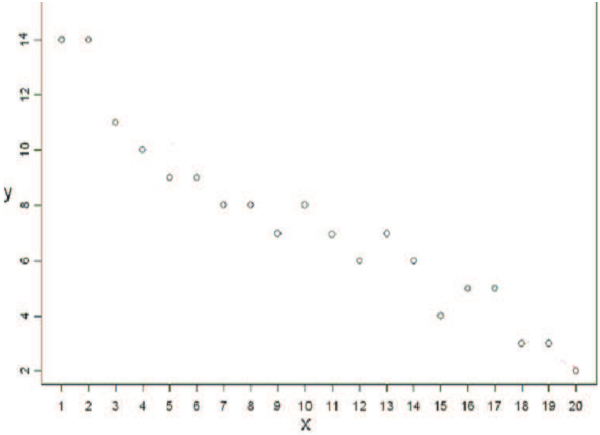
Caso Regresión Simple

En el caso más sencillo de regresión simple tendríamos

$$y_i = \beta_1 + \beta_2 x_i \quad 1 \leq i \leq n$$

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ . & . \\ . & . \\ 1 & x_n \end{pmatrix}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$



y
14
12
10
8
6
4
2
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
X

¿Cómo estimamos los parámetros?

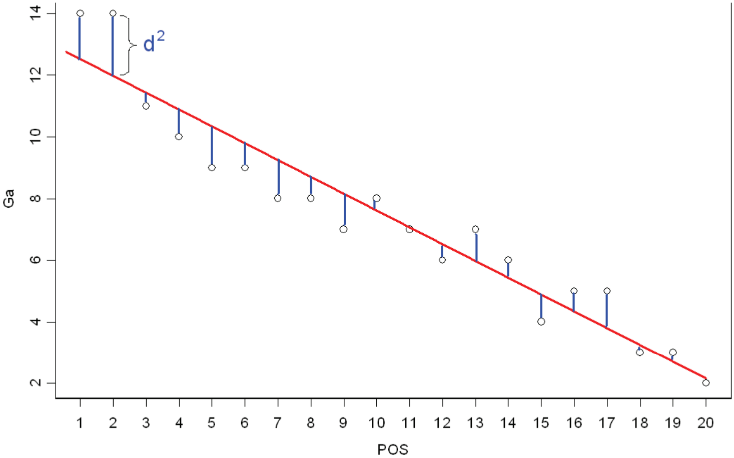
Mínimos Cuadrados

Si los puntos en un gráfico parecen seguir una recta como en el gráfico, el problema es elegir la recta que mejor ajusta los puntos.

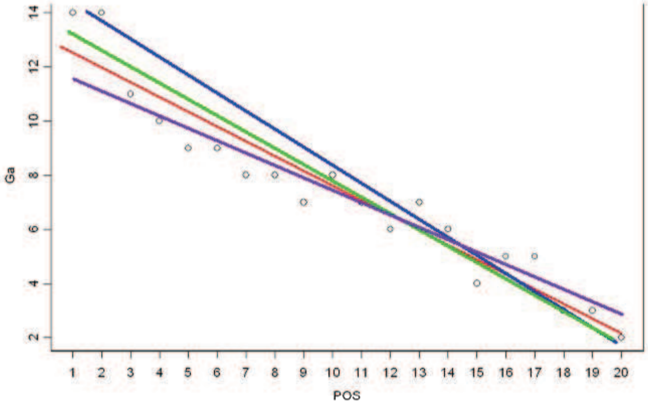
Tendremos en cuenta:

a) tomar una distancia promedio de la recta a todos los puntos
b) mover la recta hasta que esta distancia promedio sea la menor posible.

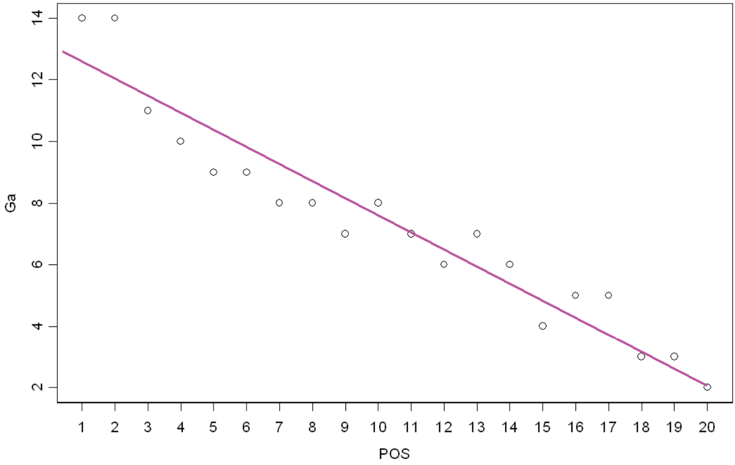
Si tenemos$(\mathbf{x}_i, y_i)$, $1 \leq i \leq n$, y queremos predecir y a partir de x usando una recta, podríamos definir el error cometido en cada punto como la distancia vertical del punto a la recta.



d2
Ga
POS
2
4
6
8
10
12
14
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20



Ga
POS
14
12
10
8
6
4
2
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20



14
12
10
8
6
4
2
Ga
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
POS

Supongamos que tenemos un modelo que depende de p parámetros. Sean $(\mathbf{x}_i, y_i)$ tales que

$$y_i = f(\mathbf{x}_i, \beta_1 \ldots \beta_p) + \varepsilon_i$$

$E(\varepsilon_i) = 0$, $V(\varepsilon_i) = \sigma^2$, ε_i son independientes y la función f es conocida salvo por los parámetros $\beta_1 \ldots \beta_p$.

Estimamos $\beta_1 \ldots \beta_p$ minimizando la suma de cuadrados residual

$\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_p)$ es el estimador de mínimos cuadrados si minimiza

$$\sum_{i=1}^{n} (y_i - f(\mathbf{x}_i, \beta_1 \ldots \beta_p))^2$$

En el caso de la regresión simple en el que $f(x, \beta_1, \beta_2) = \beta_1 + \beta_2\, x$, minimizaremos:

$$\frac{1}{n} \sum_{i=1}^{n} \left[y_i - (\beta_1 + \beta_2 x_i)\right]^2 .$$

Esta suma se llama la suma de cuadrados residual del error para la recta y la recta resultante recta de cuadrados mínimos.

Dado el vector $\mathbf{b} \in \Re^p$, el vector de residuos es

$$\mathbf{Y} - \mathbf{X}\mathbf{b}\,.$$

El estimador de mínimos cuadrados de $\beta_1 \dots \beta_p$ minimiza

$$\sum_{i=1}^{n} (y_i - b_1 x_{i1} - \cdots - b_p x_{ip})^2 = \|\mathbf{Y} - \mathbf{X}\mathbf{b}\|^2\,,$$

donde $\|\mathbf{u}\|^2 = \mathbf{u}'\mathbf{u} = \sum_{i=1}^{n} u_i^2$.

Llamemos

$$\mathcal{S}(\mathbf{b}) = \|\mathbf{Y} - \mathbf{X}\mathbf{b}\|^2 = (\mathbf{Y} - \mathbf{X}\mathbf{b})'(\mathbf{Y} - \mathbf{X}\mathbf{b})$$

Un conjunto de funciones de $\mathbf{Y}$, $\widehat{\beta}_1 = \widehat{\beta}_1(\mathbf{Y})$, $\widehat{\beta}_2 = \widehat{\beta}_2(\mathbf{Y}), \dots$
$\widehat{\beta}_p = \widehat{b}_p(\mathbf{Y})$ que minimiza $\mathcal{S}(\mathbf{b})$ es el estimador de mínimos cuadrados de $\boldsymbol{\beta}$ (LS).

Veremos que el LS siempre existe, aunque no siempre es único.

Ecuaciones Normales

Derivando e igualando a 0 obtenemos las **ecuaciones normales**.

Los estimadores de mínimos cuadrados $\widehat{\beta}_1, \ldots, \widehat{\beta}_p$ cumplen:

$$\frac{\partial \mathcal{S}(\mathbf{b})}{\partial b_k} = -2 \sum_{i=1}^{n} (Y_i - \sum_{j=1}^{p} x_{ij} b_j) x_{ik} = 0$$

Ecuaciones Normales

Derivando e igualando a 0 obtenemos las **ecuaciones normales**.

Los estimadores de mínimos cuadrados $\widehat{\beta}_1, \ldots, \widehat{\beta}_p$ cumplen:

$$\frac{\partial \mathcal{S}(\mathbf{b})}{\partial b_k} = -2\sum_{i=1}^{n}(Y_i - \sum_{j=1}^{p} x_{ij} b_j) x_{ik} = 0$$

Por lo tanto, para $1 \leq k \leq p$

$$\sum_{i=1}^{n} y_i x_{ik} = \sum_{i=1}^{n}\sum_{j=1}^{p} x_{ij} x_{ik} b_j$$

reordenando la suma

$$\sum_{i=1}^{n} y_i x_{ik} = \sum_{j=1}^{p} b_j \sum_{i=1}^{n} x_{ij} x_{ik}$$

Ecuaciones Normales

Como

$$\sum_{i=1}^{n} y_i x_{ik} = \sum_{j=1}^{p} b_j \sum_{i=1}^{n} x_{ij} x_{ik} \; 1 \leq k \leq p$$

cuando el modelo tiene intercept, y lo escribimos como antes en términos de $\beta_0, \ldots, \beta_{p-1}$, los estimadores $\widehat{\beta}_i$ cumplen

$$\begin{aligned} n\widehat{\beta}_0 + \widehat{\beta}_1 \sum_{i=1}^{n} x_{i1} + \cdots + \widehat{\beta}_{p-1} \sum_{i=1}^{n} x_{ip-1} &= \sum_{i=1}^{n} y_i \\ \widehat{\beta}_0 \sum_{i=1}^{n} x_{ik} + \widehat{\beta}_1 \sum_{i=1}^{n} x_{i1} x_{ik} + \cdots + \widehat{\beta}_{p-1} \sum_{i=1}^{n} x_{ip-1} x_{ik} &= \sum_{i=1}^{n} y_i x_{ik} \; k = 1, \ldots, p-1 \end{aligned}$$

Ecuaciones Normales

Estas p ecuaciones pueden escribirse como

$$\mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}\,,$$

que se conocen como ecuaciones normales.

Si $\mathbf{X}'\mathbf{X}$ es no singular, la solución es única y resulta

$$\boxed{\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}\,.}$$

Ejemplo

En el caso de regresión simple tendríamos

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ x_1 & x_2 & x_3 & \dots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ . & . \\ . & . \\ 1 & x_n \end{pmatrix}$$

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix}$$

El sistema sería

$$\begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix}$$

Ejemplo

La inversa resulta

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n\sum_{i=1}^n x_i^2 - n^2\overline{x}^2} \left(\begin{array}{cc} \displaystyle\sum_{i=1}^n x_i^2 & -\displaystyle\sum_{i=1}^n x_i \\ -\displaystyle\sum_{i=1}^n x_i & n \end{array} \right)$$

y además

$$\mathbf{X}'\mathbf{Y} = \left(\begin{array}{c} \displaystyle\sum_{i=1}^n y_i \\ \displaystyle\sum_{i=1}^n x_i y_i \end{array} \right)$$

y por lo tanto

$$\widehat{\boldsymbol{\beta}} = \left(\begin{array}{c} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{array} \right) = \frac{1}{n\displaystyle\sum_{i=1}^n (x_i - \overline{x})^2} \left(\begin{array}{c} (\displaystyle\sum_{i=1}^n y_i)(\displaystyle\sum_{i=1}^n x_i^2) - (\displaystyle\sum_{i=1}^n x_i)(\displaystyle\sum_{i=1}^n x_i y_i) \\ n\displaystyle\sum_{i=1}^n x_i y_i - (\displaystyle\sum_{i=1}^n y_i)(\displaystyle\sum_{i=1}^n x_i) \end{array} \right)$$

Ejemplo

Entonces

$$\boxed{\widehat{\beta}_0 = \overline{y} - \overline{x}\widehat{\beta}_1}$$

y por otro lado

$$\boxed{\widehat{\beta}_1 = \frac{\displaystyle\sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y}}{\displaystyle\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} = \frac{\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^2}}$$

Interpretación Geométrica

Asumamos que $\mathbf{X}$ tiene rango completo.

Nuestro modelo plantea

$$\begin{aligned} \Omega: \quad E(\mathbf{Y}) &= \mathbf{X}\boldsymbol{\beta} \\ \boldsymbol{\Sigma}_{\mathbf{Y}} &= \sigma^2\mathbf{I} \end{aligned}$$

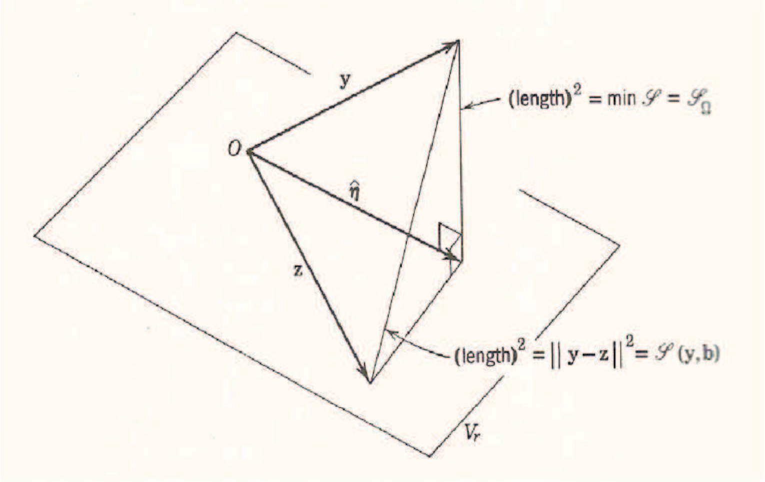
Luego, si

$$\boldsymbol{\eta} = E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$$

llamando a $\mathbf{x}^i$ la i–ésima columna de $\mathbf{X}$, Tenemos

$$\boldsymbol{\eta} = \beta_1\mathbf{x}^1 + \beta_2\mathbf{x}^2 + \cdots + \beta_p\mathbf{x}^p$$

Es decir que $\boldsymbol{\eta} \in \mathcal{V}_p =$ subespacio generado por las p columnas de $\mathbf{X}$: $\mathbf{x}^1, \ldots, \mathbf{x}^p$ y p es rg$(\mathbf{X})$.


y
$(\text{length})^2 = \min \mathscr{S} = \mathscr{S}_\Omega$
O
$\hat{\eta}$
z
$(\text{length})^2 = \|\mathbf{y}-\mathbf{z}\|^2 = \mathscr{S}(\mathbf{y},\mathbf{b})$
V_r

Entonces

$$\min_b \mathcal{S}(\mathbf{b}) = \min_b \|\mathbf{Y} - \mathbf{X}\mathbf{b}\|^2 = \min_{\mathbf{z} \in V_p} \|\mathbf{Y} - \mathbf{z}\|^2$$

Sabemos que se alcanza en $\widehat{\boldsymbol{\beta}} = b_1\mathbf{x}^1 + b_2\mathbf{x}^2 + \cdots + b_p\mathbf{x}^p$ la proyección ortogonal de Y sobre V_p, que sabemos que siempre existe y es única, aunque los b_i pueden no serlo.

En términos de las ecuaciones normales tenemos que:

$$\begin{aligned} \mathbf{X}'\mathbf{X}\mathbf{b} &= \mathbf{X}'\mathbf{Y} \\ \mathbf{X}'\widehat{\boldsymbol{\beta}} &= \mathbf{X}'\mathbf{Y} \end{aligned}$$

Caso en que $rg(\mathbf{X}) = p$

En este caso existe la inversa de $\mathbf{X}'\mathbf{X}$, pues $\text{rg}(\mathbf{X}'\mathbf{X}) = \text{rg}(\mathbf{X}) = p$.

De las ecuaciones normales queda:

$$\begin{aligned} \mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} &= \mathbf{X}'\mathbf{Y} \\ \\ \widehat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \end{aligned}$$

entonces

$$\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{P}\mathbf{Y} = \widehat{\mathbf{Y}}$$

En consecuencia el vector de residuos es:

$$\begin{aligned} \mathbf{r} &= \mathbf{Y} - \widehat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}} \\ &= \mathbf{Y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ &= \mathbf{Y} - \mathbf{P}\mathbf{Y} \\ &= (\mathbf{I} - \mathbf{P})\mathbf{Y} \end{aligned}$$

donde $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \in \Re^{n\times n}$.

Propiedades de $\mathbf{P}$

$\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \in \Re^{n \times n}$ es la matriz de **proyección sobre el espacio generado por las columnas de $\mathbf{X}$**. Suele llamarse a esta matriz $\mathbf{P}$ por ser de proyección $\mathbf{P}$ o $\mathbf{H}$ de hat matrix.

Matriz simétrica e idempotente, es decir: $\mathbf{P} = \mathbf{P}' = \mathbf{P}^2$. $\mathbf{I} - \mathbf{P}$ también es simétrica es idempotente, es decir también es una matriz de proyección y proyecta sobre el ortogonal de V_r.

Lema:

i) $\mathbf{P}$ y $\mathbf{I} - \mathbf{P}$ son simétricas e idempotentes
ii) $\text{rg}(\mathbf{P}) = \text{tr}(\mathbf{P}) = p$ y $\text{rg}(\mathbf{I} - \mathbf{P}) = \text{tr}(\mathbf{I} - \mathbf{P}) = n - p$
iii) $(\mathbf{I} - \mathbf{P})\mathbf{X} = \mathbf{0}$

Suma de Cuadrados

Tenemos que

$$\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 \quad = \quad \|\mathbf{Y} - \mathbf{P}\mathbf{Y}\|^2$$

Notemos que obtenemos el Teorema de Pitágoras. En efecto,

$$\begin{array}{rcl}
\|\mathbf{Y} - \widehat{\mathbf{Y}}\|^2 = \|\mathbf{Y} - \mathbf{P}\mathbf{Y}\|^2 & = & \|(\mathbf{I} - \mathbf{P})\mathbf{Y}\|^2 \\
& = & \mathbf{Y}'(\mathbf{I} - \mathbf{P})'(\mathbf{I} - \mathbf{P})\mathbf{Y} = \mathbf{Y}'(\mathbf{I} - \mathbf{P})\mathbf{Y} \\
& = & \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{P}\mathbf{Y} \\
& = & \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{P}'\mathbf{P}\mathbf{Y} \\
& = & \|\mathbf{Y}\|^2 - \|\mathbf{P}\mathbf{Y}\|^2 \\
& = & \|\mathbf{Y}\|^2 - \|\widehat{\mathbf{Y}}\|^2 \\
\Rightarrow & & \|\mathbf{Y}\|^2 = \|\mathbf{Y} - \widehat{\mathbf{Y}}\|^2 + \|\widehat{\mathbf{Y}}\|^2
\end{array}$$

Propiedades del Estimador de Mínimos Cuadrados

Usando la notación matricial escribimos el modelo como

$$\begin{aligned} \Omega: \quad \mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ E(\boldsymbol{\epsilon}) &= 0 \\ \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} &= \sigma^2 \mathbf{I} \end{aligned}$$

Lema: Si se cumple el modelo Ω, tenemos que

- $\widehat{\boldsymbol{\beta}}$ es un estimador insesgado de $\boldsymbol{\beta}$, es decir $E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$.
- $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}} = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$

Estimación de σ^2

Las varianzas de los estimadores dependen del diseño y σ^2, que es desconocida.

Dado que $\sigma^2 = E(\epsilon^2)$, parece natural estimarla mediante el promedio de los cuadrados de los residuos. El vector de residuos es

$$\begin{aligned}\mathbf{r} &= \mathbf{Y} - \widehat{\mathbf{Y}} \\ &= \mathbf{Y} - \mathbf{P}\mathbf{Y}\,,\end{aligned}$$

Bajo el modelo Ω, tenemos que

$$\boxed{s^2 = \frac{\|\mathbf{Y} - \widehat{\mathbf{Y}}\|^2}{n-p} = \frac{\|\mathbf{Y} - \mathbf{P}\mathbf{Y}\|^2}{n-p} = \frac{\mathbf{Y}'(\mathbf{I} - \mathbf{P})\mathbf{Y}}{n-p}}$$

es un estimador insesgado de σ^2.

Lema Auxiliar: Sea $\mathbf{x}$ un vector aleatorio n–dimensional y sea $\mathbf{A} \in \Re^{n\times n}$ una matriz simétrica. Si $E(\mathbf{x}) = \boldsymbol{\mu}$ y su matriz de covarianza es $\boldsymbol{\Sigma}_{\mathbf{x}}$ entonces

$$E(\mathbf{x}'\mathbf{A}\mathbf{x}) = tr(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$$

Errores Normales

Si los errores además de ser independientes con media 0 y varianza σ^2, son normales, es decir

$$\epsilon_i \sim N(0, \sigma^2) \, i = 1, \ldots, n$$

el modelo Ω puede formularse de la siguiente forma:

$$\Omega \quad : \qquad \mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$$

En este caso puede demostrarse que el estimador de máxima verosimilitud de $\boldsymbol{\beta}$ coincide con el de mínimos cuadrados.

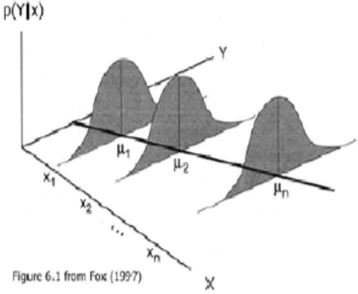


Figure 6.1 from Fox (1997)

En el caso de rango completo, es decir cuando $rg(\mathbf{X}) = p$, obtenemos el siguiente resultado:

Teorema: Supongamos que $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, rg$(\mathbf{X}) = p$, $\boldsymbol{\beta} \in \Re^p$. Entonces,

i) $\widehat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$

ii) $\dfrac{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{X}'\mathbf{X})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sigma^2} \sim \chi^2_p$

iii) $\widehat{\boldsymbol{\beta}}$ y $\dfrac{(n-p)s^2}{\sigma^2}$ son independientes

iv) $\dfrac{(n-p)s^2}{\sigma^2} \sim \chi^2_{n-p}$

Estos resultados nos permiten deducir intervalos de confianza o tests para cada uno de los coeficientes del modelo lineal:

Como $\widehat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$, entonces

$$\widehat{\beta}_i = \mathbf{e}_i'\widehat{\boldsymbol{\beta}} \sim N(\beta_i, \sigma^2\mathbf{e}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{e}_i)$$

Si llamamos $\Sigma_{\widehat{\boldsymbol{\beta}}} = \sigma^2\mathbf{D}$

$$\boxed{\widehat{\beta}_i \sim N(\beta_i, \sigma^2 d_{ii})}$$

siendo d_{ii} el i-ésimo elemento diagonal de $\mathbf{D}$.

Si para un i fijo quisiéramos testear

$$\boxed{H_o : \beta_i = 0 \quad \text{vs.} \quad H_1 : \beta_i \neq 0}$$

tenemos que bajo H_o

$$\frac{\widehat{\beta}_i}{s\sqrt{d_{ii}}} \sim t_{n-p}$$

Por lo tanto, rechazaremos H_o con nivel α si

$$\left| \frac{\widehat{\beta}_i}{s\sqrt{d_{ii}}} \right| > t_{n-p,\frac{\alpha}{2}}$$