

1.- Sea X una variable aleatoria que mide la estatura de los individuos de una población y que se distribuye según una normal de media 1,74 y de desviación estándar 0,1. Hallar

a) La probabilidad de que la estatura de un individuo, elegido al azar, esté comprendida entre 1,64 y 1,74.

Resolución

$$X = \text{estatura} \rightarrow N(1,74; 0,1) \Rightarrow Z = \frac{X - 1,74}{0,1} \rightarrow N(0, 1).$$

$$p(1,64 < X < 1,74) = p\left(\frac{1,64 - 1,74}{0,1} < \frac{X - 1,74}{0,1} < \frac{1,74 - 1,74}{0,1}\right) = p(-1 < Z < 0) =$$

$$= p(Z < 0) - p(Z > 1) = p(Z < 0) - 1 + p(Z < 1) = 0,5 - 1 + 0,8413 = 34,13\%.$$

b) Si sabemos que los individuos tienen una estatura superior a 1,64. ¿Cuál es la probabilidad de que un individuo elegido al azar tenga una estatura inferior a 1,74?

Resolución

Se pide una probabilidad condicionada: $p = p(X < 1,74 / X > 1,64) = \frac{p(1,64 < X < 1,74)}{p(X > 1,64)}$

Hallemos $p(X > 1,64) = p\left(\frac{X - 1,74}{0,1} > \frac{1,64 - 1,74}{0,1}\right) = p(Z > -1) = p(Z < 1) = 0,8413$

Por tanto, $p \cong \frac{0,3413}{0,8413} \cong 0,4057 = 40,57\%$

2.- Hallar la probabilidad de que en una familia de 4 hijos

a) Dos al menos sean varones.

Resolución

Como realizamos 4 veces el experimento aleatorio de comprobar el sexo del hijo y la probabilidad de que sea varón es 0,5, entonces la variable aleatoria $X = \text{"nº de varones"} \rightarrow B(4; 0,5)$

La ley de probabilidad es $p_k = p(X = k) = \binom{4}{k} 0,5^k 0,5^{4-k} = \binom{4}{k} 0,5^4$, con $k = 0, 1, 2, 3, 4$.

$$p(X \geq 2) = p(X = 2) + p(X = 3) + p(X = 4) = \binom{4}{2} 0,5^4 + \binom{4}{3} 0,5^4 + \binom{4}{4} 0,5^4 = (6 + 4 + 1) 0,5^4 = 6$$

b) Si el primer hijo es varón, que los otros tres también lo sean.

Resolución

Se pide una probabilidad condicionada:

$$p = p(2^\circ, 3^\circ \text{ y } 4^\circ \text{ varones} / 1^\circ \text{ hijo varón}) = \frac{p(X=4)}{p(X=1)} = \frac{\binom{4}{4} 0,5^4}{\binom{4}{1} 0,5^4} = \frac{1}{4} = 0,25 = 25\%$$

3.- Sea X una variable aleatoria que toma los valores 2 y 3 con igual probabilidad. Calcular su media y su varianza.

Resolución

Sea $p_i = p(X = x_i)$. La tabla que representa la función de probabilidad de X es:

x_i	$p_i = p(X = x_i)$	$x_i p_i$	$x_i^2 p_i$
2	0,5	1	2
3	0,5	1,5	4,5
total	1	2,5	6,5

La media de X es $\mu = \sum x_i p_i = 2,5$ y la varianza de X es $\sigma^2 = \sum x_i^2 p_i - \mu^2 = 6,5 - 2,5^2 = 0,25$

4.- Un jugador de baloncesto anota cada domingo el número de tantos que encesta en el partido de liga. Las anotaciones de los diez últimos encuentros, jugados por su equipo, se muestran en el siguiente cuadro:

Encuentros	1º	2º	3º	4º	5º	6º	7º	8º	9º	10º
Tantos	10	18	17	8		9	19	10	7	10

Hallar, sabiendo que en el quinto encuentro no pudo jugar:

a) La media y moda de las anotaciones ¿Cuál de los dos promedios describe de manera más realista las anotaciones?

Resolución

moda: 10 (valor de mayor frecuencia absoluta) ; media: $\bar{x} = \frac{10+18+17+8+9+19+7}{9} = \frac{108}{9} = 12$.

La media describe de forma más realista las anotaciones

b) Si se supone que las anotaciones siguen una tendencia lineal con el transcurso de los encuentros, estimar los puntos que hubiera encestado en el quinto encuentro

Hallamos la recta de regresión. El nº de datos es $N = 9$. Elaboramos la tabla de frecuencias adecuada:

	1	2	3	4	6	7	8	9	10	→ 50
y_i	10	18	17	8	9	19	10	7	10	→ 108
x_i^2	1	4	9	16	36	49	64	81	100	→ 360
y_i^2	100	324	289	64	81	361	100	49	100	→ 1468
$x_i y_i$	10	36	51	32	54	133	80	63	100	→ 559

$$\bar{x} = \frac{\sum x_i}{N} = \frac{50}{9} \approx 5,56 ; \quad \bar{y} = \frac{\sum y_i}{N} = \frac{108}{9} = 12 ; \quad s_x^2 = \frac{\sum x_i^2}{N} - \bar{x}^2 \approx \frac{360}{9} - 5,56^2 \approx 9,0864$$

$$s_y^2 = \frac{\sum y_i^2}{N} - \bar{y}^2 = \frac{1468}{9} - 12^2 \approx 19,1111 ; \quad s_{xy} = \frac{\sum x_i y_i}{N} - \bar{x} \bar{y} = \frac{559}{9} - 5,56 \cdot 12 \approx -4,6089$$

La recta de regresión (de Y sobre X) es $r_{YX}: y - \bar{y} = \frac{s_{xy}}{s_x^2} (x - \bar{x}) \Rightarrow r_{YX}: y - 12 = \frac{-4,6089}{9,0864} (x - 5,56)$

Operando y simplificando, la recta de regresión que se pide es $r_{YX}: y = -0,5x + 14,8$

Para $x = 5$, $y = -0,5 \cdot 5 + 14,8 = 12,3$. En el quinto encuentro estimamos una anotación de 12 puntos.

El coeficiente de correlación lineal es $r = \frac{s_{xy}}{s_x s_y} = \frac{-4,6089}{\sqrt{9,0864} \cdot \sqrt{19,1111}} \approx -0,3498$, negativo y próximo a 0.

Como el coeficiente de correlación es negativo la correlación es inversa, cuando aumenta una variable disminuye la otra y es débil pues está más próximo a 0 que a 1.

La estimación NO es buena

5.- La información estadística obtenida de una muestra de tamaño 12 sobre la relación existente entre la inversión realizada y el rendimiento obtenido en cientos de miles de pesetas para explotaciones agropecuarias, se muestra en el siguiente cuadro:

Inversión	11	14	16	15	16	18	20	21	14	20	19	11
Rendimiento	2	3	5	6	5	3	7	10	6	10	5	6

Hallar:

- La recta de regresión del rendimiento respecto de la inversión.
- La previsión de inversión que se obtendrá con un rendimiento de 1250000 ptas.

Resolución

El nº de datos es N = 12. Elaboramos la tabla de frecuencias adecuada:

	11	14	16	15	16	18	20	21	14	20	19	11	→ 195
y_i	2	3	5	6	5	3	7	10	6	10	5	6	→ 68
x_i^2	121	196	256	225	256	324	400	441	196	400	361	121	→ 3297
y_i^2	4	9	25	36	25	9	49	100	36	100	25	36	→ 454
$x_i y_i$	22	42	80	90	80	54	140	210	84	200	95	66	→ 1163

$$\bar{x} = \frac{\sum x_i}{N} = \frac{195}{12} = 16,25 \quad ; \quad \bar{y} = \frac{\sum y_i}{N} = \frac{68}{12} \cong 5,67 \quad ; \quad s_x^2 = \frac{\sum x_i^2}{N} - \bar{x}^2 = \frac{3297}{12} - 16,25^2 = 10,6875$$

$$s_y^2 = \frac{\sum y_i^2}{N} - \bar{y}^2 \cong \frac{454}{12} - 5,67^2 \cong 5,68 \quad ; \quad s_{xy} = \frac{\sum x_i y_i}{N} - \bar{x} \bar{y} \cong \frac{1163}{12} - 16,25 \cdot 5,67 \cong 4,7792$$

La recta de regresión (de X sobre Y) es $r_{XY}: x - \bar{x} = \frac{s_{xy}}{s_y^2} (y - \bar{y}) \Rightarrow r_{XY}: x - 16,25 = \frac{4,7792}{5,68} (y - 5,67)$

Operando y simplificando, la recta de regresión que se pide es $r_{XY}: x = 0,84y + 11,48$

Para $y = 12,5$, $x \cong 0,84 \cdot 12,5 + 11,48 \cong 21,98$. La inversión será aproximadamente de 2198000 ptas

El coeficiente de correlación lineal es $r = \frac{s_{xy}}{s_x s_y} = \frac{4,7792}{\sqrt{10,6875} \cdot \sqrt{5,68}} \cong 0,6134$, positivo y próximo a 1.

Como el coeficiente de correlación es positivo la correlación es directa, cuando aumenta una variable aumenta la otra y es fuerte pues está más próximo a 1 que a 0.

6.- La tabla siguiente contiene los valores de dos variables X e Y. Calcule el coeficiente de correlación lineal entre las variables y la recta de regresión de Y sobre X. ¿Cuál es el valor estimado por dicha recta para y cuando $x = 10$? ¿Cómo es esta predicción según el valor del coeficiente de correlación?

X	1	3	5	7	9	11	13	15
Y	1	5	8	14	17	21	25	29

Resolución

El nº de datos es N = 8. Elaboramos la tabla de frecuencias adecuada:

	1	3	5	7	9	11	13	15	→ 64
y_i	1	5	8	14	17	21	25	29	→ 120
x_i^2	1	9	25	49	81	121	169	225	→ 680
y_i^2	1	25	64	196	289	441	625	841	→ 2482
$x_i y_i$	1	15	40	98	153	231	325	435	→ 1298

$$\bar{x} = \frac{\sum x_i}{N} = \frac{64}{8} = 8 ; \quad \bar{y} = \frac{\sum y_i}{N} = \frac{120}{8} = 15 ; \quad s_x^2 = \frac{\sum x_i^2}{N} - \bar{x}^2 = \frac{680}{8} - 8^2 = 21$$

$$s_y^2 = \frac{\sum y_i^2}{N} - \bar{y}^2 = \frac{2482}{8} - 15^2 = 85,25 ; \quad s_{xy} = \frac{\sum x_i y_i}{N} - \bar{x} \bar{y} = \frac{1298}{8} - 8 \cdot 15 = 42,25$$

El coeficiente de correlación lineal es $r = \frac{s_{xy}}{s_x s_y} = \frac{42,25}{\sqrt{21} \cdot \sqrt{85,25}} \cong 0,9986$.

La recta de regresión (de Y sobre X) es $r_{YX}: y - \bar{y} = \frac{s_{xy}}{s_x^2}(x - \bar{x}) \Rightarrow r_{YX}: y - 15 = \frac{42,25}{21}(x - 8)$

Operando y simplificando, la recta de regresión que se pide es $r_{YX}: y = 2,01x - 1,1$

Para $x = 10, y \cong 2,01 \cdot 10 - 1,1 = 19$

Como el coeficiente de correlación es positivo la correlación es directa, cuando aumenta una variable aumenta la otra y es muy fuerte pues está mucho más próximo a 1 que a 0.

La predicción es muy buena

7.- Los siguientes datos corresponden a la altura en centímetros de los alumnos de una determinada clase: 150, 169, 171, 172, 172, 175, 176, 177, 178, 179, 181, 182, 183, 184, 184. Calcule la mediana, cuartiles, rango y moda de la variable y realice una representación gráfica Box-whisker. Indique el significado de los parámetros encontrados.

Resolución

Las alturas de los 15 alumnos ya están ordenadas de menor a mayor.

El primer cuartil, Q1, deja el 25% de los datos a la izquierda ; $25\% \text{ de } 15 = 3,75 \Rightarrow Q1 = 172 \text{ cm}$

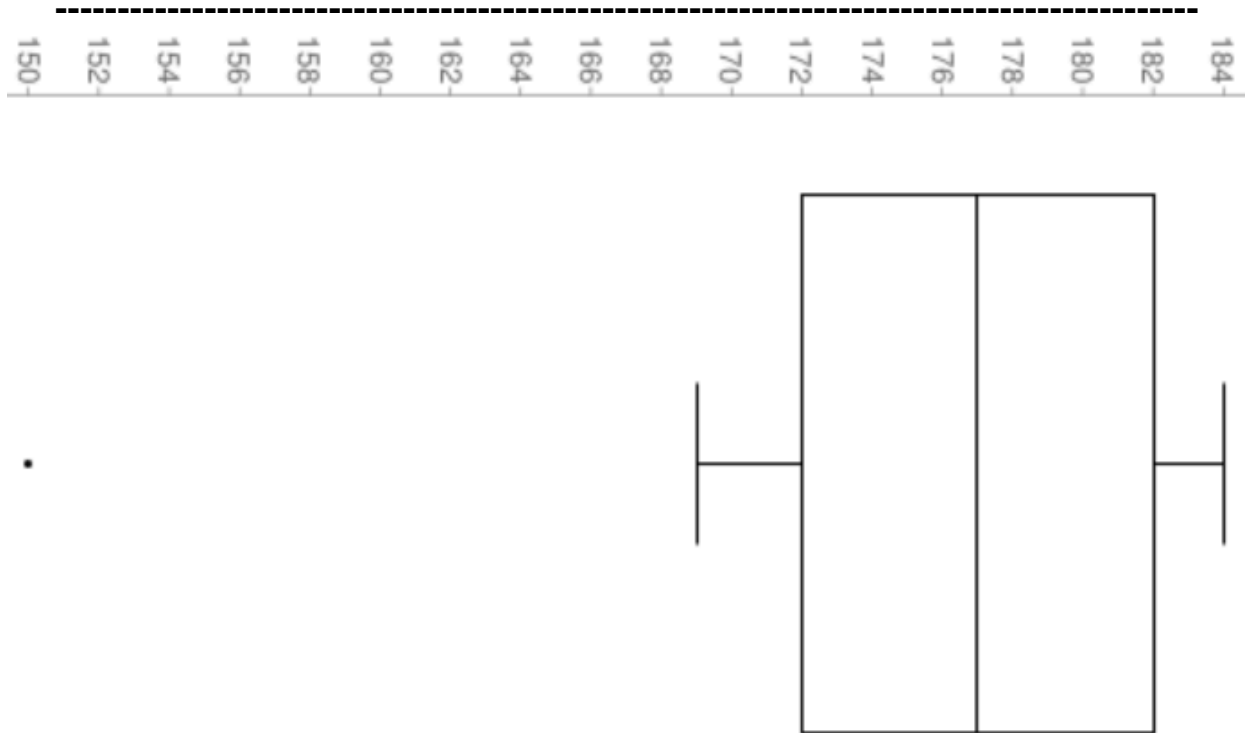
El segundo cuartil, Q2, es la mediana (el dato central), 177 cm

El tercer cuartil, Q3, deja el 75% de los datos a la izquierda ; $75\% \text{ de } 15 = 11,25 \Rightarrow Q3 = 182 \text{ cm}$

El rango es la diferencia entre la mayor y menor altura, $184 - 150 = 34 \text{ cm}$

La moda es el valor que más se repite. Hay dos modas: 172 cm y 184 cm

Diagrama Box-whisker (o diagrama de caja-bigotes)



8.- La tabla siguiente muestra el número de gérmenes patógenos (en miles por centímetro cúbico) de un determinado cultivo, según el tiempo transcurrido. Calcule una recta de regresión para predecir el número de gérmenes por centímetro cúbico en función del tiempo. ¿Qué cantidad de gérmenes por centímetro cúbico es predecible encontrar cuando hayan transcurrido seis horas? ¿Es buena esa predicción?

Nº de horas	0	1	2	3	4	5
Nº de gérmenes	20	26	33	41	47	53

Resolución

El nº de datos es $N = 6$. Elaboramos la tabla de frecuencias adecuada:

	0	1	2	3	4	5	→ 15
y_i	20	26	33	41	47	53	→ 220
x_i^2	0	1	4	9	16	25	→ 55
y_i^2	400	676	1089	1681	2209	2809	→ 8864
$x_i y_i$	0	26	66	123	188	265	→ 668

$$\bar{x} = \frac{\sum x_i}{N} = \frac{15}{6} = 2,5 \quad ; \quad \bar{y} = \frac{\sum y_i}{N} = \frac{220}{6} \cong 36,67 \quad ; \quad s_x^2 = \frac{\sum x_i^2}{N} - \bar{x}^2 = \frac{55}{6} - 2,5^2 \cong 2,92$$

$$s_y^2 = \frac{\sum y_i^2}{N} - \bar{y}^2 = \frac{8864}{6} - 36,67^2 \cong 132,64 \quad ; \quad s_{xy} = \frac{\sum x_i y_i}{N} - \bar{x} \bar{y} = \frac{668}{6} - 2,5 \cdot 36,67 \cong 19,66$$

El coeficiente de correlación lineal es $r = \frac{s_{xy}}{s_x s_y} \cong \frac{19,66}{\sqrt{2,92} \cdot \sqrt{132,64}} \cong 0,99898$.

La recta de regresión (de Y sobre X) es $r_{YX}: y - \bar{y} = \frac{s_{xy}}{s_x^2}(x - \bar{x}) \Rightarrow r_{YX}: y - 36,67 = \frac{19,66}{2,92}(x - 2,5)$

Operando y simplificando, la recta de regresión que se pide es $r_{yx}: y = 6,73x + 19,84$

Para $x = 6$, $y \cong 6,73 \cdot 6 + 19,84 \cong 60$. Cuando hayan transcurrido 6 horas habrá unos 60 gérmenes

Como el coeficiente de correlación es positivo la correlación es directa, cuando aumenta una variable aumenta la otra y es muy fuerte pues está mucho más próximo a 1 que a 0. La predicción es muy buena

9.- Las alturas, A, expresadas en centímetros, y los pesos, P, expresados en kilos, de varias personas vienen dados por la siguiente tabla:

A	164	173	187	169	175
P	68	75	90	71	77

Calcule el coeficiente de correlación y la predicción del peso de una persona de 168 centímetros de altura mediante mínimos cuadrados. Interprete cómo es esta predicción en función del valor del coeficiente de correlación.

Resolución

El nº de datos es $N = 5$. Elaboramos la tabla de frecuencias adecuada:

	164	173	187	169	175	→ 868
y_i	68	75	90	71	77	→ 381
x_i^2	26896	29929	34969	28561	30625	→ 150980
y_i^2	4624	5625	8100	5041	5929	→ 29319
$x_i y_i$	11152	12975	16830	11999	13475	→ 66431

$$\bar{x} = \frac{\sum x_i}{N} = \frac{868}{5} = 173,6 \quad ; \quad \bar{y} = \frac{\sum y_i}{N} = \frac{381}{5} = 76,2 \quad ; \quad s_x^2 = \frac{\sum x_i^2}{N} - \bar{x}^2 = \frac{150980}{5} - 173,6^2 = 59,04$$

$$s_y^2 = \frac{\sum y_i^2}{N} - \bar{y}^2 = \frac{29319}{5} - 76,2^2 = 57,36 \quad ; \quad s_{xy} = \frac{\sum x_i y_i}{N} - \bar{x} \bar{y} = \frac{66431}{5} - 173,6 \cdot 76,2 = 57,88$$

El coeficiente de correlación lineal es $r = \frac{s_{xy}}{s_x s_y} = \frac{57,88}{\sqrt{59,04} \cdot \sqrt{57,36}} \cong 0,9946$.

Por el método de ajuste por mínimos cuadrados, la recta que mejor se ajusta a los datos es $r: y = mx + b$

$$\text{siendo } m = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{N}}{\sum x_i^2 - \frac{(\sum x_i)^2}{N}}, \quad b = \bar{y} - m \bar{x}. \text{ Sustituyendo, } m = \frac{66431 - \frac{868 \cdot 381}{5}}{150980 - \frac{868^2}{5}} = \frac{289,4}{295,2} \cong 0,98035$$

$b \cong 76,2 - 0,98035 \cdot 173,6 \cong -93,98876$. La recta es $r: y = 0,98035x - 93,98876$

Para $x = 168$, $y \cong 0,98035 \cdot 168 - 93,98876 \cong 70,71$. Es decir, pesaría 70,71 kg.

Como el coeficiente de correlación es positivo la correlación es directa, cuando aumenta una variable aumenta la otra y es muy fuerte pues está mucho más próximo a 1 que a 0. La predicción es muy buena.

10.- Los pesos de cincuenta personas vienen dados por la tabla:

Intervalos	50-60	60-70	70-80	80-90	90-100
Frecuencias	10	15	20	4	1

Calcule la mediana y el intervalo intercuartílico (el que contiene al 50% central de la distribución).

¿Qué miden estos parámetros?

Resolución

clases (pesos)	[50, 60)	[60, 70)	[70, 80)	[80, 90)	[90, 100)	total
n_i (frecuencia absoluta)	10	15	20	4	1	50 = N
N_i (frecuencia absoluta acumulada)	10	25	45	49	50	-

La mediana Me deja la mitad de los datos a la izquierda. Como $\frac{50}{2} = 25$, el intervalo mediano

es $[70, 80)$. La mediana es $Me = L_i + \frac{c\left(\frac{N}{2} - N_{i-1}\right)}{n_i}$

siendo L_i = límite inferior de la clase mediana, c = amplitud del intervalo, N = número total de datos.

N_{i-1} = frecuencia absoluta acumulada de la clase anterior a la clase mediana.

n_i = frecuencia absoluta de la clase mediana. Luego, $Me = 70 + \frac{10(25 - 25)}{45} = 70$

El primer cuartil, Q_1 , deja el 25% de los datos a la izquierda. Como $25\% \text{ de } N = 25\% \text{ de } 50 = 12,5$, la clase que contiene a Q_1 es $[60, 70)$.

Aplicando la expresión del primer cuartil, similar a la de la mediana, se tiene:

El primer cuartil es $Q_1 = L_i + \frac{c\left(\frac{N}{4} - N_{i-1}\right)}{n_i} = 60 + \frac{10(12,5 - 10)}{25} = 61$

El tercer cuartil, Q_3 , deja el 75% de los datos a la izquierda. Como $75\% \text{ de } N = 75\% \text{ de } 50 = 37,5$, la clase que contiene a Q_3 es $[70, 80)$.

Aplicando la expresión del tercer cuartil, similar a la de la mediana, se tiene:

El tercer cuartil es $Q_3 = L_i + \frac{c\left(\frac{3N}{4} - N_{i-1}\right)}{n_i} = 70 + \frac{10(37,5 - 25)}{45} \cong 72,78$

El intervalo intercuartílico es $[Q_1, Q_3] = [61 ; 72,78]$