r_{XY} es un estimador (construido usando la muestra $(x_1, y_1), \ldots, (x_N, y_N)$) del coeficiente de correlación lineal de Pearson (Galton 1880)

$$R_{XY} = \frac{\mathrm{COV}(X,Y)}{\sqrt{\mathrm{VAR}(X)}\,\sqrt{\mathrm{VAR}(Y)}}$$

R_{XY} es una medida de la asociación lineal que pueda existir en las variables X y Y, un resultado que ilustra este hecho es

<u>Proposición</u>: $(r_{XY})^2 = 1$ si y sólo si existen dos constantes a y b tales que $y_i = a + b x_i$, $i = 1, 2, \ldots, N$. Además si $b > 0$ entonces $r_{XY} = 1$ y si $b < 0$ $r_{XY} = -1$.

Dadas estas observaciones, se recomienda seguir los siguientes pasos para hacer un ajuste de un modelo lineal de regresión

1.- Producir un diagrama de dispersión de las parejas $(x_1, y_1), \ldots, (x_N, y_N)$, si no existe evidencia de una tendencia lineal se busca otro modelo ó se intenta alguna transformación de los variables[1]. En caso contrario, pasemos a 2 a continuación.

2.- Calcular el coeficiente de correlación lineal de Pearson. Si r_{XY}^2 no esta cercano (suficiente-mente cercano) a 1, el modelo lineal no es adecuado. En caso contrario se pasa a 3 a continuación

3.- Ajustar la linea recta de mínimos cuadrados e interpretar los valores de los coeficientes $\hat{\beta}_0$ y $\hat{\beta}_1$.

Una pregunta inmediata con respecto a este algoritmo surge si al calcular r_{XY}^2 este valor es tal que $0 < r_{XY}^2 < 1$, ¿Cómo sabemos qué significa que r_{XY}^2 este cercano a 1 ó que r_{XY}^2 este cercano a 0?

(1) El capítulo 4 del libro "Applied Regression Analysis and Generalized Linear Models" de John Fox contiene una discusión (3ª edición, SAGE editores). También en la sección 6, capítulo 7 de "A second course in Statistics Regression Analysis" Mendenhall, Sincich (7ª edición, PEARSON editores).

En otras palabras ¿Cuándo está r_{XY}^2 cercano a 1 ó a 0?

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

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

consecuencia

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

es decir

$$(SSQ) \ldots\ldots \frac{1}{N}\sum_{i=1}^{N}(y_i-\bar{y})^2 = \overbrace{\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}^{A} + \overbrace{\frac{1}{N}\hat{\beta}_1^2 \sum_{i=1}^{N}(x_i-\bar{x})^2}^{B},$$

lo cual nos indica que la varianza de la variable de respuesta se descompone en dos partes:

A = la parte de la variabilidad de Y que la variable X <u>no</u> puede explicar a través del modelo $\boxed{\text{Lineal}}$.

B = la parte de la variabilidad de Y que la variable X <u>sí</u> logra explicar a través del modelo Lineal.

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

ó bien

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

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

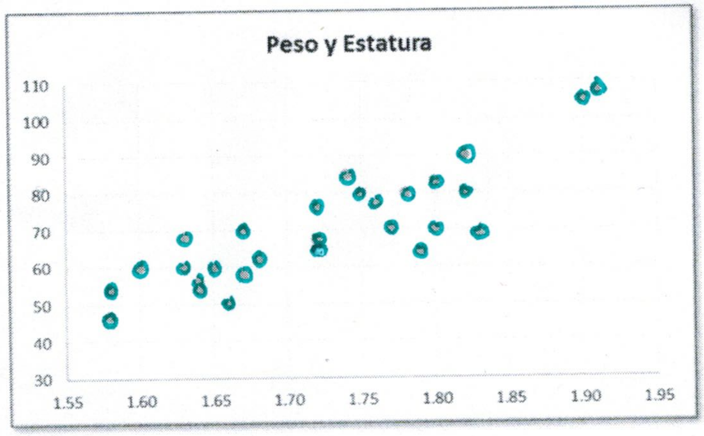
$$= r_{XY}^2 .$$

Entonces al dividir ambos lados de (SSQ) por $\frac{1}{N}\sum_{i=1}^{N}(y_i-\bar{y})^2$ tenemos

$$1 = \underbrace{\frac{\sum_{i=1}^{N}(y_i-\hat{y})^2}{\sum_{i=1}^{N}(y_i-\bar{y})^2}}_{A'} + \underbrace{r_{XY}^2}_{B'}$$

De forma que r_{XY}^2 representa la proporción de la variabilidad de Y, que logra explicar la variable X a través del modelo. La cantidad r_{XY}^2 recibe el nombre de "Coeficiente de Determinación" y representa el porcentaje de la varianza de la variable de respuesta que explica el modelo

Para el caso de los datos de Peso y estatura tenemos



Los datos presentan una tendencia monótona creciente y una posible relación lineal entre X y Y no parece inapropiada. Para calcular el coeficiente de correlación lineal

$$\bar{X} = 1.72 \quad \bar{Y} = 70.3$$

$$r_{XY} = \frac{\sum_{i=1}^{N}(y_i-\bar{y})(x_i-\bar{x})}{\sqrt{\sum_{i=1}^{N}(y_i-\bar{y})^2}\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2}} = \frac{31.309}{\sqrt{0.225}\sqrt{6010.37}}$$

$$= 0.8521$$

cuya magnitud no resulta lejana de 1, de hecho $r_{XY}^2 = 0.726$ de manera que una regresión lineal de Y como función de X explica un 72% de la variabilidad de la variable Y = Peso

Como ya se mencionó, existen parejas (x_i, y_i) tales que este par ordenado no yace sobre la linea recta $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$. Lo anterior nos sugiere considerar el modelo

$$Y = \overbrace{\beta_0 + \beta_1 X}^{(i)} + \overbrace{\varepsilon}^{(ii)},$$

donde ε es un término que da cuenta del efecto en Y de todos los factores distintos a X, los cuales no se controlan durante el estudio. Este efecto aparece en forma aditiva en el modelo, es decir se asume que todos los factores que pueden influir(1) en Y producen un efecto que queda incorporado por el término ε en el modelo, en forma aditiva.

(i) = componente determinista
(ii) = componente aleatoria

(1) Que no sean X!

RESULTADO

Sea $f: \mathbb{R}^2 \to \mathbb{R}$ una función continua y con derivadas

$$\frac{\partial^2 f(x,y)}{\partial x^2}, \quad \frac{\partial^2 f(x,y)}{\partial y^2}, \quad \frac{\partial^2 f(x,y)}{\partial y \partial x}, \quad \frac{\partial^2 f(x,y)}{\partial x \partial y}$$ que son

continuas. La función $f(x,y)$ tiene un mínimo relativo en el punto (a,b) si

1.- $\frac{\partial f}{\partial x}(a,b) = 0$ y $\frac{\partial f}{\partial y}(a,b) = 0$

2.- Para $D(a,b) = \frac{\partial^2 f}{\partial x^2}(a,b) \cdot \frac{\partial^2 f}{\partial y^2}(a,b) - \left(\frac{\partial^2 f}{\partial x \partial y}(a,b)\right)^2$

se tiene $D(a,b) > 0$ y además $\frac{\partial^2 f}{\partial x^2}(a,b) > 0$.

D es el determinante de la Matriz Hessiana H

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}$$

★ Si $\Lambda(\beta_0, \beta_1) \equiv \sum_{i=1}^{N} (y_i - (\beta_0 + \beta_1 x_i))^2$, entonces

$$\frac{\partial^2 \Lambda(\hat{\beta}_0, \hat{\beta}_1)}{\partial \beta_0^2} = 2N > 0 \qquad \frac{\partial^2 \Lambda(\hat{\beta}_0, \hat{\beta}_1)}{\partial \beta_1^2} = 2\sum_{i=1}^{N} x_i^2 > 0$$

$$\frac{\partial^2 \Lambda(\hat{\beta}_0, \hat{\beta}_1)}{\partial \beta_1 \partial \beta_0} = 2\sum_{i=1}^{N} x_i \qquad \frac{\partial^2 \Lambda(\hat{\beta}_0, \hat{\beta}_1)}{\partial \beta_0 \partial \beta_1} = 2\sum_{i=1}^{N} x_i$$

DONDE:

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

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

Calculando $D(\hat{\beta}_0, \hat{\beta}_1)$

$$D(\hat{\beta}_0, \hat{\beta}_1) = 4N \sum_{i=1}^{N} x_i^2 - 4\left(\sum_{i=1}^{N} x_i\right)^2$$

$$= 4N \sum_{i=1}^{N} x_i^2 - 4N^2 \bar{x}^2 = 4N\left(\sum_{i=1}^{N} x_i^2 - N\bar{x}^2\right)$$

$$= 4N \sum_{i=1}^{N} (x_i - \bar{x})^2 \geq 0$$

$D(\hat{\beta}_0, \hat{\beta}_1)$ es 0, sólo cuando $\sum_{i=1}^{N} (x_i - \bar{x})^2 = 0$, que como ya hemos discutido no es un caso de interés práctico. $\therefore D(\hat{\beta}_0, \hat{\beta}_1) > 0$

y usando el resultado de cálculo diferencial multivariado que se enunció tenemos que el valor $\Delta(\hat{\beta}_0, \hat{\beta}_1)$ es un mínimo.

Gauss propuso que se considerase ε una variable aleatoria. Por otra parte, se ha considerado asumir que

$$\mathbb{E}(\varepsilon) = 0 \qquad VAR(\varepsilon) = \sigma^2 \quad \text{y que}$$

$$(RL) \; \cdots \quad Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad ; \; i = 1, 2, \ldots, N.$$

Si se asume que la variable explicativa X tiene valores que son fijados en forma determinista por quienes estudian el experimento, entonces podemos considerar que $\beta_0 + \beta_1 X_i$ <u>no</u> es un término aleatorio en el modelo y por tanto el carácter aleatorio de ε_i se hereda de forma directa a Y_i, de forma que

$$\begin{aligned}
\mathbb{E}(Y_i) &= \mathbb{E}(\beta_0 + \beta_1 X_i + \varepsilon_i) \\
&= \mathbb{E}(\beta_0 + \beta_1 X_i) + \mathbb{E}(\varepsilon_i) \\
&= \beta_0 + \beta_1 X_i + 0 = \beta_0 + \beta_1 X_i \; ; \; \forall i = 1, \ldots, N.
\end{aligned}$$

También

$$\begin{aligned}
VAR(Y_i) &= VAR(\beta_0 + \beta_1 X_i + \varepsilon_i) \\
&= VAR(\varepsilon_i) = \sigma^2 \; ; \; \forall i = 1, \ldots, N
\end{aligned}$$

Para simplificar el modelo, también se considera que el error en la medición Y_i no tiene que ver, en un sentido estocástico, con el error para la medición Y_j, para cada $i \neq j$. Aunque la correlación cero <u>no</u> implica independencia de variables aleatorias, por el momento asumiremos que $\text{cov}(\varepsilon_i, \varepsilon_j) = 0 \ \forall\, i \neq j$.

Como consecuencia de este supuesto

$$\text{COV}(Y_i, Y_j) = \text{cov}(\beta_0 + \beta_1 X_i + \varepsilon_i,\ \beta_0 + \beta_1 X_j + \varepsilon_j)$$

$$= \text{cov}(\beta_0 + \beta_1 X_i,\ \beta_0 + \beta_1 X_j + \varepsilon_j) + \text{cov}(\varepsilon_i,\ \beta_0 + \beta_1 X_j + \varepsilon_j)$$

$$= \text{cov}(\varepsilon_i,\ \beta_0 + \beta_1 X_j + \varepsilon_j)$$

$$= \text{cov}(\varepsilon_i, \varepsilon_j) + \text{cov}(\varepsilon_i,\ \beta_0 + \beta_1 X_j)$$

$$= \text{cov}(\varepsilon_i, \varepsilon_j) = 0 \quad ; \ \forall\, i \neq j$$

Usaremos estas propiedades para evaluar la calidad de los coeficientes ajustados por mínimos cuadrados $\hat{\beta}_0$ y $\hat{\beta}_1$. Recordemos que

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

Como ya mencionamos antes, los valores de la variable explicativa se consideran como constantes conocidas, de forma que $\sum_{i=1}^{N}(x_i-\bar{x})^2$ tambien es una constante conocida para nosotros. Por otra parte $\sum_{i=1}^{N}(y_i-\bar{y})(x_i-\bar{x})$ involucra valores de la variable de respuesta, dado el modelo (RL), estos valores son realizaciones de las variables aleatorias $Y_1, \ldots Y_N$ y por ende, $\hat{\beta}_1$ también se puede pensar como una variable aleatoria. A continuación analizaremos a $\hat{\beta}_1$ con más detalle, observemos que

$$\begin{aligned}
\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y}) &= \sum_{i=1}^{N}(x_i-\bar{x})y_i - \sum_{i=1}^{N}(x_i-\bar{x})\bar{y} \\
&= \sum_{i=1}^{N}(x_i-\bar{x})y_i - \bar{y}\sum_{i=1}^{N}(x_i-\bar{x}) \\
&= \sum_{i=1}^{N}(x_i-\bar{x})y_i - \bar{y}\cdot 0 \\
&= \sum_{i=1}^{N}(x_i-\bar{x})y_i .
\end{aligned}$$

De forma que $\hat{\beta}_1 = \sum_{i=1}^{N} a_i Y_i$, donde $a_i = \dfrac{X_i-\bar{X}}{\sum_{i=1}^{N}(X_i-\bar{X})^2}$; $i=1,2,\ldots,N$. Lo anterior nos dice que $\hat{\beta}_1$ es una función lineal de las variables aleatorias $Y_1,\ldots,Y_N$, como consecuencia de esto se tiene que

$$\mathbb{E}(\hat{\beta}_1) = \mathbb{E}\left(\sum_{i=1}^{N} a_i Y_i\right) = \sum_{i=1}^{N} \mathbb{E}(a_i Y_i)$$

$$= \sum_{i=1}^{N} a_i \,\mathbb{E}(Y_i) = \sum_{i=1}^{N} a_i(\beta_0 + \beta_1 X_i)$$

$$= \sum_{i=1}^{N} a_i\beta_0 + \sum_{i=1}^{N} a_i\beta_1 X_i$$

$$= \beta_0 \sum_{i=1}^{N} a_i + \beta_1 \sum_{i=1}^{N} a_i X_i .$$

Pero $\sum_{i=1}^{N} a_i = \frac{1}{\sum_{i=1}^{N}(X_i-\bar{X})^2} \cdot \sum_{i=1}^{N}(X_i - \bar{X}) = 0$ y además

$$\sum_{i=1}^{N} a_i X_i = \frac{1}{\sum_{i=1}^{N}(X_i-\bar{X})^2} \sum_{i=1}^{N}(X_i-\bar{X})X_i$$

$$= \frac{1}{\sum_{i=1}^{N}(X_i-\bar{X})^2}\left\{\sum_{i=1}^{N}(X_i-\bar{X})(X_i-\bar{X}) + \sum_{i=1}^{N}(X_i-\bar{X})\bar{X}\right\}$$

$$= \frac{1}{\sum_{i=1}^{N}(X_i-\bar{X})^2} \cdot \sum_{i=1}^{N}(X_i-\bar{X})^2 = 1$$

$\therefore \quad \mathbb{E}(\hat{\beta}_1) = \beta_1$ "$\hat{\beta}_1$ es un estimador insesgado y lineal de β_1"

Por otra parte, debido a que $COV(Y_i, Y_j) = 0 \;\; \forall\, i \neq j$

$$VAR(\hat{\beta}_1) = VAR\left(\sum_{i=1}^{N} a_i Y_i\right) = \sum_{i=1}^{N} a_i^2 \, VAR(Y_i)$$

$$VAR(\hat{\beta}_1) = \sum_{i=1}^{N} \frac{(x_i-\bar{x})^2}{\left(\sum_{i=1}^{N}(x_i-\bar{x})^2\right)^2}\, \sigma^2$$

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

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

De este cálculo, observamos que: La precisición de la estimación aumenta en la medida en que se incrementa el valor de $\sum_{i=1}^{N}(x_i-\bar{x})^2$ (1). Entonces, cuando es posible elegir los valores de la variable explicativa, en el diseño del experimento, resulta conveniente que estos valores sean tales que su varianza sea lo más grande posible.

El caso de $\hat{\beta}_0$ es similar

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} = \bar{Y} - \left(\sum_{i=1}^{N} a_i Y_i\right)\bar{X}$$

$$= \sum_{i=1}^{N}\frac{1}{N}Y_i - \left(\sum_{i=1}^{N} a_i Y_i\right)\bar{X} = \sum_{i=1}^{N}\left(\frac{1}{N} - a_i\bar{X}\right)Y_i$$

(1) Si $\sum_{i=1}^{N}(x_i-\bar{x})^2$ crece, entonces $VAR(\hat{\beta}_1) = \mathbb{E}\left((\hat{\beta}_1-\beta_1)^2\right)$ decrece y en promedio la "distancia" $(\hat{\beta}_1-\beta_1)^2$ es más pequeña.

$$\hat{\beta}_0 = \sum_{i=1}^{N} b_i Y_i \,, \quad \text{donde} \quad b_i = \frac{1}{N} - a_i \bar{X}$$

Por tanto, $\hat{\beta}_0$ es también, un estimador lineal de β_0.

Como variable aleatoria tenemos que $\hat{\beta}_0$ satisface

$$\begin{aligned}
\mathbb{E}(\hat{\beta}_0) &= \sum_{i=1}^{N} b_i \mathbb{E}(Y_i) \\
&= \sum_{i=1}^{N} b_i (\beta_0 + \beta_1 X_i) \\
&= \sum_{i=1}^{N} b_i \beta_0 + \sum_{i=1}^{N} b_i \beta_1 X_i \\
&= \beta_0 \sum_{i=1}^{N} b_i + \beta_1 \sum_{i=1}^{N} b_i X_i \\
&= \beta_0 \sum_{i=1}^{N} \left(\frac{1}{N} - a_i \bar{X}\right) + \beta_1 \sum_{i=1}^{N} \left(\frac{1}{N} - a_i \bar{X}\right) X_i .
\end{aligned}$$

Pero como $\sum_{i=1}^{N} a_i X_i = 1$ entonces

$$\sum_{i=1}^{N} \left(\frac{1}{N} - a_i \bar{X}\right) X_i = \sum_{i=1}^{N} \frac{1}{N} X_i - \bar{X} \sum_{i=1}^{N} a_i X_i = \bar{X} - \bar{X} = 0 .$$

Por otra parte $\sum_{i=1}^{N} a_i = 0$ implica que

$\sum_{i=1}^{N} \left(\frac{1}{N} - a_i \bar{X}\right) = 1 - \bar{X} \sum_{i=1}^{N} a_i = 1$ .. Por tanto

$$\mathbb{E}(\hat{\beta}_0) = \beta_0 \times 1 + \beta_1 \times 0 = \beta_0 ,$$

de manera que $\hat{\beta}_0$ es un estimador lineal e insesgado de β_0. Por otra parte (debido a que $COV(Y_i, Y_j) = 0$ $\forall i \neq j$)

$$VAR(\hat{\beta}_0) = \sum_{i=1}^{N} b_i^2 VAR(Y_i)$$

$$= \sum_{i=1}^{N} b_i^2 \sigma^2$$

$$= \sigma^2 \sum_{i=1}^{N} b_i^2$$

Pero $\sum_{i=1}^{N} b_i^2 = \sum_{i=1}^{N} \left(\frac{1}{N} - a_i \bar{X}\right)^2 = \sum_{i=1}^{N} \left(\frac{1}{N^2} - \frac{2a_i\bar{X}}{N} + a_i^2 \bar{X}^2\right)$

$$= \frac{1}{N} + \bar{X}^2 \sum_{i=1}^{N} a_i^2 = \frac{1}{N} + \bar{X}^2 \sum_{i=1}^{N} \left\{ \frac{(X_i - \bar{X})^2}{\left(\sum_{i=1}^{N}(X_i - \bar{X})^2\right)^2} \right\}$$

$$= \frac{1}{N} + \bar{X}^2 \frac{1}{\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

$$= \frac{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2 + \bar{X}^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2} = \frac{\frac{1}{N}\left(\sum_{i=1}^{N} X_i^2 - N\bar{X}^2\right) + \bar{X}^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

$$= \frac{\frac{1}{N}\sum_{i=1}^{N} X_i^2 - \bar{X}^2 + \bar{X}^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2} = \frac{1}{N} \frac{\left(\sum_{i=1}^{N} X_i^2\right)}{\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

$$\therefore\ VAR(\hat{\beta}_0) = \sigma^2\left(\sum_{i=1}^{N} X_i^2\right) \Big/ N\left(\sum_{i=1}^{N}(X_i - \bar{X})^2\right)$$

Esta varianza es más pequeña, a medida que los valores de la variable explicativa produzcan un valor más pequeño del cociente

$$h(X_1, \ldots, X_N) \equiv \frac{\sum_{i=1}^{N} X_i^2}{N \sum_{i=1}^{N}(X_i-\bar{X})^2}$$

Tarea: demuestre que para un tamaño de muestra fijo N la función h se minimiza si $\bar{X}=0$.

La conclusión es que $\hat{\beta}_0$ y $\hat{\beta}_1$ son estimadores lineales e insesgados para β_0 y β_1, cuyas varianzas dependen de los valores de la variable explicativa X. El siguiente teorema establece que $\hat{\beta}_0$ y $\hat{\beta}_1$ son estimadores; que dentro de la clase o conjunto de los estimadores lineales insesgados para β_0 y β_1, tienen la varianza más pequeña

TEOREMA (GAUSS-MARKOV)

Sean $Y_1, \ldots, Y_N$ v.a. tales que

(i) $\mathbb{E}(Y_i) = \beta_0 + \beta_1 x_i$; $i=1,2,\ldots,N$,

(ii) $VAR(Y_i) = \sigma^2$; $i=1,2,\ldots,N$,

(iii) $COV(Y_i, Y_j) = 0$ $\forall i \neq j$ $(i,j=1,2,\ldots,N)$, donde $x_1, \ldots, x_N$ son constantes conocidas. Entonces los estimadores de mínimos cuadrados $\hat{\beta}_0$ y $\hat{\beta}_1$ son los Mejores Estimadores Lineales Insesgados (MELI) de β_0 y β_1, en el sentido de varianza mínima.

<u>NOTA</u>: En el idioma Inglés se dice que $\hat{\beta}_0$ y $\hat{\beta}_1$ son los BLUE (Best Linear Unbiased Estimators)

NOTA: El teorema no establece que $\hat{\beta}_0$ y $\hat{\beta}_1$ son óptimos en general. Lo que establece es que $\hat{\beta}_0$ y $\hat{\beta}_1$ son los mejores, dentro de una clase particular de estimadores, la clase ó conjunto de los estimadores lineales e insesgados.

Si quisieramos comparar $\hat{\beta}_0$ y $\hat{\beta}_1$ con los estimadores que se obtienen con otros métodos estadisticos, podríamos pensar en estimar β_0 y β_1 vía momentos o vía Máxima Verosimilitud. El método de momentos no funcionará, toda vez que este supone que las variables aleatorias que se observan son idénticamente distribuidas, que en nuestro caso no se tiene ($\mathbb{E}(Y_i)$ no es igual a $\mathbb{E}(Y_j)$).

Entonces pensaremos como estimar β_0 y β_1 vía el método de Máxima Verosimilitud. Lo anterior requiere tener la distribución conjunta para las variables que se observan.

Como la variable aleatoria ε describe el efecto de todos los factores, distintos de la variable explicativa, que no están fijos en el estudio, no parecería verosimil que tome valores en un conjunto finito ó numerable, así que suena razonable asumir que ε es una v.a. continua.

Gauss estudió el comportamiento empírico de los errores en la estimación de distancias a objetos en la boveda celeste, a partir de ese análisis sugirió que la distribución debería ser unimodal y simétrica con media 0. En particular si se piensa en ε como la suma infinita de efectos que no están bajo control en el estudio (y que no tienen relación con X) podemos conectar esta concepción de ε con el Teorema Central del Límite, que establece que bajo algunas condiciones, las sumas de variables aleatorias tienen distribución "cercana" a ser Normal.

MODELO DE REGRESION LINEAL NORMAL

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

donde ε es una v.a. con distribución $N(0,\sigma^2)$.

Los datos son generados con la relación

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \; ; \; i = 1, 2, \dots, N,$$

donde $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N$ son variables aleatorias idénticamente distribuidas e independientes según el modelo Normal $(0, \sigma^2)$.

Esta estructura conduce a la siguiente formulación alternativa del Modelo Lineal Normal, el cual está vinculado con la discusión en las primeras páginas de estas notas

" $Y_1, Y_2, \dots, Y_N$ son v.a. independientes y tales que $Y_i | X_i = x_i \sim \text{Normal}(\beta_0 + \beta_1 x_i, \sigma^2)$ "

Bajo esta premisa, la verosimilitud (la función de densidad conjunta de $Y_1, \ldots, Y_N$ condicional a $X_1 = x_1; \ldots; X_N = x_N$ es

$$L(\beta_0, \beta_1, \sigma^2) = f_{Y_1 \ldots Y_N | X_1 \ldots X_N}(y_1, \ldots, y_N | x_1, \ldots, x_N, \beta_0, \beta_1, \sigma^2)$$

$$= \prod_{i=1}^{N} f_{Y_i | X_i}(y_i | x_i, \beta_0, \beta_1, \sigma^2)$$

$$= \prod_{i=1}^{N} \left\{ (2\pi\sigma^2)^{-1/2} e^{-\frac{1}{2\sigma^2}[y_i - (\beta_0 + \beta_1 x_i)]^2} \right\}$$

$$= (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \beta_0 - \beta_1 x_i)^2}$$

de densidad conjunta de $Y_1, \dots, Y_N$ condicional a $X_1 = x_1; \dots; X_N = x_N$ es

$$L(\beta_0, \beta_1, \sigma^2) = f_{Y_1 \dots Y_N | X_1 \dots X_N}(y_1, \dots, y_N | x_1, \dots, x_N, \beta_0, \beta_1, \sigma^2)$$

$$= \prod_{i=1}^{N} f_{Y_i|X_i}(y_i | x_i, \beta_0, \beta_1, \sigma^2)$$

$$= \prod_{i=1}^{N} \left\{ (2\pi\sigma^2)^{-1/2} e^{-\frac{1}{2\sigma^2}[y_i - (\beta_0 + \beta_1 x_i)]^2} \right\}$$

$$= (2\pi\sigma^2)^{-N/2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x_i)^2}$$

$$l(\beta_0, \beta_1, \sigma^2) = \operatorname{Log} L(\beta_0, \beta_1, \sigma^2)$$

$$= -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x_i)^2$$

Es claro que maximizar $l(\beta_0, \beta_1, \sigma^2)$ c.r. a β_0 y β_1 corresponde con minimizar la suma de cuadrados

$$\sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x_i)^2 \quad \left(= \Delta(\beta_0, \beta_1) \right)$$

En consecuencia, los estimadores de máxima verosimilitud para β_0 y β_1 coinciden con los estimadores de mínimos cuadrados. El sistema de ecuaciones que se obtiene cuando $\frac{\partial l}{\partial \beta_0}$ y $\frac{\partial l}{\partial \beta_1}$ se igualan a 0, es el mismo que analizamos cuando estabamos buscando los estimadores de mínimos cuadrados. Por esta razón el sistema de ecuaciones (EN) en la página 36 se llama sistema Normal (ó ecuaciones normales). En cuanto al tercer parámetro σ^2, la derivada parcial correspondiente es

$$\frac{\partial l}{\partial \sigma^2} = -\frac{N}{2}\frac{2\pi}{2\pi\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{N}(y_i - \beta_0 - \beta_1 x_i)^2.$$

Igualando a 0 obtenemos

$$-\frac{n}{\hat{\sigma}^2} + \frac{1}{\hat{\sigma}^4}\sum_{i=1}^{N}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2,$$

de donde

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

En resumen, los estimadores de máxima verosimilitud $\hat{\theta}^{(mv)} = \left(\hat{\beta}_0^{(mv)}, \hat{\beta}_1^{(mv)}, \hat{\sigma}^{2\,(mv)}\right)$ para $\theta = (\beta_0, \beta_1, \sigma^2)$ están dados por

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

$$\hat{\beta}_0^{(mv)} = \bar{y} - \hat{\beta}_1\bar{x},$$

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

donde $\hat{\beta}_0^{(mv)}$ y $\hat{\beta}_1^{(mv)}$ coinciden con los estimadores de mínimos cuadrados para β_0 y β_1. Debido a lo anterior podemos escribir

$$\hat{\beta}_1^{(mv)} = \sum_{i=1}^{N} a_i Y_i \qquad \hat{\beta}_0^{(mv)} = \sum_{i=1}^{N} b_i Y_i \;\; ^{(1)} \quad \cdots\cdots (1)$$

y además sabemos que

$$\mathbb{E}\left(\hat{\beta}_1^{(mv)}\right) = \beta_1 \quad y \quad VAR\left(\hat{\beta}_1\right) = \frac{\sigma^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2} \cdots\cdots (2)$$

Además

$$\mathbb{E}\left(\hat{\beta}_0^{(mv)}\right) = \beta_0 \quad y \quad VAR\left(\hat{\beta}_0\right) = \frac{\sigma^2 \sum_{i=1}^{N} x_i^2}{N\sum_{i=1}^{N}(x_i-\bar{x})^2} \cdots\cdots (3)$$

(1) donde $a_i = \frac{x_i - \bar{x}}{\sum_{i}^{N}(x_i-\bar{x})^2}$ y $b_i = \frac{1}{N} - a_i\bar{x}$

Usando las relaciones (1), (2) y (3)(1), podemos deducir que

$$\hat{\beta}_1^{(MV)} \sim \text{Normal}\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}\right)$$

$$\hat{\beta}_0^{(MV)} \sim \text{Normal}\left(\beta_0, \frac{\sigma^2 \sum_{i=1}^{N} x_i^2}{N\sum_{i=1}^{N}(x_i-\bar{x})^2}\right)$$

<u>NOTA</u>: $\hat{\beta}_0^{(MV)}$ y $\hat{\beta}_1^{(MV)}$ vistos como v.a. <u>no</u> necesariamente son independientes:

$$\text{cov}(\hat{\beta}_0, \hat{\beta}_1) = \text{cov}\left(\sum_{i=1}^{N} b_i Y_i, \sum_{i=1}^{N} a_i Y_i\right)$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N} a_i b_j \,\text{cov}(Y_i, Y_j)$$

$$\overset{(1)}{=} \sum_{i=1}^{N} a_i b_i \,\text{VAR}(Y_i)$$

$$= \sigma^2 \sum_{i=1}^{N} a_i b_i$$

$$= \sigma^2 \sum_{i=1}^{N} a_i \left(\frac{1}{N} - \bar{x} a_i\right)$$

$$= \sigma^2 \left[\frac{1}{N}\sum_{i=1}^{N} a_i - \bar{x}\sum_{i=1}^{N} a_i^2\right] = -\sigma^2 \frac{\bar{x}}{\sum_{i=1}^{N}(x_i-\bar{x})^2}$$

(1) Y tomando en cuenta que $Y_i | X_i = x_i \sim \text{Normal}(\beta_0 + \beta_1 x_i, \sigma^2)$ $i=1,2,\ldots,N$, con $f_{Y_1 \cdots Y_N | X_1 = x_1 \cdots X_N = x_N} = \prod_{i=1}^{N} f_{Y_i|X_i=x_i}$ y con $X_1, \ldots, X_N$ v.a. independ. ($\Rightarrow Y_1 \cdots Y_N$ v.a. independ.)

☆ **PROPOSICIÓN**: Si $\mathbb{Y}' = (Y_1, \dots, Y_N)$ es un vector aleatorio tal que $\mathbb{Y} \sim N_N(\mu_{\mathbb{Y}}; \Sigma_{\mathbb{Y}})$ y $\mathbb{A}$ es una matriz de dimensiones $p \times N$ entonces

$$\mathbb{X} \equiv \mathbb{A}\mathbb{Y} \sim N_p(\mu_X, \Sigma_X) \quad \text{donde}$$

$$\mu_X = \mathbb{A}\mu_{\mathbb{Y}} \quad \text{y} \quad \Sigma_X = \mathbb{A}\Sigma_Y\mathbb{A}'.$$

Usaremos este resultado con $\mathbb{Y}' = (Y_1, \dots, Y_N)$,

$$\mu_{\mathbb{Y}}' = (\beta_0 + \beta_1 x_1, \dots, \beta_0 + \beta_1 x_N) \; ; \; \Sigma_{\mathbb{Y}} = \sigma^2 \mathbb{I}_{N \times N}$$

y $\mathbb{A} = \begin{pmatrix} a_1 & \cdots & a_N \\ b_1 & \cdots & b_N \end{pmatrix}$ de dimensiones $2 \times N$. Por el resultado

tenemos que

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_0 \end{pmatrix} = \mathbb{A}\mathbb{Y} \sim N_2\left(\mathbb{A}\mu_Y, \mathbb{A}\cdot\sigma^2\mathbb{I}\cdot\mathbb{A}'\right),$$

con

$$\mathbb{A}\mu_{\mathbb{Y}} = \begin{pmatrix} a_1(\beta_0+\beta_1 x_1) + \cdots + a_N(\beta_0+\beta_1 x_N) \\ b_1(\beta_0+\beta_1 x_1) + \cdots + a_N(\beta_0+\beta_1 x_N) \end{pmatrix}$$

$$= \begin{pmatrix} \beta_0 \sum_{i=1}^N a_i + \beta_1 \sum_{i=1}^N a_i x_i \\ \beta_0 \sum_{i=1}^N b_i + \beta_1 \sum_{i=1}^N b_i x_i \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_0 \end{pmatrix} \quad \text{ya que}$$

$$\sum_{i=1}^N a_i = 0 \; ; \; \sum_{i=1}^N a_i x_i = 1 \; ; \; \sum_{i=1}^N b_i = \sum_{i=1}^N \left(\tfrac{1}{N} - a_i\bar{x}\right) = 1 \; \text{y} \; \sum_{i=1}^N b_i x_i = \bar{x} - \bar{x} = 0$$

$$\mathbb{A}\cdot\sigma^2\mathbb{I}\,\mathbb{A}' = \sigma^2\mathbb{A}\mathbb{A}' = \sigma^2\begin{pmatrix} a_1,\dots,a_N \\ b_1,\dots,b_N\end{pmatrix}\begin{pmatrix} a_1 & b_1 \\ \vdots & \vdots \\ a_N & b_N\end{pmatrix}$$

$$= \sigma^2\begin{pmatrix} \sum_1^N a_i^2 & \sum_1^N a_i b_i \\ \sum_{i=1}^N b_i a_i & \sum_1^N b_i^2 \end{pmatrix}$$

$$= \sigma^2\begin{pmatrix} \dfrac{1}{\sum_{i=1}^N (x_i-\bar{x})^2} & -\dfrac{\bar{x}}{\sum_1^N (x_i-\bar{x})^2} \\ -\dfrac{\bar{x}}{\sum_1^N (x_i-\bar{x})^2} & \dfrac{\sum_1^N x_i^2}{N\sum_1^N (x_i-\bar{x})^2} \end{pmatrix}$$

como $\begin{pmatrix}\hat{\beta}_1 \\ \hat{\beta}_0\end{pmatrix}$ es <u>conjuntamente</u> normal (1) y con

matriz de varianzas covarianzas $\mathbb{A}(\sigma^2\mathbb{I})\mathbb{A}'$, tenemos

que $\hat{\beta}_1$ y $\hat{\beta}_0$ son independientes, si y sólo si

$\bar{x}=0$.

Por otra parte, como

$$\hat{\sigma}^{2\,(mv)} = \frac{1}{N}\sum_{i=1}^N (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

(1) $\hat{\beta}_1 \sim N_1$ y $\hat{\beta}_0 \sim N_1$ con $\operatorname{cov}(\hat{\beta}_1, \hat{\beta}_0) = 0 \not\Rightarrow$ independencia

se puede probar que

$$\frac{N}{\sigma^2}\hat{\sigma}^{2(MV)} \sim \chi^2_{(N-2)} \quad \cdots\cdots \quad (DV)$$

Usando este resultado tenemos que

$$\mathbb{E}\left(\frac{N}{\sigma^2}\hat{\sigma}^{2(MV)}\right) = N-2 \quad y \quad \mathbb{E}\left(\hat{\sigma}^{2(MV)}\right) = \left(\frac{N-2}{N}\right)\sigma^2$$

$$VAR\left(\frac{N}{\sigma^2}\hat{\sigma}^{2(MV)}\right) = 2(N-2) \quad y \quad VAR\left(\hat{\sigma}^{2(MV)}\right) = 2\frac{(N-2)}{N^2}\sigma^4$$

A partir de estos cálculos se puede definir un estimador <u>insesgado</u> para la varianza σ^2

$$\tilde{\sigma}^2 = \frac{1}{N-2}\sum_{i=1}^{N}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

PROPOSICION: $\hat{\beta}_0$ y $\hat{\beta}_1$ son independientes de $\hat{\sigma}^2$

<u>Dem</u>

$$D_j \equiv Y_j - \hat{\beta}_0 - \hat{\beta}_1 x_i \quad j=1,\ldots,N$$

cada D_j es una combinacion lineal de las v.a. Normales $Y_1, Y_2, \ldots, Y_N$, por lo que D_j es es una v.a. Normal $\forall j = 1,\ldots,N$. De hecho usando la proposición ★ en la página 68 el vector aleatorio $\mathbb{D}_j = (D_j, \hat{\beta}_1, \hat{\beta}_0)$; $j=1,\dots,N$, tiene distribución Normal Multivariada. En estas condiciones, para verificar la independencia entre los coeficientes $\hat{\beta}_0$, $\hat{\beta}_1$ y $\hat{\sigma}^2$ basta comprobar que

$$\operatorname{cov}(D_j, \hat{\beta}_1) = 0 \quad y \quad \operatorname{cov}(D_j, \hat{\beta}_0) = 0 \quad \cdots \quad (A)$$

$$\forall j = 1, 2, \dots, N.$$

Si sucede lo que indican las relaciones en (A), entonces $\hat{\beta}_1$ es independiente de cada D_j $\forall j=1,\dots,N$ luego $\hat{\beta}_1$ es independiente de D_j^2 $\forall j = 1, 2, \dots, N$ y por tanto $\hat{\beta}_1$ es independiente de $\frac{1}{N}\sum_{j=1}^{N} D_j^2 = \hat{\sigma}^2$. El argumento es análogo para $\hat{\beta}_0$.

En estas condiciones, para verificar la independencia entre los coeficientes estimados y $\hat{\sigma}^2$ basta con probar que

$$Cov(D_j, \hat{\beta}_1) = 0 \text{ y } Cov(D_j, \hat{\beta}_0) = 0 \;; \quad j = 1, \dots, n.$$

Considere el caso de $\hat{\beta}_1$:

$$\begin{aligned} Cov(D_j, \hat{\beta}_1) &= Cov(\, Y_j - \hat{\beta}_0 - \hat{\beta}_1 x_j, \hat{\beta}_1 \,) \\ &= Cov(\, Y_j, \hat{\beta}_1) - Cov(\hat{\beta}_0, \hat{\beta}_1 \,) - x_j Var(\hat{\beta}_1) \\ &= Cov(\, Y_j, \Sigma_{i=1}^{n} \, a_i Y_i) - Cov(\hat{\beta}_0, \hat{\beta}_1 \,) - x_j Var(\hat{\beta}_1) \\ &= Cov(\, Y_j, a_j Y_j) - Cov(\hat{\beta}_0, \hat{\beta}_1 \,) - \, x_j \, Var(\hat{\beta}_1) \\ &= \sigma^2 \, a_j + \sigma^2 \, \frac{\overline{x}}{S_{XX}} - \, x_j \, \sigma^2 \, \frac{1}{S_{XX}} \,. \end{aligned}$$

En resumen,

$$Cov\left(D_j, \hat{\beta}_1\right) = \sigma^2 a_j + \sigma^2 \frac{\overline{x}}{S_{XX}} - x_j\, \sigma^2 \frac{1}{S_{XX}}$$

$$= \sigma^2 \frac{\left(x_j - \overline{x}\right)}{S_{XX}} + \sigma^2 \frac{\overline{x}}{S_{XX}} - x_j\, \sigma^2 \frac{1}{S_{XX}}$$

$$= 0.$$

Por lo tanto, $\hat{\beta}_1$ es independiente de D_j $\forall\, j = 1, \dots, n$ y como consecuencia, $\hat{\beta}_1$ es independiente tanto de $\hat{\sigma}^2$ como de $\tilde{\sigma}^2$.

Tarea. *Pruebe que $\hat{\beta}_0$ también es independiente de $\hat{\sigma}^2$ y de $\tilde{\sigma}^2$.*

Ejercicio Probar que $\hat{\beta}_0$ es independiente de $\hat{\sigma}^2$ y de $\tilde{\sigma}^2$

Ejercicio: Probar que $\tilde{\sigma}^2$ en la página 70 de las notas es un estimador insesgado de σ^2. Encuentre la distribución de

$$\frac{(N-2)}{\sigma^2}\tilde{\sigma}^2.$$

¿Cuánto vale $VAR(\tilde{\sigma}^2)$?

Compare $VAR(\hat{\sigma}^2)$ con $VAR(\tilde{\sigma}^2)$

PROPOSICION : $\hat{\beta}_0$ y $\hat{\beta}_1$ son independientes de $\tilde{\sigma}^2$

Dem

La demostración es análoga a la de la proposición donde se prueba que $\hat{\beta}_0$ y $\hat{\beta}_1$ son independientes de $\hat{\sigma}^2$ $\blacksquare$

Tenemos que $\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}\right)$, de donde

$$Z \equiv \frac{\hat{\beta}_1 - \beta_1}{\left(\frac{\sigma^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}\right)^{1/2}} \sim N(0,1).$$

Por otra parte

$$U \equiv \frac{N-2}{\sigma^2}\tilde{\sigma}^2 \sim \chi^2_{(N-2)}$$

y además Z y U son v.a. independientes.

Entonces

$$t \equiv \frac{Z}{\sqrt{\frac{U}{N-2}}} \sim \text{Student}_{(N-2)}$$

$$t = \frac{\frac{(\hat{\beta}_1 - \beta_1)}{\tilde{\sigma}}}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2}} = \frac{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2}\,(\hat{\beta}_1-\beta_1)}{\tilde{\sigma}}$$

Sea $t_{(N-2)}^{1-\alpha/2}$ el cuantil de orden $(1-\alpha/2)$ de una distribución Student con N-2 grados de libertad. Entonces

$$\mathbb{P}\left(-t_{(N-2)}^{1-\alpha/2} \leq \frac{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2}\,(\hat{\beta}_1-\beta_1)}{\tilde{\sigma}} \leq t_{(N-2)}^{1-\alpha/2}\right) = 1-\alpha$$

$$\mathbb{P}\left(\hat{\beta}_1 - \sqrt{\frac{\tilde{\sigma}^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}}\, t_{(N-2)}^{1-\alpha/2} \leq \beta_1 \leq \hat{\beta}_1 + \sqrt{\frac{\tilde{\sigma}^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}}\, t_{(N-2)}^{1-\alpha/2}\right) = 1-\alpha$$

Usando los valores de la muestra $(x_1,y_1),\dots,(x_N,y_N)$ para calcular $\hat{\beta}_1$ y $\tilde{\sigma}^2$, se obtiene un intervalo de confianza, a nivel $(1-\alpha)*100\%$, para β_1

$$\left[\hat{\beta}_1 \pm \frac{\tilde{\sigma}}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2}}\, t_{(N-2)}^{1-\alpha/2}\right] \quad \dots\dots (I)$$

Notemos que en virtud de que $\tilde{\sigma}^2$ es un estimador de σ^2, entonces

$$\frac{\tilde{\sigma}^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}$$ es un estimador[(1)] de $VAR(\hat{\beta}_1)$.

Denotando $\widehat{VAR}(\hat{\beta}_1) \equiv \dfrac{\tilde{\sigma}^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}$

(1) Insesgado

Se puede re-escribir el intervalo de confianza para β_1 en (I) como

$$\left[\hat{\beta}_1 \pm \sqrt{\widetilde{VAR}(\hat{\beta}_1)} \cdot t_{(N-2)}^{1-\alpha/2} \right] \quad \cdots\cdots (II)$$

<u>EJERCICIO</u>: Demuestre que con un procedimiento similar, se puede probar que un intervalo de confianza a nivel $(1-\alpha)\times 100\%$ para β_0, está dado por

$$\left[\hat{\beta}_0 \pm \sqrt{\widetilde{VAR}(\hat{\beta}_0)} \; t_{(N-2)}^{1-\alpha/2} \right]$$

con $$\widetilde{VAR}(\hat{\beta}_0) = \frac{\tilde{\sigma}^2 \sum_{i=1}^{N} x_i^2}{N \sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Para el caso de los datos de estaturas y pesos tenemos

$\bar{x} = 1.724$; $\bar{y} = 70.28$; $\sum_{i=1}^{N}(x_i - \bar{x})^2 = 0.2246$; $\sum_{i=1}^{N}(y_i - \bar{y})^2 = 6010.37$

$\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) = 31.309$ de donde

$\hat{\beta}_0 = -169.94$; $\hat{\beta}_1 = 139.37$; $r_{xy} = 0.8521$ $(r_{xy}^2 = 0,726)$

Para calcular el estimador insesgado $\tilde{\sigma}^2$, es necesario calcular

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

Se puede hacer directamente, pero tambien se puede recordar la identidad

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

$$=6010.37-(139.37)^2 \times 0.225$$

$$=1647.74,$$

de forma que $\tilde{\sigma}^2=63.37$ y en consecuencia

$$\widetilde{VAR}(\hat{\beta}_1)=\frac{63.37}{0.225}=282.15$$

$$\sqrt{\widetilde{VAR}(\hat{\beta}_1)}=16.78.$$

Para el cálculo de $\widetilde{VAR}(\hat{\beta}_0)$, se puede usar que

$$VAR(\hat{\beta}_0)=\sigma^2\sum_{i=1}^{N}\left(\frac{1}{N}-a_i\bar{x}\right)^2=\sigma^2\left[\frac{1}{N}+\frac{\bar{x}^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}\right]$$

(veáse página 59 de las notas)