COVARIANZA

La covarianza de una variable bidimensional es la media aritmética de los productos de las desviaciones de cada una de las variables respecto a sus medias respectivas.

La covarianza se representa por sxy o σxy.

$$\sigma_{xy} = \frac{\sum f_i(x_i - \bar{x})(y_i - \bar{y})}{N}$$

$$\sigma_{xy} = \frac{\sum f_i x_i y_i}{N} - \bar{x}\bar{y}$$

La covarianza indica el sentido de la correlación entre las variables

Si σxy > 0 la correlación es directa.

Si σxy < 0 la correlación es inversa.

La covarianza presenta como inconveniente, el hecho de que su valor depende de la escala elegida para los ejes.

Es decir, la covarianza variará si expresamos la altura en metros o en centímetros. También variará si el dinero lo expresamos en euros o en dólares.

CORRELACIÓN

La correlación trata de establecer la relación o dependencia que existe entre las dos variables que intervienen en una distribución bidimensional.

Es decir, determinar si los cambios en una de las variables influyen en los cambios de la otra. En caso de que suceda, diremos que las variables están correlacionadas o que hay correlación entre ellas.

Coeficiente de correlación lineal

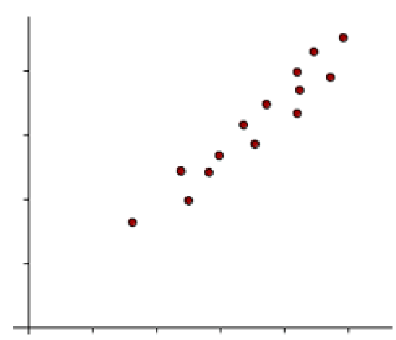
$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

covarianza

varianza de x y de y

Tipos de correlación

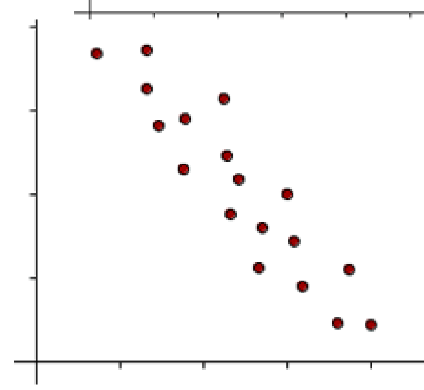
Correlación directa



La correlación directa se da cuando al aumentar una de las variables la otra aumenta.

La recta correspondiente a la nube de puntos de la distribución es una recta creciente.

Correlación inversa



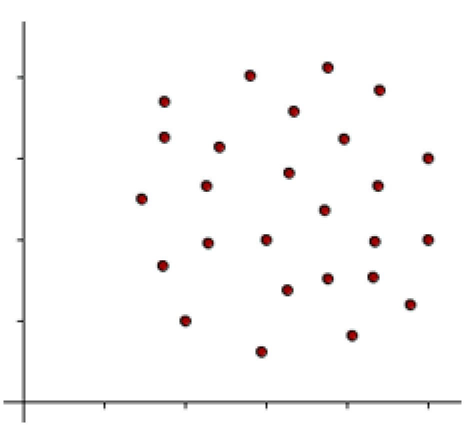
La correlación inversa se da cuando al aumentar una de las variables la otra disminuye.

La recta correspondiente a la nube de puntos de la distribución es una recta decreciente.

Correlación nula

La correlación nula se da cuando no hay dependencia de ningún tipo entre las variables.

En este caso se dice que las variables son incorreladas y la nube de puntos tiene una forma redondeada.

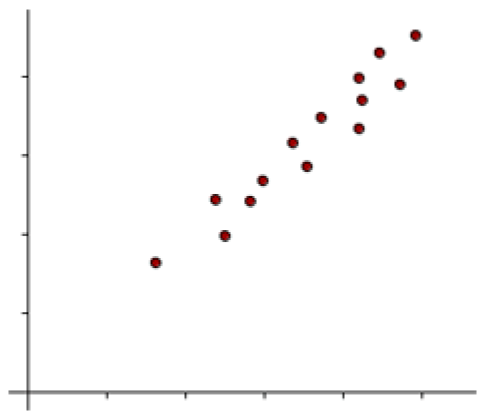


Grado de correlación

El grado de correlación indica la proximidad que hay entre los puntos de la nube de puntos. Se pueden dar tres tipos:

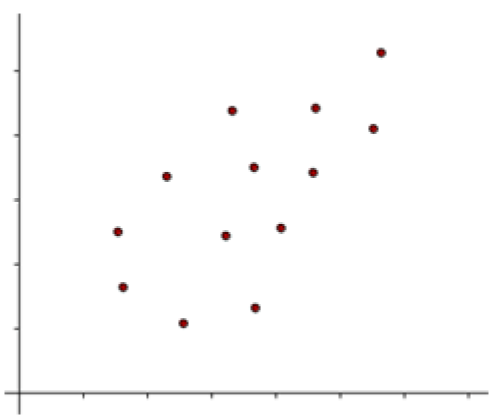
1. Correlación fuerte

La correlación será fuerte cuanto más cerca estén los puntos de la recta.



2. Correlación débil

La correlación será débil cuanto más separados estén los puntos de la recta.



EJEMPLO1: Un centro comercial sabe en función de la distancia, en kilómetros, a la que se sitúe de un núcleo de población, acuden los clientes, en cientos, que figuran en la tabla:

Nº de Clientes (X)	Distancia (Y)
8	15
7	19
6	25
4	23
2	34
1	40

a. Calcular el coeficiente de correlación lineal.
b. Si el centro comercial se sitúa a 2 km, ¿cuántos clientes puede esperar?
c. Si desea recibir a 5 clientes, ¿a qué distancia del núcleo de población debe situarse?

FORMULA DE LA VARIANZA

$$\sigma^2 = \sum_{i=1}^{n} \frac{x_i^2}{N} - \bar{x}^2$$

DESVIACIÓN TÍPICA

$$\sigma = \sqrt{\sum_{i=1}^{n} \frac{x_i^2 f_i}{N} - \bar{x}^2}$$

coeficiente de correlación lineal

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

Solución

Sea X = N° de clientes y Y=Distancia

x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
8	15	120	64	225
7	19	133	49	361
6	25	150	36	625
4	23	92	16	529
2	34	68	4	1 156
1	40	40	1	1 600
28	156	603	170	4 496

$$\bar{x} = \frac{28}{6} = 4.67 \qquad \bar{y} = \frac{156}{6} = 26$$

$$\sigma_x^2 = \frac{170}{6} - 4.67^2 = 6.53 \qquad \sigma_y^2 = \frac{4496}{6} - 26^2 = 73.33$$ Varianza

$$\sigma_x = \sqrt{6.53} = 2.55 \qquad \sigma_y = \sqrt{73.33} = 8.56$$ Desviación típica

$$\sigma_{xy} = \frac{603}{6} - 4.677 \cdot 26 = -20.92$$ Covarianza

a) $$r = \frac{-20.92}{2.55 \cdot 8.56} = -0.96$$

Correlación negativa muy fuerte.

b) $$x - 4.67 = \frac{-20.92}{73.33}(y - 26) \qquad x = -0.29y + 12.09$$

$$x = -0.29 \cdot 2 + 12.09 = 11.51 \simeq 12 \text{ clientes}$$

c) $$y - 26 = \frac{-20.92}{6.53}(x - 4.67) \qquad y = -3.2x + 40.96$$

$$y = -3.2 \cdot 5 + 40.96 = 24.96 \text{ km}$$

REGRESIÓN

Recta de regresión

La recta de regresión es la que mejor se ajusta a la nube de puntos.

La recta de regresión pasa por el punto $(\bar{x}, \bar{y})$ llamado centro de gravedad.

Recta de regresión de Y sobre X

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

La pendiente de la recta es el cociente entre la covarianza y la varianza de la variable X.

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

Recta de regresión de X sobre Y

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

La pendiente de la recta es el cociente entre la covarianza y la varianza de la variable Y.

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

Si la correlación es nula, r = 0, las rectas de regresión son perpendiculares entre sí, y sus ecuaciones son:

$y = \bar{y}$ $\qquad\qquad$ $x = \bar{x}$

EJEMPLO1.

Las notas de 12 estudiantes de una clase en matemáticas y física son las siguientes

Matemáticas	Física
2	1
3	3
4	2
4	4
5	4
6	4
6	6
7	4
7	6
8	7
10	9
10	10

Hallar las **rectas de regresión** y representarlas.

Solución: sea x_1 =matemática y y_1= física

x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
2	1	2	4	1
3	3	9	9	9
4	2	8	16	4
4	4	16	16	16
5	4	20	25	16
6	4	24	36	16
6	6	36	36	36
7	4	28	49	16
7	6	42	49	36
8	7	56	64	49
10	9	90	100	81
10	10	100	100	100
72	60	431	504	380

Paso 1: Hallamos las medias aritméticas para *x* y para *y*

$$\bar{x} = \frac{72}{12} = 6 \qquad\qquad \bar{y} = \frac{60}{12} = 5$$

Paso:2 Calculamos la covarianza

$$\sigma_{xy} = \frac{431}{12} - 6 \cdot 5 = 5.92$$

Paso3: Calculamos las varianzas.

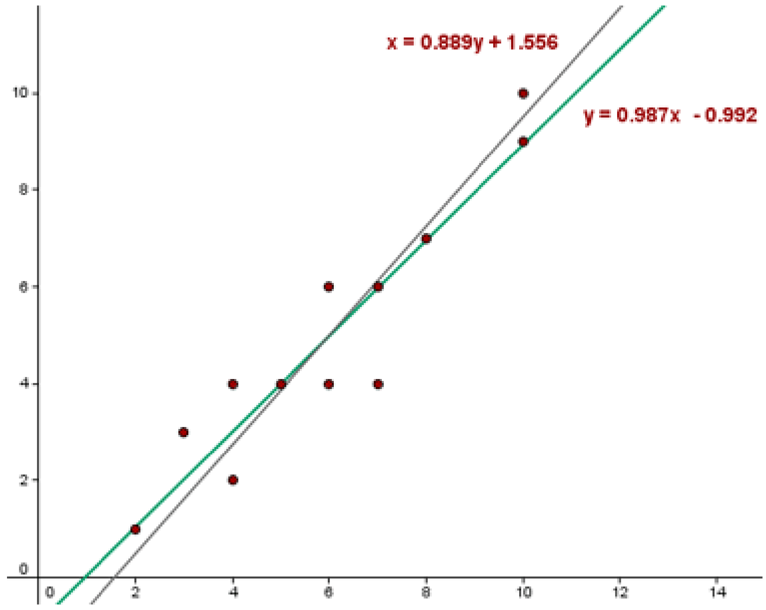
$$\sigma_x^2 = \frac{504}{12} - 6^2 = 6 \qquad\qquad \sigma_y^2 = \frac{380}{12} - 25 = 6.66$$

Recta de regresión de Y sobre X.

$$y - 5 = \frac{5.92}{6}(x - 6) \qquad\qquad y = 0.987x - 0.922$$

Recta de regresión de X sobre Y.

$$x - 6 = \frac{5.92}{6.66}(y - 5) \qquad\qquad x = 0.889y + 1.556$$



EJEMPLO2

Cinco niños de 2, 3, 5, 7 y 8 años de edad pesan, respectivamente, 14, 20, 32, 42 y 44 kilos.

a. Hallar la ecuación de la recta de regresión de la edad sobre el peso.
b. ¿Cuál sería el peso aproximado de un niño de seis años?

Solución: sea **X** la edad *y* **Y** los pesos de los niños

x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
2	14	4	196	28
3	20	9	400	60
5	32	25	1 024	160
7	42	49	1 764	294
8	44	64	1 936	352
25	152	151	5 320	894

$\bar{x} = \frac{25}{5} = 5$ $\qquad$ $\bar{y} = \frac{152}{5} = 30.4$

$\sigma_x^2 = \frac{151}{5} - 5^2 = 5.2$ $\qquad$ $\sigma_y^2 = \frac{5320}{5} - 30.4^2 = 139.84$

$\sigma_{xy} = \frac{894}{5} - 5 \cdot 30.\ 4 = 26.8$

$x - 5 = 0.192\ (y - 30)$ $\qquad$ $x = 0.192y - 0.76$

$y - 30.4 = 5.15\ (x - 5)$ $\qquad$ $y = 5.15x + 4.65$

$y = 5.15 \cdot 6 + 4.65 = 35.55\ \text{Kg}$

ACTIVIDAD

1. Las notas obtenidas por cinco alumnos en Matemáticas y Química son:

Matemáticas	Química
6	6. 5
4	4. 5
8	7
5	5
3. 5	4

Determinar las rectas de regresión y calcular la nota esperada en Química para un alumno que tiene 7.5 en Matemáticas.

2. Las estaturas y pesos de 10 jugadores de baloncesto de un equipo son:

Estatura (X)	Peso (Y)
186	85
189	85
190	86
192	90
193	87
193	91
198	93
201	103
203	100
205	101

Calcular:

a. La recta de regresión de Y sobre X.
b. El coeficiente de correlación.
c. El peso estimado de un jugador que mide 208 cm.