CONTENIDO

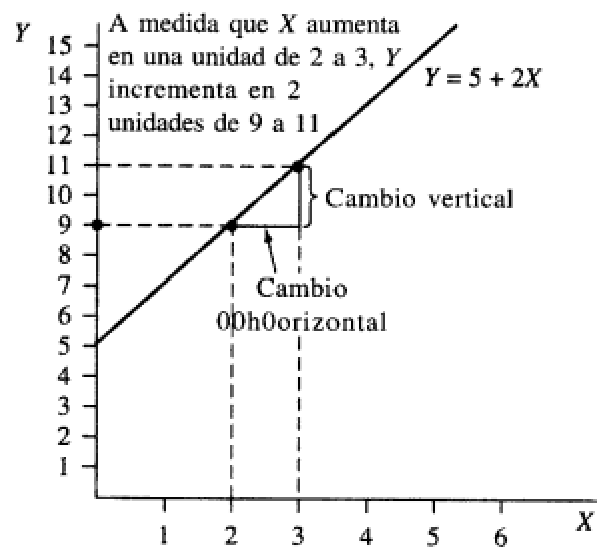
Objetivo 3.2. Análisis de regresión lineal simple y correlación

a) Ecuación de la recta de estimación por el método de mínimos cuadrados,
b) Coeficientes de correlación y determinación,
c) Error estándar de estimación.

(Desarrollo)

a) Ecuación de la recta de estimación por el método de mínimos cuadrados

La ecuación de una recta puede expresarse como Y = aX + b, en donde 'a' es la pendiente de la recta y 'b' el intercepto con el eje Y. Por ejemplo, Y = 2X + 5 es una línea recta con pendiente positiva, que se representa así:



La ecuación de la recta representa una relación entre las variables X e Y. Es una relación determinística, lo que significa que conociendo el valor de X se obtiene el valor de Y mediante el cálculo sencillo de la ecuación Y = aX + b. Esta ecuación matemática representa un modelo el cual permite conocer cada valor de Y a partir de un valor conocido de X.

Desafortunadamente, muy pocas relaciones en el mundo de los negocios y las financias son así de exactas. Lo más común es que siempre exista un grado de aleatoriedad en dicha relación. En este caso, se dice que la relación es de naturaleza estocástica o no determinístico.

El modelo de regresión lineal es un método por el cual se puede encontrar una ecuación de la recta que se ajusta de la mejor manera a un conjunto de datos aleatorios indicando la posible relación que existe entre las variables implicadas.

Para encontrar la ecuación de la recta en el modelo de regresión lineal se deben hallar los valores de la pendiente 'a' y el intercepto 'b' que la definen.

Una forma de obtener estos valores es mediante el método de Mínimos Cuadrados. No estudiaremos aquí este método, el cual puede ser revisado en la bibliografía sugerida. Solo indicaremos utilizaremos las fórmulas resultantes de la aplicación del método mencionado, las cuales expresamos a continuación:

Pendiente de la recta de regresión: $a = \dfrac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2} = \dfrac{n\,\Sigma XY - (\Sigma X)(\Sigma Y)}{n\,\Sigma X^2 - (\Sigma X)^2}$

Intercepto de la recta de regresión: $b = \bar{Y} - a\,\bar{X}$

Donde $\bar{X}$, $\bar{Y}$ son las medias de los valores de X y los valores de Y.

Ejercicio 1

Dados los siguientes datos para X e Y:

X	28, 54, 67, 37, 41, 69, 76
Y	14, 21, 36, 39, 18, 54, 52

i. Haga un diagrama de dispersión para los datos
ii. ¿Qué sugieren los datos sobre una relación entre X y Y?
iii. Construya la recta de regresión lineal por el método explicado.

---

b) Coeficientes de correlación y determinación

El coeficiente de correlación lineal, r, es la medida numérica de la fuerza de la relación lineal entre dos variables. El coeficiente de correlación lineal, r, siempre tiene un valor entre –1 y +1. Un valor de +1 significa una correlación positiva perfecta, y un valor de –1 muestra una correlación negativa perfecta. El valor de r está definido por la fórmula:

$$r = \frac{SCxy}{\sqrt{(SCx)(SCy)}}$$

Donde, los valores de *SCx*, *SCy* y *SCxy* son las sumas de los cuadrados de x, y, y los productos xy, respectivamente:

$$SCx = \Sigma X^2 - \frac{(\Sigma X)^2}{n} \qquad SCy = \Sigma Y^2 - \frac{(\Sigma Y)^2}{n} \qquad SCxy = \Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{n}$$

Por otra parte, el coeficiente de determinación, r^2, es otra medida importante de la bondad de ajuste. Se halla mediante la fórmula:

$$r^2 = \sqrt{\frac{\text{desviación explicada}}{\text{desviación total}}} = \frac{(SCxy)^2}{\sqrt{(SCx)(SCy)}}$$

Este coeficiente proporciona una medida de la bondad de ajuste porque revela qué porcentaje del cambio en Y se explica por un cambio en X.

<u>Ejercicio 2</u>

Para los datos del Ejercicio 1, encuentre el coeficiente de correlación r y el coeficiente de determinación r^2.

---

c) Error estándar de estimación.

El error estándar de estimación, *Se*, es una medida del grado de dispersión de los valores Y_i alrededor de la recta de regresión. Mide la variación promedio de los puntos de datos alrededor de la recta de regresión que se utiliza para estimar Y y por ende proporciona una medida del error que se presentará en dicha estimación. A continuación, se presenta su fórmula:

$$Se = \sqrt{\frac{\Sigma(Y_i - \hat{Y}_i)^2}{n-2}}$$

Sin embargo, para efectos de facilitar el cálculo, la fórmula anterior puede ser reemplazada por:

$$Se = \sqrt{CME}$$

En este caso, *CME* representa el cuadrado medio del error, el cual se calcula como:

$$CME = \frac{SCE}{n-2}$$

Y *SCE* es la suma de cuadrados del error, que se obtiene mediante:

$$SCE = SCy - \frac{(SCxy)^2}{SCx}$$

Ejercicio 3

Para los datos del Ejercicio 1, calcule el error estándar de estimación.

Ayuda: Todo lo que necesitamos calcular puede ser obtenido fácilmente si se anexan a nuestra tabla de frecuencias las columnas X, Y, XY, X^2, Y^2. Al totalizar cada una de estas columnas encontraremos las sumas requeridas en las fórmulas presentadas.