Curso de Probabilidad y Estadística
Regresión Lineal. Propiedades de los estimadores de mínimos cuadrados

D.Bertsekas$_{.}$*Tsitsiklis*$_{.}$*IntroductionProbability*
wackerly $_{.}$*Mendenhall*$_{.}$*Sheaffer*$_{.}$*MathematicalStatistics*

March 14, 2022

Recordando...

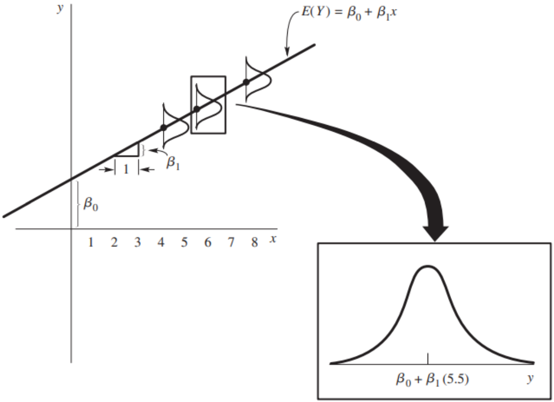
Modelo de regresión simple:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

El objetivo del modelo de regresión es tratar de explicar la relación que existe entre una variable dependiente y un conjunto de variables independientes $(X_1, ..., X_n)$, donde β_0 y β_1 son constante y ε es una variable aleatoria con media cero, llamada término del error.

Ilustración del modelo de regresión

La siguiente figura representa el modelo probabilístico $Y = \beta_0 + \beta_1 X + \varepsilon$. Cuando $x = x_0$, hay una población de posibles valores para y, la distribución de ésta población está centrada en la recta $E(y|X = x_0) = \beta_0 + \beta_1 X_0$



supuestos para el modelo de regresión lineal

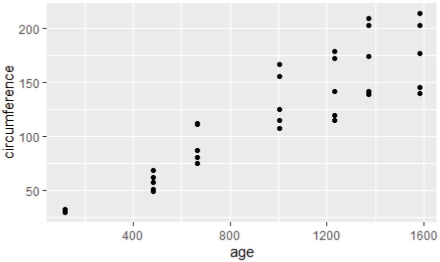
Denotamos la recta verdadera de la regresión por $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ y asumamos que se disponen de n pares de observaciones.

1. Cada x_i es un número fijo o es la realización de una v.a X_i, independiente del término de error ε_i
2. Los términos de error ε_i son variables aleatorias con media 0.
3. Las variables aleatorias ε_i tienen todas las mismas varianzas σ^2
4. Las variables aleatorias ε_i son incorrelacionadas.

El objetivo es obtener estimaciones $\hat{\beta}_0$ y $\hat{\beta}_1$ de β_0 y β_1 respectivamente para calcular la recta de regresión

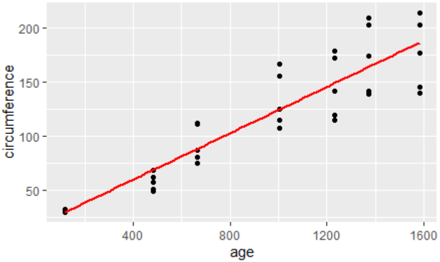
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

Ejemplo: *Se tiene una muestra de la edad en días (X) de 35 arboles de naranja asociado al tamaño de la circunferencia del tronco (mm)*



Supongamos que la recta de regresión del ejemplo anterior es:

$$circunferencia = 16.604 + 7.816 edad$$



Se estima que la circunferencia del tronco del árbol que tiene una edad de 800 días es

Estimadores de mínimos cuadrados

El método consiste en minimizar la suma de los cuadrados de las distancias verticales entre los datos y las estimaciones, es decir, minimizar la suma de los residuos al cuadrado.

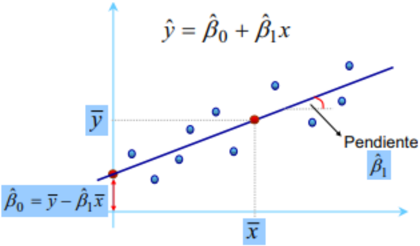
$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Estimadores de mínimos cuadrados

El resultado que se obtiene es:

$$\boxed{\hat{\beta}_1 = \frac{cov(x,y)}{s_x^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

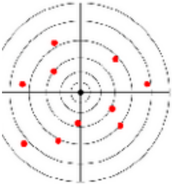
$$\boxed{\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}}$$



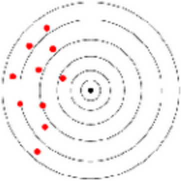
Propiedades de los estimadores del modelo de regresión

Recordando....

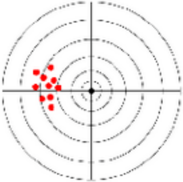
1. Un estimador (estadístico) es insesgado si el promedio (media) de todos los valores posibles de todas las muestras posibles de tamaño n de una población es igual al parámetro.
2. Un estimador es consistente si a medida que aumenta el tamaño de la muestra, las estimaciones que ésta proporciona son cada vez más próximas al valor del parámetro.
3. La eficiencia se refiere a la precisión que alcanzan los estadísticos en la estimación de los parámetros, es decir, un estimador será tanto más eficiente cuanto menos varíe de muestra a muestra de una misma población.



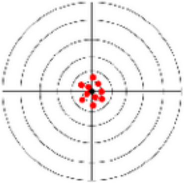
Tirador A



Tirador B



Tirador C



Tirador D

Estimadores insesgados

Funciones marginales:

$$f_{\mathbf{Y_1}}(y_1) = \int_{-\infty}^{\infty} \frac{1}{2} f_{\mathbf{X}}\left(\frac{y_1 + y_2}{2},\ \frac{y_1 - y_2}{2}\right)\, dy_2$$

Haciendo cambio de variable:

$$u = \frac{y_1 + y_2}{2}; \quad du = 1/2 dy_2$$

Luego:

Función de densidad marginal para y_1

$$f_{\mathbf{Y_1}} = \int_{-\infty}^{\infty} f_{\mathbf{X}}\Big(u, y_1 - u\Big)\, du$$