

EJERCICIOS RESUELTOS DE ESTADÍSTICA BIDIMENSIONAL.

1. Cinco niñas de 2,3,5,7 y 8 años de edad pesan respectivamente 14, 20, 30, 42 y 44 kilos.
- a) Hallar la ecuación de la recta de regresión del peso sobre la edad.
- b) Cual será el peso aproximado de una niña de 6 años de edad.

SOLUCIÓN

Construir la tabla de frecuencias.

	X_i	Y_i	f_i	$X_i \cdot f_i$	$Y_i \cdot f_i$	$f_i X_i^2$	$f_i Y_i^2$	$f_i \cdot X_i \cdot Y_i$
	2	14	1	2	14	4	196	28
	3	20	1	3	20	9	400	60
	5	30	1	5	30	25	900	150
	7	42	1	7	42	49	1764	294
	8	44	1	8	44	64	1936	352
Σ	25	150	5	25	150	151	5196	884

- I. Hallamos las **medias** de X y de Y.

$$\bar{x} = \frac{\sum_{i=1}^n x_i f_i}{\sum f_i} = \frac{25}{5} = 5$$

$$\bar{y} = \frac{\sum_{i=1}^n y_i f_i}{\sum f_i} = \frac{150}{5} = 30$$

- II. Calculamos la **covarianza**. (La covarianza indica el sentido de la correlación entre las variables):

Si $\sigma_{xy} > 0$ La correlación es directa.

Si $\sigma_{xy} < 0$ La correlación es inversa.

$$\sigma_{xy} = \frac{\sum_{i=1}^n x_i y_i}{\sum f_i} - \bar{x} \cdot \bar{y} = \frac{884}{5} - 5 \cdot 30 = 26,8$$

III. Calculamos las **desviaciones típicas**.

$$s_x = \sqrt{\sum_{i=1}^n \frac{x_i^2 f_i}{\sum f_i} - \bar{x}^2} = \sqrt{\frac{151}{5} - 5^2} = 2,28$$

$$s_y = \sqrt{\sum_{i=1}^n \frac{y_i^2 f_i}{\sum f_i} - \bar{y}^2} = \sqrt{\frac{5196}{5} - 30^2} = 2,28$$

La recta de regresión de Y sobre X se utiliza para estimar los valores de la Y a partir de los de X y viceversa.

2. En una empresa se seleccionaron cinco trabajadores, se anotaron sus años de servicio y el tiempo en horas solicitado en el último mes. Los resultados obtenidos fueron:

x	1	3	2	4	5	4
y	1	1	3	4	6	5

- a) Representa los datos anteriores. Razonar si los datos muestran correlación positiva o negativa.
b) Calcular el coeficiente de correlación e interprétalo en términos de la situación real.

SOLUCIÓN

Construir la tabla de frecuencias.

	X_i	Y_i	f_i	X_i · f_i	Y_i · f_i	f_i · X_i²	f_i · Y_i²	f_i · X_i · Y_i
	1	1	1	1	1	1	1	1
	3	1	1	3	1	9	1	3
	2	3	1	2	3	4	9	6
	4	4	1	4	4	16	16	16

	5	6	1	5	6	25	36	30
	4	5	1	4	5	16	25	20
Σ	19	20	6	19	20	71	88	86

- a) Representa los datos anteriores. Razonar si los datos muestran correlación positiva o negativa.
- b) Calcular el coeficiente de correlación e interprétalo en términos de la situación real.

I. Hallamos las **medias** de X y de Y.

$$\bar{x} = \frac{\sum_{i=1}^n x_i f_i}{\sum f_i} = \frac{19}{6} = 3,16$$

$$\bar{y} = \frac{\sum_{i=1}^n y_i f_i}{\sum f_i} = \frac{20}{6} = 3,33$$

II. Calculamos la **covarianza**. (La covarianza indica el sentido de la correlación entre las variables):

$$\sigma_{xy} = \frac{\sum_{i=1}^n x_i y_i}{\sum f_i} - \bar{x} \cdot \bar{y} = \frac{76}{6} - 3,16 \cdot 3,33 = 2,11$$

III. Calculamos las **desviaciones típicas**.

$$s_x = \sqrt{\sum_{i=1}^n \frac{x_i^2 f_i}{\sum f_i} - \bar{x}^2} = \sqrt{\frac{71}{6} - 3,16^2} = 1,34$$

$$s_y = \sqrt{\sum_{i=1}^n \frac{y_i^2 f_i}{\sum f_i} - \bar{y}^2} = \sqrt{\frac{88}{6} - 3,33^2} = 1,88$$

IV. Calculamos el **coeficiente de correlación**.

$$r = \frac{\sigma_{xy}}{s_x \cdot s_y} = \frac{2,11}{1,34 \cdot 1,88} = 0,83$$

Existe una correlación directa entre los años de servicio y el número de vacaciones.

3. La tabla adjunta da el índice de mortalidad de una muestra de población en función del consumo diario de X cigarrillos.

<i>Nº Cigarrillos</i>	3	5	6	15	20	40	45
<i>Índice de Mortalidad</i>	0,2	0,3	0,3	0,5	0,7	1,4	1,5

- a) Determinar el coeficiente de correlación entre x e y. Predecir la mortalidad para un consumidor de 60 cigarrillos diarios.

SOLUCIÓN

Construir la tabla de frecuencias.

	X_i	Y_i	f_i	$X_i \cdot f_i$	$Y_i \cdot f_i$	$f_i \cdot X_i^2$	$f_i \cdot Y_i^2$	$f_i \cdot X_i \cdot Y_i$
	3	0,2	1	3	0,2	9	0,04	0,6
	5	0,3	1	5	0,3	25	0,09	1,5
	6	0,3	1	6	0,3	36	0,09	1,8
	15	0,5	1	15	0,5	225	0,25	7,5
	20	0,7	1	20	0,7	400	0,49	14
	40	1,4	1	40	1,4	1600	1,96	56
	45	1,5	1	45	1,5	2025	2,25	67,5
Σ	134	4,9	7	134	4,9	4320	5,17	148,9

- I. Hallamos las **medias** de X y de Y.

$$\bar{x} = \frac{\sum_{i=1}^n x_i f_i}{\sum f_i} = \frac{134}{7} = 19,14$$

$$\bar{y} = \frac{\sum_{i=1}^n y f_i}{\sum f_i} = \frac{4,9}{7} = 0,7$$

- II. Calculamos la **covarianza**. (La covarianza indica el sentido de la correlación entre las variables):

$$\sigma_{xy} = \frac{\sum_{i=1}^n x_i y_i}{\sum f_i} - \bar{x} \cdot \bar{y} = \frac{148,9}{7} - 19,14 \cdot 0,7 = 7,87$$

- III. Calculamos las **desviaciones típicas**.

$$s_x = \sqrt{\sum_{i=1}^n \frac{x_i^2 f_i}{\sum f_i} - \bar{x}^2} = \sqrt{\frac{4320}{7} - 19,14^2} = 15,83$$

$$s_y = \sqrt{\sum_{i=1}^n \frac{y_i^2 f_i}{\sum f_i} - \bar{y}^2} = \sqrt{\frac{5,17}{7} - 0,7^2} = 0,498$$

- IV. Calculamos el **coeficiente de correlación**.

$$r = \frac{\sigma_{xy}}{s_x \cdot s_y} = \frac{7,87}{15,83 \cdot 0,498} = 0,997$$

- V. Calculamos la **recta de regresión** para poder predecir la mortalidad para un consumidor de 60 cigarrillos. (En este caso de **y** sobre **x**).

$$(y - \bar{y}) = \frac{\sigma_{xy}}{s_x^2} (x - \bar{x})$$

$$(y - 0,7) = \frac{7,87}{15,83^2} (60 - 19,14)$$

$$y = 1,98$$

El índice de mortalidad es de 1,98.

4. La media de pesos de una población es de 65 kg y la de sus alturas de 170 cm, mientras que las desviaciones típicas son de 5 kg y 10 cm respectivamente y la covarianza de ambas variables es 40.

- a) Calcular la recta de regresión de los pesos respecto de las estaturas.
b) ¿Cuánto se estima que pesará un individuo que mide 180 cm de altura?

SOLUCIÓN

- b) Calcular la recta de regresión de los pesos respecto de las estaturas.

De los datos del problema se que:

- I. Hallamos las **medias** de X(Pesos) y de Y(Alturas).

$$\bar{x} = 65$$

$$\bar{y} = 170$$

- II. Calculamos la **covarianza**. (La covarianza indica el sentido de la correlación entre las variables):

$$\sigma_{xy} = 40$$

- III. Calculamos las **desviaciones típicas**.

$$s_x = 5$$

$$s_y = 10$$

- IV. Calculamos la **recta de regresión** de los pesos respecto a las alturas.

$$(x - \bar{x}) = \frac{\sigma_{xy}}{s_y^2} (y - \bar{y})$$

- b) ¿Cuánto se estima que pesará un individuo que mide 180 cm de altura?

$$(x - 65) = \frac{40}{170^2} (180 - 170)$$

$$x = 65,1$$

Un individuo que midiera 180 cm de altura se estima que pesará 65,1 kg.

5. Una compañía de seguros considera que el número de vehículos (Y) que circulan por una autopista, puede ponerse en función del número de accidentes (X) que ocurren en ella. Durante cinco días se obtuvo los siguiente resultado.

x	5	7	2	1	9
y	15	18	10	8	20

- Calcula el coeficiente de correlación.
- Si ayer se produjeron 6 accidentes. ¿ Cuantos vehículos podemos suponer que circulaban por la autopista?
- ¿Es buena esta predicción?

SOLUCIÓN

Construir la tabla de frecuencias.

	X_i	Y_i	f_i	$X_i \cdot f_i$	$Y_i \cdot f_i$	$f_i \cdot X_i^2$	$f_i \cdot Y_i^2$	$f_i \cdot X_i \cdot Y_i$
	5	15	1	5	15	25	225	75
	7	18	1	7	18	49	324	126
	2	10	1	2	10	4	100	20
	1	8	1	1	8	1	64	8
	9	20	1	9	20	81	400	180
Σ	24	71	5	24	71	160	1113	409

- I. Calcula el coeficiente de correlación.

- I. Hallamos las **medias** de X y de Y.

$$\bar{x} = \frac{\sum_{i=1}^n x_i f_i}{\sum f_i} = \frac{24}{5} = 4,8$$

$$\bar{y} = \frac{\sum_{i=1}^n y f_i}{\sum f_i} = \frac{71}{5} = 14,2$$

- II. Calculamos la **covarianza**. (La covarianza indica el sentido de la correlación entre las variables):

$$\sigma_{xy} = \frac{\sum_{i=1}^n x_i y_i}{\sum f_i} - \bar{x} \cdot \bar{y} = \frac{409}{5} - 4,8 \cdot 14,2 = 13,64$$

- III. Calculamos las **desviaciones típicas**.

$$s_x = \sqrt{\sum_{i=1}^n \frac{x_i^2 f_i}{\sum f_i} - \bar{x}^2} = \sqrt{\frac{160}{5} - 4,8^2} = 2,99$$

$$s_y = \sqrt{\sum_{i=1}^n \frac{y_i^2 f_i}{\sum f_i} - \bar{y}^2} = \sqrt{\frac{1113}{5} - 14,2^2} = 4,57$$

- IV. Calculamos el **coeficiente de correlación**.

$$r = \frac{\sigma_{xy}}{s_x \cdot s_y} = \frac{13,64}{2,99 \cdot 4,57} = 0,995$$

- b) Si ayer se produjeron 6 accidentes. ¿ Cuantos vehículos podemos suponer que circulaban por la autopista?

En este apartado lo que nos piden es calcular la recta de regresión de y (Número de vehículos) sobre x (Número de accidentes).

$$(y - \bar{y}) = \frac{\sigma_{xy}}{s_x^2} (x - \bar{x})$$

$$(y - 14,2) = \frac{13,64}{2,99^2} (6 - 4,8)$$

$$y = 16,03$$

Si hubo 6 accidentes se supone que circulaban 16 vehículos por la autopista.

- c) ¿Es buena esta predicción?

Si es buena esta predicción pues el coeficiente de correlación entre el número de accidentes y la cantidad de vehículos circulando es muy alto casi pegado a 1. ($r=0,995$)

6. Un conjunto de datos bidimensionales tiene coeficiente de correlación $r=0,8$ y las medias de las distribuciones marginales $x=3$ y $y=10$. Sin efectuar cálculos, razonar por qué las siguientes ecuaciones no pueden corresponder a la recta de regresión de y sobre x :

$$y = -2x + 16 \quad ; \quad y = 1,5x + 1 \quad ; \quad y = -3,5x - 1$$

La primera y la tercera recta no podrán ser puesto que si r (coeficiente de correlación) es positivo la covarianza también es positiva, por lo que en la recta de regresión siempre la pendiente de la recta será positiva.

7. Considera la siguiente distribución.

x	2	5	7	3	5	4
y	5	12	15	10	15	10

- a) Halla las dos rectas de regresión y represéntalas.
b) Observando el grado de proximidad entre las dos rectas, ¿Cómo crees que será la correlación entre dos variables?

SOLUCIÓN

Construir la tabla de frecuencias.

	X_i	Y_i	f_i	$X_i \cdot f_i$	$Y_i \cdot f_i$	$f_i \cdot X_i^2$	$f_i \cdot Y_i^2$	$f_i \cdot X_i \cdot Y_i$
	2	5	1	2	5	4	25	10
	5	12	1	5	12	25	144	60
	7	15	1	7	15	49	225	105
	3	10	1	3	10	9	100	30
	5	15	1	5	15	25	225	75

	4	10	1	4	10	16	100	40
Σ	26	67	6	26	67	128	819	320

V. Calcula el coeficiente de correlación.

II. Hallamos las **medias** de X y de Y.

$$\bar{x} = \frac{\sum_{i=1}^n x_i f_i}{\sum f_i} = \frac{26}{6} = 4,33$$

$$\bar{y} = \frac{\sum_{i=1}^n y_i f_i}{\sum f_i} = \frac{67}{6} = 11,16$$

VI. Calculamos la **covarianza**. (La covarianza indica el sentido de la correlación entre las variables):

$$\sigma_{xy} = \frac{\sum_{i=1}^n x_i y_i}{\sum f_i} - \bar{x} \cdot \bar{y} = \frac{409}{5} - 4,8 \cdot 14,2 = 4,94$$

VII. Calculamos las **desviaciones típicas**.

$$s_x = \sqrt{\sum_{i=1}^n \frac{x_i^2 f_i}{\sum f_i} - \bar{x}^2} = \sqrt{\frac{160}{5} - 4,8^2} = 2,99$$

$$s_y = \sqrt{\sum_{i=1}^n \frac{y_i^2 f_i}{\sum f_i} - \bar{y}^2} = \sqrt{\frac{1113}{5} - 14,2^2} = 4,57$$

VIII. Calculamos el **coeficiente de correlación**.

$$r = \frac{\sigma_{xy}}{s_x \cdot s_y} = \frac{13,64}{2,99 \cdot 4,57} = 0,995$$

8. De una distribución bidimensional conocemos los siguientes datos:

- Recta de regresión de Y sobre X: $y = -0,95x + 0,97$

- Coeficiente de correlación $r = 0,98$
- Media de x: $\bar{x} = 5,4$

- a) Calcula la recta de regresión de X sobre Y.
- b) Estima el valor de y para $x=6$ y para $x=15$. ¿Qué estimación te parece mas fiable?.