

Regresión y Correlación

Regresión

La regresión mide en forma funcional, a través de una ecuación, la posible relación entre las variables con el objetivo de predecir una de ellas en función de la/s otra/s.

Correlación

La correlación se dirige sobre todo a medir la intensidad de la asociación entre variables numéricas.

¿Con qué datos contamos para llevar a cabo un análisis?

Disponemos de n observaciones de dos variables aleatorias medidas en los mismos individuos, del mismo conjunto.

Individuo	Variable X	Variable Y
1	X_1	Y_1
2	X_2	Y_2
\vdots	\vdots	\vdots
n	X_n	Y_n

¿Cómo puede visualizarse la información recabada?

Gráfico de dispersión

Es una gráfica que se representa en el sistema de ejes cartesianos los pares ordenados, correspondientes a los datos apareados que resultan de las mediciones.

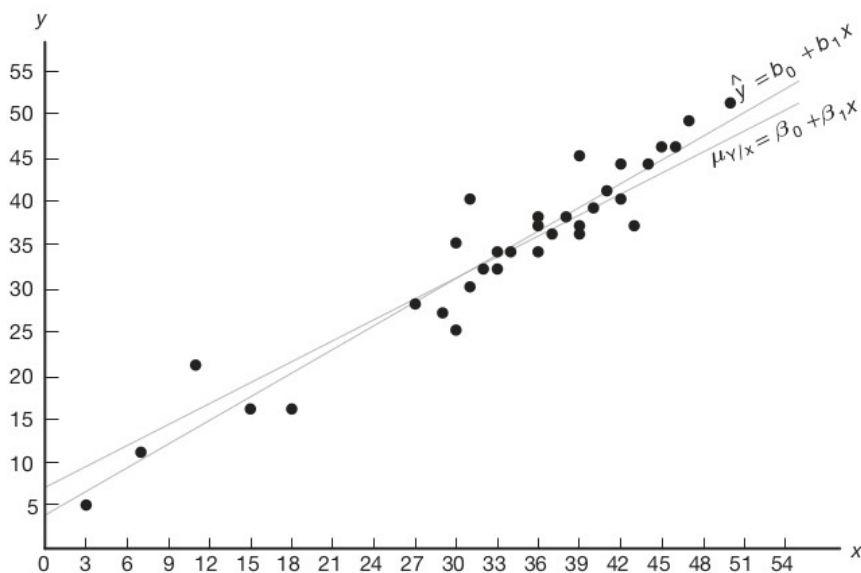


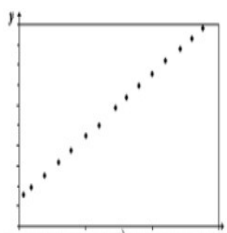
Figura 11.3: Diagrama de dispersión con rectas de regresión.

Variables que intervienen

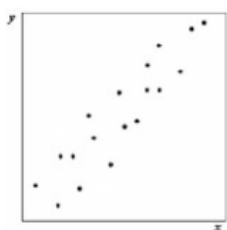
- x e y determinísticas: Conocido el valor de X, el valor de Y queda perfectamente establecido.
- x: determinística e y: probabilística (Análisis de regresión lineal).
- x e y probabilísticas (Análisis de correlación): Conocido el valor de X, el valor de Y no queda perfectamente establecido.

Tipos de relaciones

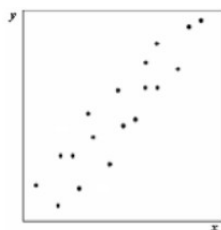
Perfectas



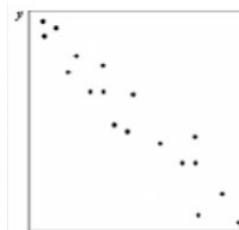
Imperfectas



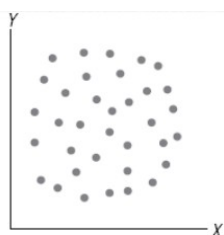
Positivas



Negativas



Sin Relación



Regresión Lineal

1º Realizar el diagrama de dispersión Lineal.

Recta de estimación o regresión: $y = bx + a$

Correlación lineal

- Coeficiente de determinación (D)

$$D = r^2$$

Campo de variación de $D: 0 \leq D \leq 1$

Si $D \geq 0,5$ (50%) \rightarrow Existe una fuerte incidencia entre x e y.

- Coeficiente de correlación (r)

$$r = \frac{\text{cov}(x, y)}{\delta x \cdot \delta y} \quad \text{siendo} \quad \text{cov}(x, y) = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{r}$$

Si $\text{cov}(x, y) > 0$, x e y están relacionados positivamente.