

Problemas de Estadística Descriptiva. Relación entre variables. Soluciones

1. El Hospital River Hills está interesado en determinar la efectividad de un nuevo medicamento para reducir el tiempo requerido para la recuperación completa de la cirugía de rodilla. La recuperación completa se mide mediante una serie de pruebas de fuerza que comparan la rodilla tratada con la rodilla no tratada. El fármaco se administró en cantidades variables a 18 pacientes durante un período de 6 meses. Para cada paciente, el número de unidades de fármaco, X , y los días para la recuperación completa, Y , vienen dados por los siguientes datos (x, y) :

(5, 53), (21, 65), (14, 48), (11, 66), (9, 46), (4, 56),
(7, 53), (21, 57), (17, 49), (14, 66), (9, 54), (7, 56),
(9, 53), (21, 52), (13, 49), (14, 56), (9, 59), (4, 56).

- Calcule la covarianza.
- Calcule el coeficiente de correlación.
- Analice brevemente la relación entre el número de unidades de fármaco y el tiempo de recuperación.

Solución

- a. En primer lugar, construimos la tabla siguiente donde la última fila es la suma de las filas anteriores:

X	Y	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$Y_i - \bar{Y}$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
5	53	-6.6111111	43.7067901	-2.2222222	4.9382716	14.691358
21	65	9.3888889	88.1512346	9.7777778	95.6049383	91.802469
14	48	2.3888889	5.7067901	-7.2222222	52.1604938	-17.253086
11	66	-0.6111111	0.3734568	10.7777778	116.1604938	-6.586420
9	46	-2.6111111	6.8179012	-9.2222222	85.0493827	24.080247
4	56	-7.6111111	57.9290123	0.7777778	0.6049383	-5.919753
7	53	-4.6111111	21.2623457	-2.2222222	4.9382716	10.246914
21	57	9.3888889	88.1512346	1.7777778	3.1604938	16.691358
17	49	5.3888889	29.0401235	-6.2222222	38.7160494	-33.530864
14	66	2.3888889	5.7067901	10.7777778	116.1604938	25.746914
9	54	-2.6111111	6.8179012	-1.2222222	1.4938272	3.191358
7	56	-4.6111111	21.2623457	0.7777778	0.6049383	-3.586420
9	53	-2.6111111	6.8179012	-2.2222222	4.9382716	5.802469
21	52	9.3888889	88.1512346	-3.2222222	10.3827160	-30.253086
13	49	1.3888889	1.9290123	-6.2222222	38.7160494	-8.641975
14	56	2.3888889	5.7067901	0.7777778	0.6049383	1.858025
9	59	-2.6111111	6.8179012	3.7777778	14.2716049	-9.864197
4	56	-7.6111111	57.9290123	0.7777778	0.6049383	-5.919753
209	994	0.0000000	542.2777778	0.0000000	589.1111111	72.555556

Las medias de las variables X e Y son:

$$\bar{X} = \frac{209}{18} = 11.611, \quad \bar{Y} = \frac{994}{18} = 55.222.$$

La covarianza entre las variables X e Y será:

$$\text{cov}(X, Y) = \frac{72.5556}{17} = 4.268.$$

- b. Para calcular el coeficiente de correlación r entre X e Y , necesitamos calcular primeramente las varianzas correspondientes:

$$s_X^2 = \frac{542.2778}{17} = 31.8987, \quad s_Y^2 = \frac{589.1111}{17} = 34.6536.$$

El coeficiente de correlación será:

$$r = \frac{\text{cov}(X, Y)}{s_X \cdot s_Y} = \frac{4.268}{\sqrt{31.8987} \cdot \sqrt{34.6536}} = 0.1284.$$

- c. El coeficiente de correlación es bajo. Por tanto, la relación entre los días de recuperación y las unidades de fármaco no sería lineal.

2. Una empresa de bienes de consumo ha estado estudiando el efecto de la publicidad en los beneficios totales. Como parte de este estudio, se recopilaron datos sobre gastos publicitarios (en miles de dólares) y ventas totales (en miles de dólares) durante un período de 5 meses y son los siguientes:

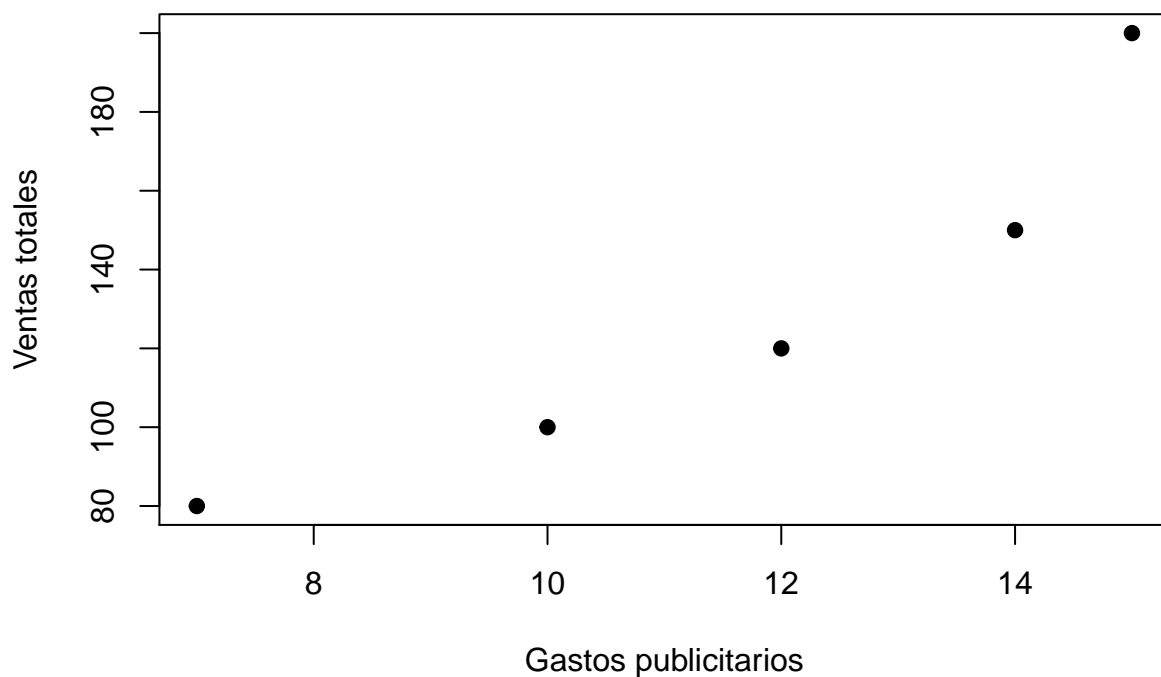
$$(10, 100), (15, 200), (7, 80), (12, 120), (14, 150).$$

El primer número son los gastos publicitarios y el segundo son las ventas totales. Haga un gráfico de los datos y calcule el coeficiente de correlación.

Solución

X	Y	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$Y_i - \bar{Y}$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
10	100	-1.6	2.56	-30	900	48
15	200	3.4	11.56	70	4900	238
7	80	-4.6	21.16	-50	2500	230
12	120	0.4	0.16	-10	100	-4
14	150	2.4	5.76	20	400	48
58	650	0.0	41.20	0	8800	560

El gráfico de los datos es el siguiente:



Para hallar el coeficiente de correlación, hacemos la tabla siguiente:

X	Y	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$Y_i - \bar{Y}$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
10	100	-1.6	2.56	-30	900	48
15	200	3.4	11.56	70	4900	238
7	80	-4.6	21.16	-50	2500	230
12	120	0.4	0.16	-10	100	-4
14	150	2.4	5.76	20	400	48
58	650	0.0	41.20	0	8800	560

Calculemos primero las medias y la covarianza entre los gastos publicitarios (X) y las ventas totales (Y):

$$\bar{X} = \frac{58}{5} = 11.6, \quad \bar{Y} = \frac{650}{5} = 130, \quad \text{cov}(X, Y) = \frac{560}{4} = 140.$$

Para calcular el coeficiente de correlación r entre X e Y , necesitamos calcular primeramente las varianzas correspondientes:

$$s_X^2 = \frac{41.2}{4} = 10.3, \quad s_Y^2 = \frac{8800}{4} = 2200.$$

El coeficiente de correlación será:

$$r = \frac{\text{cov}(X, Y)}{s_X \cdot s_Y} = \frac{140}{\sqrt{10.3} \cdot \sqrt{2200}} = 0.93.$$

Vemos que existe una relación lineal acentuada entre los gastos publicitarios y las ventas totales, tal como puede apreciarse en el gráfico anterior.

3. El presidente de Floor Coverings Unlimited desea información sobre la relación entre la experiencia minorista (años) y las ventas semanales (en cientos de euros). Obtuvo la siguiente muestra aleatoria sobre experiencia y ventas semanales:

$(2, 5), (4, 10), (3, 8), (6, 18), (3, 6), (5, 15), (6, 20), (2, 4).$

El primer número para cada observación son los años de experiencia y el segundo número son las ventas semanales. Calcule la covarianza y el coeficiente de correlación.

Solución

En primer lugar, construimos la tabla siguiente donde la última fila es la suma de las filas anteriores:

X	Y	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$Y_i - \bar{Y}$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
2	5	-1.875	3.515625	-5.75	33.0625	10.78125
4	10	0.125	0.015625	-0.75	0.5625	-0.09375
3	8	-0.875	0.765625	-2.75	7.5625	2.40625
6	18	2.125	4.515625	7.25	52.5625	15.40625
3	6	-0.875	0.765625	-4.75	22.5625	4.15625
5	15	1.125	1.265625	4.25	18.0625	4.78125
6	20	2.125	4.515625	9.25	85.5625	19.65625
2	4	-1.875	3.515625	-6.75	45.5625	12.65625
31	86	0.000	18.875000	0.00	265.5000	69.75000