Una pregunta interesante: De acuerdo a este ajuste, a una persona que mida cero metros, le corresponde un peso de -170 Kilos, es claro que este valor negativo resulta contra intuitivo ¿Cuál es el problema ó cómo podemos interpretar o entender este?

Toda la descripción del método de mínimos cuadrados dada, no asume supuestos sobre una distribución de probabilidades para Y ó X o $Y|X=x$ [1], de manera que el método de mínimos cuadrados en su formulación original NO es un procedimiento estadístico, es una técnica determinista para el ajuste de curvas.

A continuación evaluaremos que tan pequeño es el mínimo encontrado, con este fin evaluamos Λ en el punto $(\hat{\beta}_0, \hat{\beta}_1)$

$$\Lambda(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$$

(1) En todo caso, para la existencia de soluciones sólo se asume $\mathbb{P}(X = c) \neq 1$.

$$\Delta(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{N} \left\{ y_i - \left[(\bar{y} - \hat{\beta}_1 \bar{x}) + \hat{\beta}_1 x_i \right] \right\}^2$$

$$= \sum_{i=1}^{N} \left\{ (y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{x}) \right\}^2$$

$$= \sum_{i=1}^{N} \left[(y_i - \bar{y})^2 - 2\hat{\beta}_1 (y_i - \bar{y})(x_i - \bar{x}) + \hat{\beta}_1^2 (x_i - \bar{x})^2 \right]$$

$$= \sum_{i=1}^{N} (y_i - \bar{y})^2 - 2\hat{\beta}_1 \sum_{i=1}^{N} (y_i - \bar{y})(x_i - \bar{x}) + \hat{\beta}_1^2 \sum_{i=1}^{N} (x_i - \bar{x})^2$$

Pero $\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) = \hat{\beta}_1 \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})$, luego

$$\Delta(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{N} (y_i - \bar{y})^2 - 2\hat{\beta}_1^2 \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x}) + \hat{\beta}_1^2 \sum_{i=1}^{N} (x_i - \bar{x})^2$$

$$= \sum_{i=1}^{N} (y_i - \bar{y})^2 - \hat{\beta}_1^2 \sum_{i=1}^{N} (x_i - \bar{x})^2. \quad \cdots\cdots (\gamma')$$

Notando que $0 \leq \Delta(\hat{\beta}_0, \hat{\beta}_1)$ tenemos

$$0 \leq \hat{\beta}_1^2 \sum_{i=1}^{N} (x_i - \bar{x})^2 \leq \sum_{i=1}^{N} (y_i - \bar{y})^2,$$

de donde

$$0 \leq \Delta(\hat{\beta}_0, \hat{\beta}_1) \leq \sum_{i=1}^{N} (y_i - \bar{y})^2.$$

Esta desigualdad nos sugiere un método para re-escalar o estandarizar $\Delta(\hat{\beta}_0, \hat{\beta}_1)$, de manera que el resultado sea una cantidad que no dependa de las unidades de medición en la variable de respuesta Y y cuya magnitud se puede evaluar con mayor facilidad, la estandarización está dada por

$$D(\hat{\beta}_0, \hat{\beta}_1) = \frac{\Delta(\hat{\beta}_0, \hat{\beta}_1)}{\sum_{i=1}^{N}(y_i - \bar{y})^2},$$

lo cual supone que $\sum_{i=1}^{N}(y_i-\bar{y})^2 > 0$, pero está no resulta una restricción de importancia, como ya discutimos antes para la variable explicativa X, $\sum_{i=1}^{N}(y_i-\bar{y})^2=0$ sucedería sólo cuando $y_1 = y_2 = \cdots = y_N$; que es una situación que <u>núnca</u> nos interesaría modelar.

Asi entonces,

$$0 \leq D(\hat{\beta}_0, \hat{\beta}_1) \leq 1,$$

si $D=0$, entonces $\Delta(\hat{\beta}_0, \hat{\beta}_1)=0$ que se interpretaría como que la recta se ajusta perfectamente al conjunto de datos, de hecho sólo sucedería cuando todas las parejas (x_i, y_i) yacen sobre la línea recta determinada por mínimos cuadrados, es decir, cuando $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ $\forall\, i = 1, 2, \ldots, N$.

En el extremo opuesto, $D=1$ corresponde al caso en que $\Delta(\hat{\beta}_0,\hat{\beta}_1)=\sum_{i=1}^{n}(y_i-\bar{y})^2$ y este debe ser el peor escenario, el valor $\Delta(\hat{\beta}_0,\hat{\beta}_1)$ es un mínimo(1), pero es el mínimo más grande que pudimos obtener. De la relación (γ^*) en la página 42 se sigue que si $\Delta(\hat{\beta}_0,\hat{\beta}_1)=\sum_{i=1}^{n}(y_i-\bar{y})^2$ entonces

$$\hat{\beta}_1^2\cdot\sum_{i=1}^{N}(x_i-\bar{x})^2=0$$

y como la situación en que $x_1=x_2=\cdots=x_N$ está descartada, esto implicaría que $\hat{\beta}_1=0$ de forma que la recta de mínimos cuadrados es una constante lo cual sugiere que no existe una relación de asociación entre X y Y.

En resumen tenemos que $0\leq D(\hat{\beta}_0,\hat{\beta}_1)\leq 1$ y en la peor situación

$D(\hat{\beta}_0,\hat{\beta}_1)=1$ ($\hat{\beta}_1=0$ y no hay evidencia de asociación lineal)

(1) Se puede probar que $\Delta(\beta_0,\beta_1)$ tiene un mínimo en $\beta_0=\hat{\beta}_0$ y $\beta_1=\hat{\beta}_1$, APENDICE de Calculo Diferencial al final →

En la mejor situación

$$D(\hat{\beta}_0, \hat{\beta}_1) = 0 \quad \text{(ajuste perfecto)}$$

Podemos entonces usar a D como un índice que nos dice qué tan malo es el modelo (la línea recta) para describir la relación entre X y Y.

Alternativamente se puede definir un índice (de bondad de ajuste) dado por $B(\hat{\beta}_0, \hat{\beta}_1) = 1 - D(\hat{\beta}_0, \hat{\beta}_1)$

para el cual

$$0 \leq B(\hat{\beta}_0, \hat{\beta}_1) \leq 1.$$

En este caso $B(\hat{\beta}_0, \hat{\beta}_1) = 0 \Rightarrow$ no hay evidencia de asociación lineal entre X y Y ($\hat{\beta}_1 = 0$).

y $B(\hat{\beta}_0, \hat{\beta}_1) = 1 \Rightarrow$ ajuste perfecto.

La siguiente relación respecto a $B(\hat{\beta}_0, \hat{\beta}_1)$ es interesante:

$$B(\hat{\beta}_0, \hat{\beta}_1) = 1 - D(\hat{\beta}_0, \hat{\beta}_1)$$

$$= 1 - \frac{\Delta(\hat{\beta}_0, \hat{\beta}_1)}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

$$B(\hat{\beta}_0, \hat{\beta}_1) = 1 - \frac{\sum_{i=1}^{N}(y_i-\bar{y})^2 - \hat{\beta}_1^2 \sum_{i=1}^{N}(x_i-\bar{x})^2}{\sum_{i=1}^{N}(y_i-\bar{y})^2}$$

$$= \hat{\beta}_1^2 \frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{\sum_{i=1}^{N}(y_i-\bar{y})^2} \quad ,$$

pero $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{N}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{N}(x_i-\bar{x})^2}$, así que

$$B(\hat{\beta}_0, \hat{\beta}_1) = \frac{\left(\sum_{i=1}^{N}(y_i-\bar{y})(x_i-\bar{x})\right)^2}{\sum_{i=1}^{N}(y_i-\bar{y})^2 \cdot \sum_{i=1}^{N}(x_i-\bar{x})^2}$$

$$= \left(\frac{\sum_{i=1}^{N}(y_i-\bar{y})(x_i-\bar{x})}{\sqrt{\sum_{i=1}^{N}(y_i-\bar{y})^2}\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2}} \right)^2$$

$$= \left(\frac{\dfrac{\sum_{i=1}^{N}(y_i-\bar{y})(x_i-\bar{x})}{N}}{\sqrt{\dfrac{\sum_{i=1}^{N}(y_i-\bar{y})^2}{N}}\sqrt{\dfrac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N}}} \right)^2$$

$$= r_{XY}^2 ,$$

donde $r_{XY} = \dfrac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$ es la correlación muestral entre $x_1, \ldots, x_N$ y $y_1, \ldots, y_N$.