos agrupados en intervalos, calculamos la marca de clase x_i como el punto medio de cada uno de ellos.

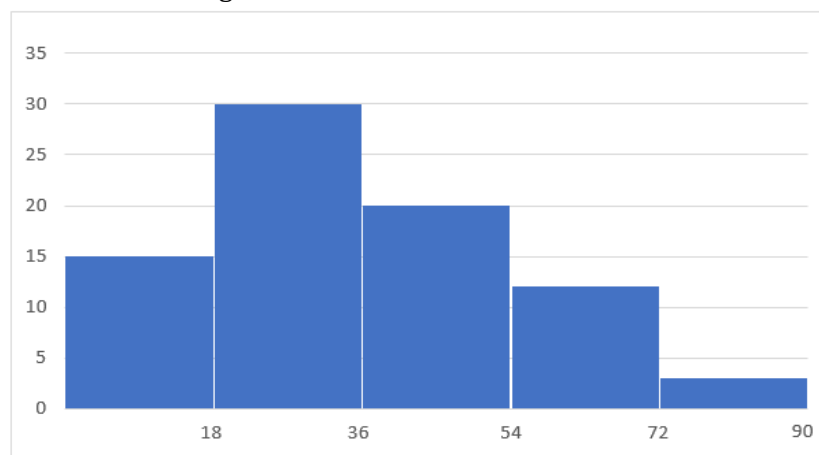
En la tabla de frecuencias indicaremos la marca de clase, las frecuencias absolutas y relativas y las frecuencias acumuladas.

intervalo	x_i	f_i	F_i	h_i	H_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
[0, 18)	9	15	15	0,1875	0,1875	135	1215
[18, 36)	27	30	45	0,375	0,5625	810	21870
[36, 54)	45	20	65	0,25	0,8125	900	40500
[54, 72)	63	12	77	0,15	0,9625	756	47628
[72, 90)	81	3	80	0,0375	1	243	19683
		80		1		2844	130896

Podemos observar que hay un total de 80 datos.

Las dos últimas columnas nos sirven para calcular las medidas de centralización y de dispersión de los siguientes apartados.

- b) Representa los datos en un histograma.



- c) Calcula las medidas de centralización: media, moda y mediana.

La media de los datos se calcula de la siguiente forma: $\bar{x} = \frac{\sum_{i=1}^{80} x_i \cdot f_i}{80} = \frac{2844}{80} = 35,55$

La moda es el dato que más se repite, es decir, aquel con mayor frecuencia absoluta. En este caso, observando la tabla de frecuencias podemos concluir que el intervalo modal sería [18, 36) y la moda la marca de clase correspondiente $M_o = 27$

La mediana es el dato central, es decir, el que deja por encima y por debajo el 50% de los datos. En este caso, al existir 80 datos no hay un dato central, sino que la mediana es la media de los datos con frecuencia absoluta 40 y 41. Ambos corresponden a la marca de clase 27, por tanto, el intervalo mediano es [18, 36) y la mediana la marca de clase $Me = 27$.

- d) Calcula las medidas de dispersión: desviación típica, percentil 85 y coeficiente de variación.

La desviación típica es la raíz cuadrada de la varianza, por tanto, calculamos en primer lugar esta medida de dispersión.

La varianza se calcula de la siguiente forma: $\sigma^2 = \frac{\sum x_i^2 \cdot f_i}{80} - \bar{x}^2 = \frac{130896}{80} - 35,55^2 = 372,3975$

De forma que la desviación típica es: $\sigma = \sqrt{372,3975} = 19,2976$

El percentil 85 (P_{85}), deja por debajo el 85% de los datos, por tanto: $85\% \cdot 80 = 68$. P_{85} es el dato con frecuencia absoluta 68, es decir, $P_{85} = 63$.

El coeficiente de variación se calcula como $CV = \frac{\sigma}{\bar{x}} = \frac{19,2976}{35,55} = 0,54283$.

2. Se ha realizado una encuesta a 6 propietarios, preguntando el número de personas que habitan el hogar familiar y el número de habitaciones que tiene la casa, obteniendo la tabla dada:

Nº de personas	3	5	4	6	5	4
Nº de habitaciones	2	4	3	5	5	4

- a) Calcula el coeficiente de correlación, indicando todos los cálculos necesarios e interpretando el resultado.

Se trata de datos de una variable bidimensional, que relaciona el número de habitaciones (X) y el número de personas que habitan en el hogar (Y) de 6 inquilinos.

El coeficiente de correlación es una medida que indica el grado de dependencia lineal que existe entre ambas variables y toma siempre un valor comprendido entre -1 y 1.

La forma de calcularlo es a partir de la covarianza y de las desviaciones típicas de las variables:

$$r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$$

Comenzamos calculando la media de habitaciones:

$$\bar{x} = \frac{2 + 4 + 3 + 5 + 5 + 4}{6} = \frac{23}{6} = 3,83 \approx 4 \text{ habitaciones}$$

La media de personas en el hogar es:

$$\bar{y} = \frac{3 + 5 + 4 + 6 + 5 + 4}{6} = \frac{27}{6} = 4,5$$

Calculamos la covarianza de X y de Y:

$$\begin{aligned} \sigma_{xy} &= \frac{\sum x_i \cdot y_i}{6} - \bar{x} \cdot \bar{y} = \\ &= \frac{2 \cdot 3 + 4 \cdot 5 + 3 \cdot 4 + 5 \cdot 6 + 5 \cdot 5 + 4 \cdot 4}{6} - 3,83 \cdot 4,5 = \frac{109}{6} - 3,83 \cdot 4,5 = 0,9317 \end{aligned}$$

Calculamos las desviaciones típicas. En primer lugar, calculamos la varianza de X y la varianza de Y:

$$\sigma_x^2 = \frac{\sum x_i^2 \cdot f_i}{6} - \bar{x}^2 = \frac{95}{6} - 3,83^2 = 1,1644 ; \text{ Luego, } \sigma_x = \sqrt{1,1644} = 1,07$$

$$\sigma_y^2 = \frac{\sum y_i^2 \cdot f_i}{6} - \bar{y}^2 = \frac{127}{6} - 4,5^2 = 0,917 ; \text{ Luego, } \sigma_y = \sqrt{0,917} = 0,9576$$

Por tanto, el coeficiente de correlación es: $r = \frac{0,9317}{1,07 \cdot 0,9576} = 0,9093$ Hay una dependencia lineal fuerte y positiva entre ambas variables.

- b) Estima el número esperado de personas que viven en el hogar, si éste tiene 6 habitaciones. ¿Es buena la estimación?

La ecuación de la recta de regresión de Y sobre X es la siguiente:

$$y = \frac{\sigma_{xy}}{\sigma^2} \cdot x + \bar{y} - \frac{\sigma_{xy}}{\sigma^2} \cdot \bar{x}$$

Sustituyendo:

$$y = \frac{0,9317}{1,1644} \cdot 4,5 - \frac{0,9317}{1,1644} \cdot 3,83 = 0,8 \cdot 6 + 1,42$$

Así, para una casa con $x=6$ habitaciones, se esperan: $y = 0,8 \cdot 6 + 1,42 = 6,22 \simeq 6$ personas

Como el coeficiente de correlación es un valor cercano a 1, podemos concluir que la estimación es fiable.

3. De una variable bidimensional (X, Y) sabemos que:

- La recta de regresión de X sobre Y es $\hat{x} = 0,6 - 0,44 \cdot y$

- $\bar{y} = 1$

Calcula la media de la variable X. (1.5 puntos)

La ecuación de la recta de regresión de X sobre Y es la siguiente:

$$\hat{x} = \frac{\sigma_{xy}}{\sigma_y^2} \cdot y + \bar{x} - \frac{\sigma_{xy}}{\sigma_y^2} \cdot \bar{y}$$

Y sabemos que, la ecuación es: $\hat{x} = 0,6 - 0,44 \cdot y$

Por tanto, identificando los coeficiente:

$$\begin{aligned} \frac{\sigma_{xy}}{\sigma_y^2} &= -0,44 ; \quad \bar{x} - \frac{\sigma_{xy}}{\sigma_y^2} \cdot \bar{y} = 0,6 \rightarrow \\ \bar{x} - (-0,44) \cdot 1 &= 0,6 \rightarrow \bar{x} + 0,44 = 0,6 \rightarrow \bar{x} = 0,16 \end{aligned}$$

PARTE 2: PROBABILIDAD

1. Sabiendo que $p(A) = 0,3$; $p(B) = 0,6$; $p(\bar{A} \cap \bar{B}) = 0,8$. Calcula:

a) $p(A \cup B)$

b) $p(A \cap \bar{B})$

a) $p(\bar{A} \cap \bar{B}) = p(\overline{A \cup B}) = 1 - p(A \cup B) \rightarrow p(A \cup B) = 1 - p(\bar{A} \cap \bar{B}) = 1 - 0,8 = 0,2$

b) $p(A \cap \bar{B}) = p(A) - p(A \cap B)$

Calculamos la probabilidad de la intersección:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) \Leftrightarrow p(A \cap B) = p(A) + p(B) - p(A \cup B) = 0,7$$

Así: $p(A - B) = 0,3 - 0,7 = -0,4$ Evidentemente esto no es posible, pues una probabilidad no puede ser negativa. Esto significa que las probabilidades de los sucesos A y B no constituyen una medida de probabilidad.

2. Según un estudio, el 75% de los trabajadores de una empresa llegan al trabajo de forma puntual y, de ellos, el 25% va en transporte público. Además, entre los trabajadores que llegan tarde, la mitad va en transporte público. Seleccionado un trabajador al azar, calcular la probabilidad de que:

a) Vaya al trabajo en transporte público.

b) Llegue tarde al trabajo y vaya en transporte público.

c) Si ha ido en transporte público al trabajo, haya sido puntual.

Definimos los sucesos:

P = " el trabajador es puntual"

$\bar{P}$ = "el trabajador llega tarde"

TP = "el trabajador va en transporte público"

$\overline{TP}$ = "el trabajador no va en transporte público"

a) Utilizamos el Teorema de la probabilidad total:

$$p(TP) = p(P) \cdot p(TP/P) + p(\bar{P}) \cdot p(TP/\bar{P}) = 0,75 \cdot 0,25 + 0,25 \cdot 0,5 = 0,3125$$

b) Si llega tarde al trabajo y va en transporte público, nos referimos al suceso $\bar{P} \cap TP$:

$$p(\bar{P} \cap TP) = p(\bar{P}) \cdot p(TP/\bar{P}) = 0,25 \cdot 0,5 = 0,125$$

c) Utilizamos el Teorema de Bayes:

$$p(P/TP) = \frac{p(TP/P) \cdot p(P)}{p(TP)} = \frac{0,25 \cdot 0,75}{0,3125} = 0,6$$

3. El 45% de la población de una ciudad son mujeres:

a) Elegidas 5 personas al azar, calcula:

i. La probabilidad de que 4 sean mujeres.

ii. La probabilidad de que al menos haya cuatro hombres.

b) Si la ciudad tiene 1500 habitantes, ¿cuál es la probabilidad de que haya entre 620 y 700 mujeres?

a) Se trata de una variable aleatoria con distribución binomial de parámetros $n=5$ y $p=0,45$. Es decir:

$$X \sim B(5, 0'45) \quad X = \text{"número de mujeres"}$$

Esta variable aleatoria indica la probabilidad de obtener un determinado número de hombres de entre las 5 personas elegidas. La función de probabilidad asociada a X es:

$$p(X = k) = \binom{5}{k} \cdot 0,45^k \cdot 0,55^{5-k} \quad k = 0, 1, \dots, 5$$

i) Nos piden la probabilidad de que 4 sean hombres:

$$p(X = 4) = \binom{5}{4} \cdot 0,45^4 \cdot 0,55 = 0,11277$$

ii) Nos piden la probabilidad de que al menos haya 4 hombres, lo que equivale es a que como mucho haya una mujer, es decir:

$$p(X \leq 1) = p(X = 0) + p(X = 1)$$

$$p(X = 0) = \binom{5}{0} \cdot 0,45^0 \cdot 0,55^5 = 0,05033$$

$$p(X = 1) = \binom{5}{1} \cdot 0,45 \cdot 0,55^4 = 0,20589$$

$$\text{Por tanto: } p(X \leq 1) = 0,05033 + 0,20589 = 0,25622$$

a) Si la ciudad tiene 1500 habitantes, se trata de una variable aleatoria con distribución binomial de parámetros $n=1500$ y $p=0,45$. Como $n \cdot p = 1500 \cdot 0,45 = 675 > 5$ vamos a utilizar la aproximación de la distribución Binomial a la distribución Normal. Así:

$$X \sim B(1500, 0'45) \xrightarrow{\text{APROXIMACIÓN}} X_N \sim N(675, 19'27)$$

Nos piden calcular $p(620 \leq X \leq 700)$, utilizaremos la aproximación anterior y las correcciones de Yates:

(En lo que sigue, consideramos que $Z \sim N(0,1)$)

$$p(620 \leq X \leq 700) = p(619,5 \leq X_N \leq 700,5) = p\left(\frac{619,5 - 675}{19,27} \leq \frac{X_N - 675}{19,27} \leq \frac{700,5 - 675}{19,27}\right) =$$

$$p(-2,88 \leq Z \leq 1,32) = p(Z \leq 1,32) - p(Z \leq -2,88) = p(Z \leq 1,32) - p(Z \geq 2,88) =$$

$$= p(Z < 1,32) - (1 - p(Z < 2,88)) = p(Z < 1,32) - 1 + p(Z < 2,88) = 0,9066 - 1 + 0,9980 = 0,9046$$

4. El tiempo de espera de un viajero en una estación de tren, en minutos, sigue una distribución normal de media 8 y desviación típica 2. Sabiendo que cada mañana 450 viajeros acceden a la estación, calcula cuántos de ellos esperan más de 7 minutos y medio.

a) Se trata de una variable aleatoria que sigue una distribución normal de media 8 y desviación típica 2.

$$X \sim N(8,2)$$

$$p(X > 7,5) = p\left(Z > \frac{7,5 - 8}{2}\right) = p(Z > 0 - 0,25) = p(Z < 0,25) = 0,5987$$

Como cada mañana acceden a la estación 450 viajeros, el número de ellos que esperan más de 9 minutos y medio es: $0,5987 \cdot 450 = 269,4 \simeq 269$ viajeros aproximadamente.