Modelo Lineal

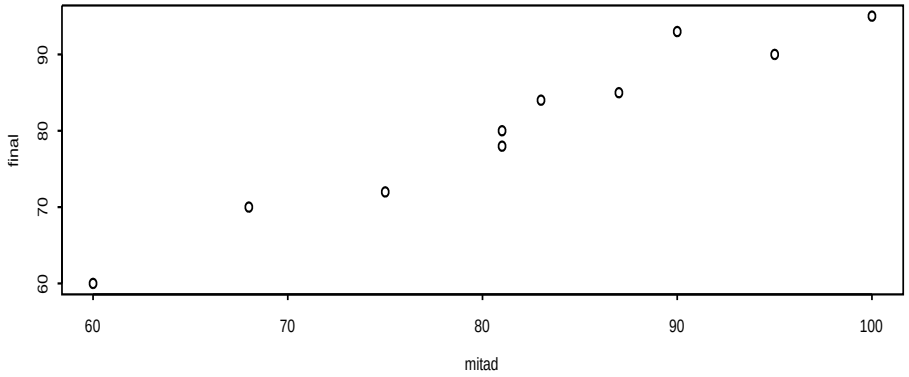
En regresión lineal se modela la relación entre una variable dependiente Y y otras variables p $X_1, \ldots, X_p$. Esta metodología es ampliamente usada en problemas de economía, de la industria y de ciencias en general. Por ejemplo:

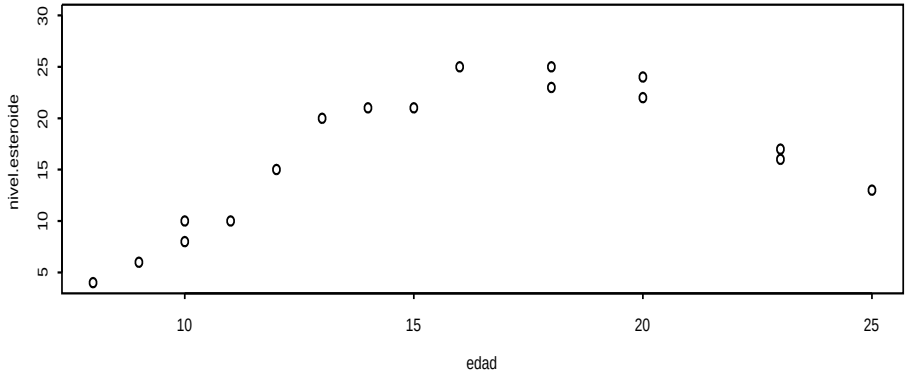
- en mujeres de 8 a 25 años de edad se desea relacionar la edad y la cantidad de esteroides presentes en plasma.
- dadas las evaluaciones de mitad y de fin de año de alumnos que participan en un estudio de rendimiento, se quiere relacionar la performance de los alumnos en los dos exámenes. El objetivo es poder predecir en situaciones similares cómo le irá a un alumno en la evaluación final a partir de lo que se observa en la evaluación de mitad de curso.

- un ingeniero podría estar interesado en la relación entre cantidad de óxido que se forma en un metal calcinado en un horno y la temperatura de horneado y el tiempo expuesto a dichas temperaturas.

En los dos primeros ejemplos podríamos tener gráficos como los siguientes:

Ejemplos


final
60
70
80
90
mitad
60
70
80
90
100


nivel.esteroide
5
10
15
20
25
30
edad
10
15
20
25

En los dos primeros ejemplos consideramos sólo dos variables, mientras que en el tercero hay 3 variables involucradas.

En general tendremos:

- y: *variable dependiente*.
- x: *variables independientes* (o *predictoras*, *regresoras* o *covariables*).

Buscaremos un modelo que exprese a la variable dependiente en términos de las variables independientes.

Cuando hablamos de un modelo nos referimos a una expresión matemática que describa en algún sentido el comportamiento de la variable de interés en función de las demás variables, es decir, las covariables.

En general, identificaremos con la letra Y (y) a la variable dependiente. El modelo pretende describir cómo el comportamiento de $E(Y)$ varía bajo condiciones cambiantes.

En nuestro caso, supondremos, al menos en un principio, que $V(Y)$ no es afectada por estas condiciones cambiantes.

Bajo el supuesto de que *otras variables* aportan información sobre la variable Y, éstas variables son incorporadas al modelo como variables independientes. Identificaremos con $\mathbf{X} = (X_1, \ldots, X_p)'$ ($\mathbf{x} = (x_1, \ldots, x_p)'$) a las variables independientes. Estas podrían ser variables aleatorias o constantes conocidas. En general, trabajaremos bajo este último caso y lo extenderemos al caso de variables aleatorias.

Una forma general de plantear esto es expresando a la media de la distribución de Y como una $g(\mathbf{x})$:

$$E(Y|\mathbf{X} = \mathbf{x}) = g(\mathbf{x}) \quad \text{para } \mathbf{x} \in D\,,$$

o también como

$$Y = g(X_1, \ldots, X_p) + \varepsilon\,,$$

donde en general la función g no es conocida y $E(\varepsilon) = 0$.

Los modelos de este tipo se llaman modelos de regresión. Las posibles funciones de regresión g pertenecen a una clase $\mathcal{G}$ tan grande que es frecuente que se simplifique el problema suponiendo cierta forma o ciertas propiedades de la función de regresión g.

Una forma de simplificar el problema suponiendo que la familia $\mathcal{G}$ puede expresarse en función de un número finito de constantes desconocidas, a estimar, llamadas *parámetros*, que controlan el comportamiento del modelo. En este sentido diremos que el modelo de regresión es paramétrico.

Se dirá que el modelo de regresión es no parámétrico si la familia $\mathcal{G}$ no puede expresarse en un número finito de parámetros.

Algunos ejemplos de modelos paramétricos y no paramétricos cuando hay dos variables independientes X_1 y X_2.

Modelos paramétricos

(i) $Y = \theta_1 X_1 + \theta_2 X_2 + \theta_3 + \varepsilon$

(ii) $Y = \theta_1 e^{\theta_2 X_1} + \theta_3 e^{\theta_4 X_2} + \varepsilon$

(iii) $Y = \theta_1 X_1^{\theta_2} X_2^{\theta_3} + \varepsilon$

(iv) $Y = \theta_1 \log X_1 + \theta_2 \log X_2 + \theta_3 X_1^3 + \theta_4 \text{ sen } X_2 + \varepsilon$

Modelos no paramétricos

(i) $Y = g(X_1, X_2) + \varepsilon$ donde $g(X_1, X_2)$ es una función continua.

(ii) $Y = g(X_1, X_2) + \varepsilon$ donde $g(X_1, X_2)$ es una función continua y derivable.

(iii) $Y = g(X_1, X_2) + \varepsilon$ donde $g(X_1, X_2)$ es monótona creciente en X_1 y X_2.

Uno de los modelos más sencillos es el **modelo lineal**, en el que los parámetros entran como simples coeficientes de las variables independientes o de funciones de éstas.

Es el caso de:

(i) $Y = \theta_1 X_1 + \theta_2 X_2 + \theta_3 + \varepsilon$

(iv) $Y = \theta_1 \log X_1 + \theta_2 \log X_2 + \theta_3 X_1^3 + \theta_4 \text{ sen } X_2 + \varepsilon$

En todos estos ejemplos $g(x)$ es **lineal** en los **parámetros**. No es el caso, por ejemplo, de $g(x) = \beta_o e^{-\beta_1 x}$, ya que no es lineal como función de los parámetros.

En situaciones más complejas Y depende de un conjunto de p variables $(x_1, \ldots, x_p)$, por lo tanto tendremos

$$g(x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_{p-1} x_{ip-1} \, .$$

Eventualmente , las $x_i's$ podrían ser funciones de otras variables, tales como: $x_1 = senz, x_2 = \log w, x_3 = zw$, etc.

Algunos ejemplos sencillos son:

$$\begin{aligned} g(x) &= \beta_o + \beta_1 x + \beta_2 x^2 \\ g(x) &= \beta_o + \beta_1 x \\ g(x) &= \beta_o + \beta_1 \log x \end{aligned}$$

También podríamos introducir variables explicativas que sean categóricas como las dummies que sólo toman los valores 0 y 1 y que sirven, como ya veremos, para indicar las distintas categorías de una variable categórica. Este caso es de especial interés pues permite tratar en el marco del modelo lineal el problema de comparar la media de más de dos poblaciones, que se conoce como Análisis de la Varianza.

Una vez "seleccionado" el modelo, nos interesará:

- Estimar los parámetros desconocidos: β_j y σ
- Testear hipótesis del tipo

$$H_o : \beta_j = 0 \qquad \text{o} \qquad H_o : c'\beta = \delta$$

- Intervalos de confianza para los parámetros (combinaciones lineales).
- Predicción
- Chequeo de supuestos
- Identificación de datos atípicos. Uso de métdos robustos.
- Medidas de ajuste

Por último, veremos

- Criterios para la selección de modelos.

Enfoque matricial

respuesta $y \longleftrightarrow p-1$ variables explicativas x_j

Por ahora, supondremos $x_j, 1 \leq j \leq p-1$ determinísticas.

Muestra $(x_{i1}, \ldots, x_{ip-1}, y_i)$, $1 \leq i \leq n$ que cumplen el modelo:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \ldots + \beta_{p-1} x_{ip-1} + \epsilon_i \quad i = 1, \ldots, n \\ & E(\varepsilon_i) = 0 \\ & V(\varepsilon_i) = \sigma^2 \\ & cov(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j \end{aligned}$$

donde, $\beta_0, \beta_1, \ldots, \beta_{p-1}$ son p parámetros desconocidos a estimar.

Este modelo tiene *intercept* u *ordenada al origen*, eventualmente podríamos saber que es 0, en cuyo caso plantearíamos

$$y_i = \beta_1 x_{i1} + \ldots + \beta_{p-1} x_{ip-1} + \epsilon_i \quad i = 1, \ldots, n$$

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p-1} \\ 1 & x_{21} & x_{22} & \dots & x_{2p-1} \\ & \dots & & \dots & \\ & \dots & & \dots & \\ 1 & x_{n1} & x_{n2} & \dots & x_{np-1} \end{pmatrix}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ \beta_{p-1} \end{pmatrix} \qquad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ . \\ . \\ \epsilon_n \end{pmatrix}$$

$$\Downarrow$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

La matriz $\mathbf{X} \in \Re^{n \times p}$ recibe el nombre de matriz de regresión o de diseño.

En general, se elige de tal forma que tenga rango máximo, es decir $rg(\mathbf{X}) = p$, sin embargo esto no siempre es posible, como en el caso de algunos diseños tratados en análisis de la varianza (ANOVA).

La teoría que veremos no necesita que la primera columna sea de 1's, es decir que el modelo tenga intercept, por lo tanto estudiaremos el caso general.

Propiedades de vectores y matrices aleatorias

Dada una matriz $\mathbf{V}$ $(r \times s)$ de variables aleatorias conjuntamente distribuidas $\{V_{ij}\}$ con esperanza finita, definimos la matriz o vector de esperanzas como:

$$\{E(\mathbf{V})\}_{ij} = E(V_{ij})$$

En nuestro caso, esto nos permite decir que el vector de errores es tal que

$$E(\boldsymbol{\epsilon}) = \mathbf{0}$$

y que

$$E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = E\begin{pmatrix} \epsilon_1\epsilon_1 & \epsilon_1\epsilon_2 & \dots & \epsilon_1\epsilon_n \\ \epsilon_2\epsilon_1 & \epsilon_2\epsilon_2 & \dots & \epsilon_2\epsilon_n \\ \dots & & \dots & \\ \dots & & \dots & \\ \epsilon_n\epsilon_1 & \epsilon_n\epsilon_2 & \dots & \epsilon_n\epsilon_n \end{pmatrix} = \sigma^2\mathbf{I}$$

Lema: Sean $\mathbf{A} \in \Re^{q\times r}$, $\mathbf{B} \in \Re^{s\times t}$ y $\mathbf{C} \in \Re^{q\times t}$ matrices constantes y $\mathbf{V}$ una matriz aleatoria de dimensión $r \times s$, entonces:

$$E(\mathbf{AVB} + \mathbf{C}) = \mathbf{A}E(\mathbf{V})\mathbf{B} + \mathbf{C}.$$

Matriz de Covarianza

Sea $\mathbf{v} = (v_1, \ldots, v_n)'$ un vector aleatorio de variables con $E(v_i) = \mu_i$ y varianza finita. Definimos la matriz de covarianza de $\mathbf{v}$ como:

$$\{\Sigma_{\mathbf{v}}\}_{ij} = Cov(\mathbf{v}_i, \mathbf{v}_j) = E[(v_i - \mu_i)(v_j - \mu_j)])$$

que podemos escribir como:

$$\Sigma_{\mathbf{v}} = \{E[(\mathbf{v} - \boldsymbol{\mu})(\mathbf{v} - \boldsymbol{\mu})'])\}$$

donde $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$.

En este sentido, como $E(\boldsymbol{\epsilon}) = \mathbf{0}$, entonces hemos visto que

$$\Sigma_{\boldsymbol{\epsilon}} = E[(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \sigma^2\mathbf{I}$$

Usaremos frecuentemente el siguiente

Lema: Sean $\mathbf{A} \in \Re^{m \times n}$, una matriz constante y $\mathbf{v}$ un vector aleatorio n–dimensional con matriz de covarianza $\Sigma_{\mathbf{v}}$. Si $\mathbf{w} = \mathbf{A}\mathbf{v}$, entonces:

$$\Sigma_{\mathbf{w}} = \mathbf{A}\Sigma_{\mathbf{v}}\mathbf{A}' .$$

El modelo que presentamos más arriba puede escribirse como:

$$\Omega : \mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\epsilon} \quad E(\boldsymbol{\epsilon}) = \mathbf{0} \quad \Sigma_{\boldsymbol{\epsilon}} = \sigma^2\mathbf{I}$$

o equivalentemente

$$\Omega : E(\mathbf{Y}) = \mathbf{X}\beta \quad \Sigma_{\mathbf{Y}} = \sigma^2\mathbf{I}$$

¿Cómo estimamos los parámetros?

Mínimos Cuadrados

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

$\Rightarrow$ solución de compromiso: acercar la recta a unos puntos la alejará de otros. 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 lo menor posible.

Gráficos 0

Si tenemos(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.

Sean (x_i, y_i) tales que

$$y_i = g(x_i, \beta_1 \dots \beta_p) + \varepsilon_i$$

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

Estimamos $\beta_1 \dots \beta_p$ minimizando *la suma de cuadrados residual*, o sea $\widehat{\beta} = (\widehat{\beta}_1, \dots, \widehat{\beta}_p)$ es el estimador de mínimos cuadrados si minimiza

$$\sum_{i=1}^{n} (y_i - g(x_i, \beta_1 \dots \beta_p))^2$$

Si $g(x, \beta_0, beta_1) = \beta_0 + \beta_1 x$, minimizaremos:

$$\frac{1}{n} \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2 .$$

Esta medida promedio se llama *la suma de cuadrados residual del error para la recta*. Fue inicialmente propuesta por Gauss. La recta de regresión así definida produce la menor suma de cuadrados residual para el error de predecir y a partir de x y por esta razón se la suele llamar *recta de mínimos cuadrados*.

Cinsideremos para cada vector $\mathbf{b} \in \Re^p$ el vector de residuos

$$\mathbf{Y} - \mathbf{Xb}\,.$$

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

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

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

Llamemos

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

Definimos un conjunto de funciones de $\mathbf{Y}$ $\widehat{\beta}_1 = \widehat{\beta}_1(\mathbf{Y})$, $\widehat{\beta}_2 = \widehat{\beta}_2(\mathbf{Y}), \ldots$ $\widehat{\beta}_p = \widehat{\beta}_p(\mathbf{Y})$ que minimice $\mathcal{S}(\mathbf{b})$ como el estimador de mínimos cuadrados de β (LS).

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

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 $1 \leq k \leq p$

$$\begin{aligned} \sum_{i=1}^{n} Y_i x_{ik} &= \sum_{i=1}^{n} \sum_{j=1}^{p} x_{ij} x_{ik} b_j \\ \sum_{i=1}^{n} Y_i x_{ik} &= \sum_{j=1}^{p} b_j \sum_{i=1}^{n} x_{ij} x_{ik} \end{aligned}$$

Si el modelo tiene intercept, los $\widehat{\beta}_i$ cumplen

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

Estas p ecuaciones pueden escribirse como

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

que se conocen como ecuaciones normales.

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

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

Ejemplo: En el caso de regresión simple tendríamos

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

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

El sistema sería

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

Tenemos que la inversa resulta

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

y además

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

y por lo tanto

$$\widehat{\beta} = \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{pmatrix} = \frac{1}{n \sum_{i=1}^{n} (x_i - \overline{x})^2} \begin{pmatrix} (\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} x_i y_i) \\ n \sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} x_i) \end{pmatrix}$$

entonces

$$b_0 = \overline{y} - \overline{x} b_1$$

y por otro lado

$$b_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

Interpretación Geométrica

En nuestro modelo tenemos que

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

Luego, si

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

si $\mathbf{x}_i$ es la i–ésima columna de $\mathbf{X}$ entonces

$$\boldsymbol{\eta} = \beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \ldots + \beta_p\mathbf{x}_p$$

es decir que $\boldsymbol{\eta} \in \mathcal{V}_r$= subespacio generado por las p columnas de $\mathbf{X}$: $\mathbf{x}_1, \ldots, \mathbf{x}_p$ y r es $rg(\mathbf{X})$.

Entonces

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

Gráfico 1

y sabemos que se alcanza en $\widehat{\boldsymbol{\eta}} = b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + \ldots + b_p\mathbf{x}_p$ la proyección ortogonal de Y sobre V_r, que sabemos que siempre existe y es única, aunque los b_i pueden no serlo.

En términos de las ecuaciones normales tenemos que:

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}$$

$$\mathbf{X}'\widehat{\boldsymbol{\eta}} = \mathbf{X}'\mathbf{Y}$$

Dados $\{b_1, \ldots, b_p\}$ funciones de $\mathbf{Y}$ serán un conjunto de estimadores de mínimos cuadrados (EMC) si y sólo si $\mathbf{X}'\mathbf{b} = \widehat{\boldsymbol{\eta}}$, es decir atisfacen las ecuaciones normales.

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

En este caso existe la inversa de $\mathbf{X}'\mathbf{X}$, pues $rg(\mathbf{X}'\mathbf{X}) = rg(\mathbf{X}) = p$.
De las ecuaciones normales queda:

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

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

entonces

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

En consecuencia el vector de residuos es:

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

donde $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \in \Re^{n\times n}$ es la matriz de **proyección**

sobre el espacio generado por las columnas de X. Suele llamarse a $\mathbf{P}$ o **H matriz hat** (hat matrix).

Propiedades de P

matriz simétrica e idempotente, es decir: $\mathbf{P} = \mathbf{P}' = \mathbf{P}^2$. $\mathbf{I}-\mathbf{P}$ también es simétrica es idempotente.

Suma de Cuadrados

Notemos que obtenemos elTeorema de Pitágoras:

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

$$\begin{aligned}
\|\mathbf{Y} - \mathbf{PY}\|^2 &= \\
&= \|(\mathbf{I} - \mathbf{P})\mathbf{Y}\|^2 \\
&= \mathbf{Y}'(\mathbf{I} - \mathbf{P})'(\mathbf{I} - \mathbf{P})\mathbf{Y} \\
&= \mathbf{Y}'(\mathbf{I} - \mathbf{P})\mathbf{Y} \\
&= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{PY} \\
&= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{P}'\mathbf{PY} \\
&= \|\mathbf{Y}\|^2 - \|\mathbf{PY}\|^2 \\
&= \|\mathbf{Y}\|^2 - \|\widehat{\mathbf{Y}}\|^2
\end{aligned}$$

Caso en que $rg(\mathbf{X}) = p$
Propiedades del Estimador de Mínimos Cuadrados

Usando la notación matricial podemos escribir el modelo como

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

Propiedades Si se cumple el modelo Ω, tenemos que

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

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

Bajo el modelo Ω

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

tenemos que

- $E(\widehat{\mathbf{Y}}) = \mathbf{X}\boldsymbol{\beta}$
- $\Sigma_{\widehat{\mathbf{Y}}} = \sigma^2\mathbf{P}$
- $E(\mathbf{r}) = \mathbf{0}$
- $\Sigma_{\mathbf{r}} = \sigma^2(\mathbf{I} - \mathbf{P})$

Si llamamos p_{ij} a los elementos de $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ tenemos que

$$p_{ij} = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_j$$

donde $\mathbf{x}_i$ representa la i–ésima fila de $\mathbf{X}$.

Luego:

$$\begin{aligned} Var(\widehat{y}_i) &= \sigma^2 p_{ii} \\ Var(r_i) &= \sigma^2(1 - p_{ii}) \\ Cov(r_i, r_j) &= -\sigma^2 p_{ij}\,, \end{aligned}$$

por lo tanto

$$Corr(r_i, r_j) = -\frac{p_{ij}}{\sqrt{1 - p_{ii}}\,\sqrt{1 - p_{jj}}}$$

Algunas propiedades de la matriz $\mathbf{P}$:
Lema:

i) $\mathbf{P}$ y $\mathbf{I}-\mathbf{P}$ son simétricas e idempotentes

ii) $rg(\mathbf{I}-\mathbf{P}) = tr(\mathbf{I}-\mathbf{P}) = n-p$ y $rg(\mathbf{P}) = tr(\mathbf{P}) = p$

iii) $(\mathbf{I}-\mathbf{P})\mathbf{X} = \mathbf{0}$

Proposición: Dados $1 \leq i, j \leq n$ tenemos que

i) $0 \leq p_{ii} \leq 1$

ii) $-\frac{1}{2} \leq p_{ij} \leq \frac{1}{2}$ si $i \neq j$

Como ya vimos $Var(\widehat{y}_i) = \sigma^2 p_{ii}$, una consecuencia inmediata es que

$$Var(\widehat{y}_i) \leq Var(y_i) = \sigma^2 .$$

Una propiedad interesante es que $\mathbf{P}$ es invariante por transformaciones lineales no singulares de la forma $\mathbf{X} \rightarrow \mathbf{XA}$, donde $\mathbf{A} \in \Re^{p \times p}$ y $rg(\mathbf{A}) = p$. Este tipo de transformaciones es útil, por ejemplo, si queremos realizar un cambio de unidades en las covariables.

Más aún, como consecuencia de esta propiedad, si el modelo $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ contiene un término constante, entonces los predichos $\widehat{\mathbf{Y}}$ y los residuos $\mathbf{r}$ son invariantes por cambio de escala y de posición de $\mathbf{X}$, mientras que si no contiene intercept, entonces es invariante por cambios de escala de $\mathbf{X}$.

Respecto a las propiedades de invariancia, podemos ver que si

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

para $\mathbf{A} \in \Re^{p \times p}$ no singular, $\lambda \in \Re$ y $\gamma \in \Re^p$, entonces

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}(\mathbf{XA}, \mathbf{Y}) &= \mathbf{A}^{-1}\widehat{\boldsymbol{\beta}} && \text{Invariancia por transformciones afines} \\
\widehat{\boldsymbol{\beta}}(\mathbf{X}, \lambda\mathbf{Y}) &= \lambda\widehat{\boldsymbol{\beta}} && \text{Invariancia por cambios de escala} \\
\widehat{\boldsymbol{\beta}}(\mathbf{X}, \mathbf{Y} + \mathbf{X}\gamma) &= \widehat{\boldsymbol{\beta}} + \gamma && \text{Invariancia por cambios de regresión}
\end{aligned}$$

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

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

es un estimador insesgado de σ^2.

Lema: 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 $\Sigma_{\mathbf{x}}$ entonces

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

Respecto del diseño

Covariables aleatorias

Si las covariables son aleatorias suponemos que tenemos los vectores $(\mathbf{x}_i, y_i)$ i.i.d. que satisfacen el modelo

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + \epsilon_i$$

donde los ϵ_i son i.i.d., con $E(\epsilon_i) = 0$ y $Var(\epsilon_i) = \sigma^2$ e independientes de $\mathbf{x}_i$

El análogo de suponer que $\mathbf{X}$ tiene rango completo es asumir que la distribución de $\mathbf{x}$ no está concentrada en ningún hiperplano, es decir $P(\mathbf{a}'\mathbf{x} = 0) < 1 \quad \forall \mathbf{a} \neq \mathbf{0}$. Esta condición se cumple por ejemplo si $\mathbf{x}$ tiene densidad.

En este caso, $\widehat{\boldsymbol{\beta}}$ está bien definido y las fórmulas que vimos para esperanza y varianza de $\widehat{\boldsymbol{\beta}}$ son válidas condicionalmente:

$$E(\widehat{\boldsymbol{\beta}}|\mathbf{X} = \mathbf{x}) = \boldsymbol{\beta} \qquad Var(\widehat{\boldsymbol{\beta}}|\mathbf{X} = \mathbf{x}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

Se puede ver que si $\mathbf{V_x} = E(\mathbf{xx'})$ existe, entonces la distribución asintótica de $\widehat{\boldsymbol{\beta}}$ será

$$N_p\left(\boldsymbol{\beta}, \frac{\sigma^2 \mathbf{V}_\mathbf{x}^{-1}}{n}\right)$$

Cuando el modelo tiene intercept, podemos escribirlo como:

$$y_i = \beta_0 + \mathbf{x}_i' \boldsymbol{\beta}_1 + \epsilon_i$$

donde β_0 es la intercept y $\boldsymbol{\beta}_1$ es el vector de pendientes. En este caso resulta

$$\sigma^2 \mathbf{V}_\mathbf{x}^{-1} = \sigma^2 \begin{pmatrix} 1 + \boldsymbol{\mu}_\mathbf{x}' \Sigma_\mathbf{x}^{-1} \boldsymbol{\mu}_\mathbf{x} & -\boldsymbol{\mu}_\mathbf{x}' \Sigma_\mathbf{x}^{-1} \\ -\Sigma_\mathbf{x}^{-1} \boldsymbol{\mu}_\mathbf{x} & \Sigma_\mathbf{x}^{-1} \end{pmatrix}$$

con $\boldsymbol{\mu}_\mathbf{x} = E(\mathbf{x})$ y $\Sigma_\mathbf{x}$ matriz de covarianza de $\mathbf{x}$.

Estructura Ortogonal en la matriz de Diseño

Supongamos que podemos dividir a la matriz $\mathbf{X}$ en k conjuntos de columnas ortogonales:$\mathbf{X}_1, \ldots, \mathbf{X}_k$, de manera que

$$\mathbf{X} = [\mathbf{X}_1 \ldots \mathbf{X}_k]$$

La correspondiente división en los parámetros daría

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k)'$$

Luego podemos escribir:

$$E(\mathbf{Y}) = \mathbf{X}_1\boldsymbol{\beta}_1 + \ldots + \mathbf{X}_k\boldsymbol{\beta}_k$$

Como las comulmnas de $\mathbf{X}_i$ son ortogonales a las de $\mathbf{X}_j$ si $i \neq j$, tenemos que $\mathbf{X}_i'\mathbf{X}_j = 0$, luego

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{pmatrix} \mathbf{X}_1'\mathbf{X}_1 & 0 & \ldots & 0 \\ 0 & \mathbf{X}_2'\mathbf{X}_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \mathbf{X}_k'\mathbf{X}_k \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}_1'\mathbf{Y} \\ \mathbf{X}_2'\mathbf{Y} \\ \ldots \\ \mathbf{X}_k'\mathbf{Y} \end{pmatrix}$$

entonces

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{pmatrix} (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y} \\ (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{Y} \\ \cdots \\ (\mathbf{X}_k'\mathbf{X}_k)^{-1}\mathbf{X}_k'\mathbf{Y} \end{pmatrix} = \begin{pmatrix} \widehat{\boldsymbol{\beta}}_1 \\ \widehat{\boldsymbol{\beta}}_2 \\ \cdots \\ \widehat{\boldsymbol{\beta}}_k \end{pmatrix}$$

en consecuencia el estimador de $\boldsymbol{\beta}_i$ no cambiará si alguno de los otros $\boldsymbol{\beta}_j$ se iguala a 0, es decir si se remueve del modelo.

¿ Cómo resulta la suma de cuadrados?

$$\mathbf{Y}'\mathbf{Y} - \widehat{\mathbf{Y}}'\widehat{\mathbf{Y}} \; = \; \mathbf{Y}'\mathbf{Y} - \widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} = \mathbf{Y}'\mathbf{Y} - \sum_{j=1}^{k} \widehat{\boldsymbol{\beta}}_j'\mathbf{X}_j'\mathbf{Y}$$

Por lo tanto si en el modelo ponemos algún $\boldsymbol{\beta}_i = 0$, el único cambio

en la suma de cuadrados es que el término de $\widehat{\boldsymbol{\beta}}_i' \mathbf{X}_i' \mathbf{Y}$ no aparece:

$$\mathbf{Y}'\mathbf{Y} - \sum_{\substack{j=1 \\ j \neq i}}^{k} \widehat{\boldsymbol{\beta}}_j' \mathbf{X}_j' \mathbf{Y}$$

En el caso más sencillo, cada $\mathbf{X}_i$ consta de una única columna y resulta:

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

y la suma de cuadrados queda

$$\mathbf{Y}'\mathbf{Y} - \sum_{j=1}^{k} \widehat{\boldsymbol{\beta}}_j' \mathbf{X}_j' \mathbf{Y} = \mathbf{Y}'\mathbf{Y} - \sum_{j=1}^{k} \widehat{\boldsymbol{\beta}}_j^2 \mathbf{X}_j' \mathbf{X}_j$$

Teorema de Gauss–Markov

En muchas aplicaciones estamos más interesado en estimar funciones lineales de $\boldsymbol{\beta}$ que en estimar al mismo $\boldsymbol{\beta}$.

Estas funciones incluyen elvalor esperado de y en una futura observación $\mathbf{x}_o$, por ejemplo.

Si bien puede haber muchos estimadores de una función lineal $\mathbf{c}'\boldsymbol{\beta}$ o $\mathbf{C}\boldsymbol{\beta}$, estudiaremos los estimadores lineales, es decir funciones lineales de las observaciones $y_1, \ldots, y_n$.

Primero veremos cuando una función paramétrica es estimable.

Definición: Una función paramétrica ψ se dice que es una función lineal de los paramétros $\{\beta_1, \ldots, \beta_p\}$ si existen $\{c_1, \ldots, c_p\}$ constantes conocidas tal que

$$\psi = \mathbf{c}'\boldsymbol{\beta} = \sum_{j=1}^{p} c_j \beta_j$$

donde $\mathbf{c} = (c_1, \ldots, c_p)'$.

Definición: Decimos que una función paramétrica $\psi = \mathbf{c}'\boldsymbol{\beta}$ es estimable si tiene un estimador lineal (en $\mathbf{Y}$) insesgado, es decir si existe $\mathbf{a} \in \Re^n$ tal que

$$E(\mathbf{a}'\mathbf{Y}) = \psi = \mathbf{c}'\boldsymbol{\beta} \quad \forall \boldsymbol{\beta} \in \Re^p$$

¿Hay funciones que no son estimables? Veamos el siguiente resultado

Teorema: La función paramétrica $\psi = \mathbf{c}'\boldsymbol{\beta}$ es estimable si y sólo si $\mathbf{c}$ es una combinación lineal de las filas de $\mathbf{X}$, o sea si existe $\mathbf{a} \in \Re^n$ tal que

$$\mathbf{c}' = \mathbf{a}'\mathbf{X}$$

Veamos un ejemplo de una función paramétrica no estimable.

Supongamos que queremos comparar la respuesta media de dos tratamientos y un control y que para ello observamos

$$\begin{array}{lll} \text{T1:} & y_{11}, y_{12}, \ldots, y_{1k} & y_{1j} \sim N(\beta_1, \sigma^2) \\ \text{T2:} & y_{21}, y_{22}, \ldots, y_{2k} & y_{2j} \sim N(\beta_2, \sigma^2) \\ \text{Co:} & y_{31}, y_{32}, \ldots, y_{3k} & y_{3j} \sim N(\beta_3, \sigma^2) \end{array}$$

Suponemos igual cantidad de observaciones por tratamiento para simplificar la notación.

Podemos esciribir esto como

$$y_{ij} = \beta_i + \epsilon_{ij}$$

Podríamos escribir esto como un modelo lineal:

$$\mathbf{Y} = \begin{pmatrix} y_{11} \\ y_{12} \\ \dots \\ y_{1k} \\ y_{21} \\ y_{22} \\ \dots \\ y_{2k} \\ y_{31} \\ y_{32} \\ \dots \\ y_{3k} \end{pmatrix}; \mathbf{X} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ . & . & . \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ . & . & . \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ . & . & . \\ 0 & 0 & 1 \end{pmatrix}; \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

Por ejemplo, T1, T2 y el control podrían ser distintas dosis de una droga de manera que T1 es menor que la dosis del control y T2 mayor

que la dosis control. Tendría sentido preguntarse si

$$\beta_3 = \frac{\beta_1 + \beta_2}{2}$$

lo que implicaría cierta linealidad en el efecto medio. En ese caso nos interesaría saber si

$$\left(-\tfrac{1}{2}, -\tfrac{1}{2}, 1 \right) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = 0$$

Otra manera de escribir el modeo sería

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

donde μ es el efecto general y α_i es el efecto del tratamiento i. En ese caso tendríamos

$$\mathbf{Y} = \begin{pmatrix} y_{11} \\ y_{12} \\ \dots \\ y_{1k} \\ y_{21} \\ y_{22} \\ \dots \\ y_{2k} \\ y_{31} \\ y_{32} \\ \dots \\ y_{3k} \end{pmatrix}; \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ . & . & . & . \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ . & . & . & . \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ . & . & . & . \\ 1 & 0 & 0 & 1 \end{pmatrix}; \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$$

¿Son todas las funciones estimables en este modelo?

Consideremos

$$\alpha_1 = \left(0, 1, 0, 0 \right) \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$$

Veremos que α_1 no es estimable.

Lema: Supongamos que vale el modelo Ω. Sean $\psi = \mathbf{c}'\boldsymbol{\beta}$ una función estimable y $\mathcal{V}_r$ el espacio generado por las columnas de $\mathbf{X}$ ($r = rg(\mathbf{X}) \leq p$). Luego, existe un único estimador lineal insesgado de ψ, digamos $\mathbf{a}^{*'}\mathbf{Y}$ con $\mathbf{a}^* \in \mathcal{V}_r$. Más aún, si $\mathbf{a}'\mathbf{Y}$ es un estimador insesgado de ψ, $\mathbf{a}^*$ es la proyección ortogonal de $\mathbf{a}$ sobre $\mathcal{V}_r$.

Teorema de Gauss–Markov:
Supongamos que vale el modelo $\Omega : E(\mathbf{Y}) = \mathbf{X}\beta \quad \Sigma_{\mathbf{Y}} = \sigma^2\mathbf{I}$.
Toda función estimable $\psi = \mathbf{c}'\boldsymbol{\beta}$ tiene un único estimador $\widehat{\psi}$ lineal insesgado de mínima varianza (BLUE). Este estimador $\widehat{\psi}$ se puede obtener reemplazando a $\boldsymbol{\beta}$ en $\mathbf{c}'\boldsymbol{\beta}$ por $\widehat{\boldsymbol{\beta}}$, el estimador de mínimos cuadrados.

Definición: Dada una función estimable ψ su único estimador lineal insesgado de mínima varianza $\widehat{\psi}$, cuya existencia y cálculo están dados por el Teorema de Gauss–Markov, es el estimador de mínimos cuadrados de ψ.

Tenemos el siguiente resultado:

Corolario: Si $\{\psi_1, \ldots, \psi_q\}$ son q funciones estimables toda combinación lineal $\Psi = \Sigma_{i=1}^{q} h_i \psi_i$ es estimable y su estimador de mínimos cuadrado está dado por $\Sigma_{i=1}^{q} h_i \widehat{\psi}_i$.

¿ Qué ocurre cuando el $rg(\mathbf{X}) < p$

Si $rg(\mathbf{X}) = r < p$ tenemos que $\widehat{\beta}_1, \ldots, \widehat{\beta}_p$ no son únicos. Esta misma indeterminación afecta a los parámetros $\beta_1, \ldots, \beta_p$, en el sentido de que distintos conjuntos $b_1, \ldots, b_p$ darían origen al mismo η y por lo tanto al mismo modelo

$$\mathbf{Y} = \eta + \boldsymbol{\epsilon} = E(\mathbf{Y}) + \boldsymbol{\epsilon}\,.$$

Sin embargo, tal como vimos si $\mathbf{c}'\boldsymbol{\beta}$ es una función estimable tendrá el mismo valor independientente del $\boldsymbol{\beta}$ que usemos, en tanto

$$\mathbf{c}'\boldsymbol{\beta} = \mathbf{a}'\mathbf{X}\boldsymbol{\beta} = \mathbf{a}'\eta$$

que depende de η que es único.

¿ Cómo podemos eliminar esta indeterminación?

a) Considerar un problema reducido con sólo r parámetros

Podríamos considerar r columnas l.i. de $\mathbf{X}$ que generen a $\mathcal{V}_r$ y mantener en el modelo sólo aquellos β_j asociados.

Así tendríamos una nueva matriz de diseño $\mathbf{X}_1 \in \Re^{n\times r}$ con rango máximo. En este caso tendríamos el modelo

$$\mathbf{Y} = \eta + \boldsymbol{\epsilon} \quad \text{con } \eta \in \mathcal{V}_r$$

El estimador sería

$$\boldsymbol{\alpha} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}$$

y la matriz de proyección corresponiente $\mathbf{P} = \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$.

Si asumimos, s.p.g., que las columnas elegidas son las primeras r, tendríamos que

$$\mathbf{X} = [\mathbf{X}_1\mathbf{X}_2]$$

donde $\mathbf{X}_2 \in \Re^{n\times(p-r)}$ y además $\mathbf{X}_2 = \mathbf{X}_1\mathbf{B}$. Por lo tanto

$$\mathbf{X} = \mathbf{X}_1[\mathbf{I}_r \quad \mathbf{B}] = \mathbf{K}\mathbf{L}$$

con $\mathbf{K} \in \Re^{n\times r}$, $\quad \mathbf{L} \in \Re^{n\times p}$ y $rg(\mathbf{L}) = r$.

Por lo tanto el modelo original se obtiene como:

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{KL}\boldsymbol{\beta} = \mathbf{K}\boldsymbol{\alpha}$$

b) Considerar condiciones de contorno adecuadas para los β_j's y sus estimadores

Así podríamos pedir que $\boldsymbol{\beta}_{r+1} = \ldots = \boldsymbol{\beta}_p = 0$ y en este caso obtendíamos el mismo que en la situación a) (suponiendo que las r primeras son las columnas l.i.).

Sin embargo, en otras situaciones, como en el ANOVA, es frecuente que se impongan otras restricciones lineales de manera de obtener la unicidad.

Consideremos el caso en que imponemos $t \geq p - r$ restricciones lineales a los $\boldsymbol{\beta}_j$, es decir

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{0} \quad \text{con } \mathbf{H} \in \Re^{\mathbf{t}\times\mathbf{p}}$$

Queremos encontrar dentro del conjunto de soluciones de $\mathbf{X}\boldsymbol{\beta} = \eta$ una sola que cumpla $\mathbf{H}\boldsymbol{\beta} = \mathbf{0}$, es decir buscamos $\tilde{\boldsymbol{\beta}}$ que sea única solución de

$$\begin{aligned} \mathbf{X}\tilde{\boldsymbol{\beta}} &= \mathbf{X}\boldsymbol{\beta} \quad (= \eta) \\ \mathbf{H}\tilde{\boldsymbol{\beta}} &= \mathbf{0} \end{aligned}$$

De manera que las primeras ecuaciones establecen que encontraremos una solución del sistema que nos interesa y las segundas que esta solución es única.

Lo que queremos es que

- toda función estimable del nuevo sistema lo sea en el viejo problema,
- un único conjunto de estimadores de mínimos cuadrados que satisfaga las condiciones de contorno.

El siguiente teorema nos dice como elegir $\mathbf{H}$ para cumplir con este objetivo:

Teorema: Sean $\mathbf{X} \in \Re^{n \times p}$ y $\mathbf{H} \in \Re^{t \times p}$ con $rg(\mathbf{X}) = r$, $p > r$ y $t \geq p - r$. Consideremos $\mathcal{V}_{\mathbf{X}}$ el espacio generado por las columnas de $\mathbf{X}$. El sistema

$$\begin{aligned} \mathbf{Xb} &= \mathbf{z} \\ \mathbf{Hb} &= \mathbf{0} \end{aligned}$$

tiene solución única $\mathbf{b}$ para todo $\mathbf{z} \in \mathcal{V}_{\mathbf{X}}$ si y sólo si se cumplen las siguientes dos condiciones:

i) si $rg(\mathbf{G}) = rg \begin{pmatrix} \mathbf{X} \\ \mathbf{H} \end{pmatrix} = p$

ii) ninguna combinación lineal de las filas de $\mathbf{H}$ es combinación lineal de las de $\mathbf{X}$, excepto el 0.

Observaciones:

- las condiciones i) y ii) son conjuntamente equivalentes a **1)** $rg(\mathbf{G}) = p$ y **2)** $rg(\mathbf{H}) = p - r$
- la condición ii) del Teorema nos dice que si $\mathbf{h}_i$ es la iésima fila de $\mathbf{H}$, entonces no existe $\mathbf{a}$ tal que $\mathbf{h}_i = \mathbf{a}'\mathbf{X}$, por lo tanto las $\mathbf{h}_i'\boldsymbol{\beta}$ no es una función estimable de los parámetros.
- Si se cumplen las condiciones i) y ii) del Teorema, entonces los $\tilde{\beta}_j$ son funciones estimables.
- dada una función estimable ψ, para cualquier $\mathbf{H}$ que elijamos en las condiciones del Teorema anterior, $Var(\widehat{\psi})$ es la misma.

c) Computar una inversa generalizada de X′X: $(\mathbf{X}'\mathbf{X})^-$

En este caso tendríamos que $(\mathbf{X}'\mathbf{X})^-\mathbf{X}\mathbf{Y}$ es solución de las ecuaciones normales, por lo tanto otra forma de solucionar nuestro problema. En realidad puede verse que la opción b) y c) quedan ligadas a través del siguiente resultado:

Proposición: Sea $\mathbf{G} = \begin{pmatrix} \mathbf{X} \\ \mathbf{H} \end{pmatrix}$ una matriz que satisface las condiciones del Teorema anterior. Luego $(\mathbf{G}'\mathbf{G})^{-1}$ es una inversa generalizada de $\mathbf{X}'\mathbf{X}$, por lo tanto:

$$(\mathbf{X}'\mathbf{X})(\mathbf{G}'\mathbf{G})^{-1}(\mathbf{X}'\mathbf{X}) = \mathbf{X}'\mathbf{X}$$

En efecto, $\forall \mathbf{Y}$:

$$\begin{aligned} (\mathbf{G}'\mathbf{G})(\mathbf{G}'\mathbf{G})^{-1}\mathbf{H}'\mathbf{Y} &= \mathbf{H}'\mathbf{Y} \\ (\mathbf{X}'\mathbf{X} + \mathbf{H}'\mathbf{H})(\mathbf{G}'\mathbf{G})^{-1}\mathbf{H}'\mathbf{Y} &= \mathbf{H}'\mathbf{Y} \\ \mathbf{X}'\mathbf{X}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{H}'\mathbf{Y} &= \mathbf{H}'(\mathbf{I} - \mathbf{H}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{H}')\mathbf{Y} \end{aligned}$$

entonces como $\mathbf{X}'\boldsymbol{\alpha} = \mathbf{H}'\boldsymbol{\beta}$ tenemos que

$$\mathbf{X}'\mathbf{X}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{H}'\mathbf{Y} = \mathbf{0}$$

luego

$$\mathbf{X}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{H}'\mathbf{Y} \in \mathcal{V}_r^{\perp}$$

y al mismo tiempo

$$\mathbf{X}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{H}'\mathbf{Y} \in \mathcal{V}_r$$

por lo tanto

$$\mathbf{X}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{H}' = \mathbf{0}$$

Finalmente:

$$(\mathbf{X}'\mathbf{X})(\mathbf{G}'\mathbf{G})^{-1}(\mathbf{X}'\mathbf{X}) = (\mathbf{X}'\mathbf{X} + \mathbf{H}'\mathbf{H})(\mathbf{G}'\mathbf{G})^{-1}(\mathbf{X}'\mathbf{X}) = \mathbf{X}'\mathbf{X}$$

Mínimos Cuadrados Pesados y Mínimos Cuadrados Generalizados

¿ Qué ocurre cuando $\Sigma_{\mathbf{Y}} = \sigma^2\mathbf{V}$ donde $\mathbf{V} \neq \mathbf{I}$?

Supongamos que $\mathbf{V} \in \Re^{n\times n}$ es una matriz definida positiva de constantes. Podemos entonces escribir: $\mathbf{V} = \mathbf{K}\mathbf{K}'$ con $\mathbf{K}$ una matriz invertible.

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ \mathbf{K}^{-1}\mathbf{Y} &= \mathbf{K}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{K}^{-1}\boldsymbol{\epsilon} \end{aligned}$$

donde $E(\mathbf{K}^{-1}\boldsymbol{\epsilon}) = \mathbf{0}$ y $\Sigma_{\mathbf{K}^{-1}\boldsymbol{\epsilon}} = \sigma^2\mathbf{I}$.

Por lo tanto tenemos un nuevo problema:

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\epsilon}}$$

Hallar el estimador de mínimos cuadrados en el problema transformado equivale a:

$$\begin{aligned}\min_{\mathbf{b}} \|\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{b}\|^2 &= \min_{\mathbf{b}} (\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{b})'(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{b}) \\ &= \min_{\mathbf{b}} (\mathbf{Y} - \mathbf{X}\mathbf{b})'\mathbf{K}^{-1'}\mathbf{K}^{-1}(\mathbf{Y} - \mathbf{X}\mathbf{b}) \\ &= \min_{\mathbf{b}} (\mathbf{Y} - \mathbf{X}\mathbf{b})'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\mathbf{b})\end{aligned}$$

Si $\mathbf{V}$ es una matriz diagonal decimos que tenemos un problema de Mínimos Cuadrados Pesados, mientras que si $\mathbf{V}$ es una matriz definida positiva cualquiera, es de Mínimos Cuadrados Generalizados.

Las ecuaciones normales quedan:

$$\begin{aligned}\tilde{\mathbf{X}}\tilde{\mathbf{X}}'\mathbf{b} &= \tilde{\mathbf{X}}\tilde{\mathbf{Y}} \\ \mathbf{X}'\mathbf{K}^{-1'}\mathbf{K}^{-1}\mathbf{X}\mathbf{b} &= \mathbf{X}'\mathbf{K}^{-1'}\mathbf{K}^{-1}\mathbf{Y} \\ \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\mathbf{b} &= \mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}\end{aligned}$$

Observemos que si $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$ tiene inversa, por lo tanto

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

y además

- $\tilde{\boldsymbol{\beta}}$ es un estimador insesgado de $\boldsymbol{\beta}$, es decir $E(\tilde{\boldsymbol{\beta}}) = \beta$.
- $\Sigma_{\tilde{\boldsymbol{\beta}}_e} = \sigma^2(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} = \sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$

Veamos un ejemplo.

Consideremos el caso sencillo de una regresión simple por el origen:

$$\mathbf{Y} = \mathbf{x}\beta + \boldsymbol{\epsilon}$$

donde $\mathbf{Y} = (y_1, \ldots, y_n)'$, $\mathbf{x} = (x_1, \ldots, x_n)'$ y $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)'$ con $E(\boldsymbol{\epsilon}) = \mathbf{0}$ y $\Sigma_{\boldsymbol{\epsilon}} = \sigma^2\mathbf{V} = \sigma^2 diag(w_1, \ldots, w_n)$ con $w_i > 0$.

Probaremos que

$$\tilde{\boldsymbol{\beta}} = \frac{\Sigma_{i=1}^n y_i x_i / w_i}{\Sigma_{i=1}^n x_i^2 / w_i}$$

y además

$$\Sigma_{\tilde{\boldsymbol{\beta}}} = \sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2/w_i}$$

Si $rg(\mathbf{X}) = p$ se puede probar fácilmente que el estimador $\tilde{\boldsymbol{\beta}}$ conserva las propiedades del estimador de mínimos cuadrados: dada una función lineal estimable $\mathbf{c}'\boldsymbol{\beta}$ tenemos que

- $\mathbf{c}'\tilde{\boldsymbol{\beta}}$ es el estimador lineal insesgado de $\mathbf{c}'\boldsymbol{\beta}$ de menor varianza.

Una pregunta muy natural sería:

¿ Hay situaciones en las que $\tilde{\boldsymbol{\beta}}$ y $\widehat{\boldsymbol{\beta}}$ coinciden?

El siguiente resultado nos da la respuesta

Teorema: Una condición necesaria y suficiente para que $\tilde{\boldsymbol{\beta}}$ y $\widehat{\boldsymbol{\beta}}$ coincidan es que $\mathcal{V}_{\mathbf{V}^{-1}\mathbf{X}} = \mathcal{V}_{\mathbf{X}}$.

Corolario: $\tilde{\boldsymbol{\beta}}$ y $\widehat{\boldsymbol{\beta}}$ coinciden $\Longleftrightarrow \mathcal{V}_{\mathbf{V}\mathbf{X}} = \mathcal{V}_{\mathbf{X}}$.

Corolario: Si tenemos un modelo de regresión simple por el origen, $\mathbf{Y} = \mathbf{x}\beta + \boldsymbol{\epsilon}$, entonces

$$\tilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} \quad \forall \mathbf{x} \Longleftrightarrow \mathbf{V} = \mathbf{c}\mathbf{I}_n$$

Forma Canónica del Modelo Ω

Dada una base ortonormal de $\mathcal{V}_r = \mathcal{V}_{\mathbf{X}}$, digamos $\{\boldsymbol{\alpha}_1, \ldots \boldsymbol{\alpha}_r\}$, sabemos que podemos extenderla a una base ortonormal de $\Re^n$: $\{\boldsymbol{\alpha}_1, \ldots \boldsymbol{\alpha}_n\}$.

Por lo tanto,

$$\mathbf{y} \in \Re^n : \mathbf{y} = \sum_{j=1}^{n} z_j \boldsymbol{\alpha}_j \, .$$

tenemos que

$$\boldsymbol{\alpha}_i' \mathbf{y} = \sum_{j=1}^{n} z_j \boldsymbol{\alpha}_i' \boldsymbol{\alpha}_j = z_i \boldsymbol{\alpha}_i' \boldsymbol{\alpha}_i = z_i \forall i = 1, \ldots, n$$

Por lo tanto, si la matriz $\mathbf{T}$ tiene filas $\boldsymbol{\alpha}_i'$ entonces

$$\mathbf{z} = \mathbf{T}\mathbf{y}$$

Observemos que

$$E(z_i) = \begin{cases} \boldsymbol{\alpha}_i' \eta = \xi_i & \text{si } 1 \leq i \leq r \\ 0 & \text{si } r+1 \leq i \leq n \end{cases}$$

$$\Sigma_{\mathbf{z}} = \mathbf{T} \Sigma_{\mathbf{y}} \mathbf{T}' = \sigma^2 \mathbf{I}$$

Por lo tanto ahora podemos reescribir a Ω como

$$\Omega: \quad E(z_i) = \begin{cases} \xi_i & \text{si } 1 \leq i \leq r \\ 0 & \text{si } r+1 \leq i \leq n \end{cases}$$

$$\Sigma_{\mathbf{z}} = \sigma^2 \mathbf{I}$$

donde ξ y σ^2 son parámetros desconocidos.

En términos de esta forma caónica es sencillo demostrar que

$$\boxed{s^2 = \frac{\|\mathbf{Y} - \widehat{\mathbf{Y}}\|^2}{n-r} = \frac{\|\mathbf{Y} - \mathbf{P}\mathbf{Y}\|^2}{n-r}}$$

es un estimador insesgado de σ^2. Sólo habíamos demostrado hasta ahora el caso de rango completo.

Distribución Normal Multivariada

Definición 1: Se dice que un vector $\mathbf{X}$, $k-$dimensional tiene distribución normal multivariada $N(\boldsymbol{\mu}, \mathbf{Q})$ donde $\boldsymbol{\mu}$ es un vector $k-$dimensional, Q una matriz de $k \times k$ definida positiva, si su densidad es de la forma

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^k |\mathbf{Q}|^{1/2}} \, e^{-1/2(\mathbf{x}-\boldsymbol{\mu})'\mathbf{Q}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

donde $|Q|$ indica determinante de Q.

Por ejemplo, si X_i son k variables aleatorias normales independientes con varianza σ_i y media μ_i; entonces el vector $\mathbf{X}' = (X_1, \ldots, X_k)$

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^k \, \Pi_{j=1}^k (\sigma_i^2)^{1/2}} \, e^{-1/2\Sigma_{i=1}^k (x_i-\mu_i)^2/\sigma_i^2}$$

Luego resulta que $\mathbf{X}$ es $N(\boldsymbol{\mu}, \mathbf{Q})$ donde $\boldsymbol{\mu}' = (\mu_1, \ldots, \mu_k)$ y

$$\mathbf{Q} = diag(\sigma_1^2, \ldots, \sigma_k^2) = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_k^2 \end{pmatrix}$$

Más aún, en el caso en que las X_i sean todas $N(0,1)$, $\mathbf{X}$ es $N(\underline{0}, \mathbf{I})$ donde $\underline{0}' = (0, \ldots, 0)$ y $\mathbf{I}$ es la matriz identidad de $k \times k$.

Recordemos el Teorema de Cambio de Variable:

Sean $\mathbf{x}$ *es un vector aleatorio con densidad* f *y* $\mathbf{y} = g(\mathbf{x})$, *tal que* $g^{-1}(\mathbf{y}) = \mathbf{x}$. *Supongamos que en un abierto* $\mathcal{G}$ *existen las derivadas parciales* $\dfrac{\partial x_i}{\partial y_j}$ *y sea* $J = det\left\{\dfrac{\partial x_i}{\partial y_j}\right\}$, *entonces*

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(g^{-1}(\mathbf{y}))|J|$$

Teorema N1: *Si* $\mathbf{X}$ *es un vector aleatorio* $k-$*dimensional con distribución* $N(\boldsymbol{\mu}, \mathbf{Q})$, $\mathbf{A}$ *es una matriz no singular de* $k \times k$ *y* $\mathbf{b}$ *un vector* $k-$*dimensional, entonces*

$$\mathbf{Y} = \mathbf{AX} + \mathbf{b} \quad es \quad N(\mathbf{A}\boldsymbol{\mu} + \mathbf{b},\ \mathbf{AQA}')$$

Teorema N2:

i) *Un vector aleatorio* $k-$*dimensional* $\mathbf{X}$ *es* $N(\boldsymbol{\mu}, \mathbf{Q})$ *si y sólo si* $\mathbf{X} = \mathbf{BY} + \boldsymbol{\mu}$, *donde* $\mathbf{Y}$ *es* $N(\mathbf{0_k}, \mathbf{I_k})$ *y* $\mathbf{B}$ *es una matriz de* $k \times k$ *no singular tal que* $\mathbf{BB'} = \mathbf{Q}$.

ii) *Si* $\mathbf{X}$ *es* $N(\boldsymbol{\mu}, \mathbf{Q})$ *entonces*

$$E(\mathbf{X}) = \boldsymbol{\mu} \quad y \quad \Sigma_{\mathbf{X}} = \mathbf{Q}$$

Teorema N3: *Sea* $\mathbf{X}$ *un vector aleatorio* $k-$*dimensional* $N(\boldsymbol{\mu}, \mathbf{Q})$ *y* $\mathbf{A}$ *una matriz de* $h \times k$ *con rango* h, *luego si* $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$ *entonces* $\mathbf{Y}$ *es* $N(A\boldsymbol{\mu} + \mathbf{b}, \mathbf{AQA'})$.

Teorema N4: *Sea* $\mathbf{X}' = (X_1, \ldots, X_k)$ *un vector* $k-$*dimensional con distribución normal multivariada, luego la distribución marginal de cualquier subconjunto de componentes tiene distribución normal multivariada. En particular cada componente es normal.*

Demostración: Sea $\mathbf{X}^* = (X_{k_1}, \ldots, X_{k_h})$, $k_1 < k_2 < \ldots < k_h$, luego se tiene que $\mathbf{X}^* = A\mathbf{X}$, donde A es la matriz de $h \times k$ dada por:

$$a_{ij} = \begin{cases} 1 & \text{si } j = k_i \\ 0 & \text{si } j \neq k_i \end{cases}$$

$1 \leq i \leq h$, $1 \leq j \leq k$. Es fácil ver que A es una matriz de rango h.

Teorema N5: *Si* $\mathbf{X}$ *es un vector* $k-$*dimensional con distribución* $N(\boldsymbol{\mu}, \mathbf{Q})$, *luego*

$$(\mathbf{X} - \boldsymbol{\mu})'\mathbf{Q}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim \chi^2_k$$

.

Demostración: Por lo ya visto, resulta que $\mathbf{X} = \mathbf{BY} + \boldsymbol{\mu}$ donde $\mathbf{Y}$ es $N(\mathbf{o}_k, \mathbf{I}_k)$

$$\mathbf{Y} = \mathbf{B}^{-1}(\mathbf{X} - \boldsymbol{\mu})$$

y además

$$\mathbf{BB}' = \mathbf{Q}$$

Luego

$$\mathbf{YY}' = (\mathbf{X} - \boldsymbol{\mu})'\mathbf{B}'^{-1}\mathbf{B}^{-1}(\mathbf{X} - \boldsymbol{\mu}) = (\mathbf{X} - \boldsymbol{\mu})'\mathbf{Q}^{-1}(\mathbf{X} - \boldsymbol{\mu})$$

Luego el teorema resulta del hecho que

$$\mathbf{Y}'\mathbf{Y} = \sum_{i=1}^{k} Y_i^2$$

tiene distribución χ^2_k, ya que las Y_i son variables aleatorias independientes con distribución $N(0, 1)$.

Teorema N6: *Si* $\mathbf{X}$ *es un vector* $k-$*dimensional con distribución* $N(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ *y* $\mathbf{P}$ *una matriz simétrica e idempotente de rango* r. *Luego,*

$$\frac{(\mathbf{X}-\boldsymbol{\mu})'\mathbf{P}(\mathbf{X}-\boldsymbol{\mu})}{\sigma^2} \sim \chi^2_r$$

.

Tests y Regiones de Confianza

Hasta ahora hemos trabajado sólo con las hipótesis Ω. Sin embargo para deducir tests y regiones de conianza con nivel exacto será necesario que hagamos un supuesto adicional: **normalidad conjunta de los errores**

Supondremos que las y_i's se distribuyen conjuntamente según una normal multivariada.

Podremos deducir:

- intervalos de confianza de nivel exacto para funciones paramétricas estimables
- tests de hipótesis de nivel exacto que involucren los parámetros
- conjuntos o regiones de confianza para la estimación simultánea de más de una función paramétrica estimable.

Nuestro nuevo modelo será:

$$\Omega: \quad \mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{Y}) \quad rg(\mathbf{X}) = r \quad \boldsymbol{\beta} \in \Re^p$$

Observemos que en este caso suponer que $\Sigma_{\mathbf{Y}} = \sigma^2\mathbf{I}$ es equivalente a asumir que las $y_i, 1 \leq i \leq n$ son independientes.

Ver gráfico

Teorema: *Supongamos que se tiene el modelo*

$$\Omega: \quad \mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{Y}) \quad rg(\mathbf{X}) = p \quad \boldsymbol{\beta} \in \Re^p \, .$$

Luego, $\widehat{\boldsymbol{\beta}}$ *y* s^2 *son estadísticos suficientes y completos y por lo tanto* $\widehat{\boldsymbol{\beta}}$ *y* s^2 *son estimadores IMVU de* $\boldsymbol{\beta}$ *y* σ^2*, respectivamente.*

Si nuestro modelo es

$$E(\mathbf{Y}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

nos podría interesar testear

$$H_o : \beta_j = 0 \text{ vs. } H_1 : \beta_j \neq 0$$
$$H_o : \beta_1 - \beta_6 = 0 \text{ vs. } H_1 : \beta_1 - \beta_6 \neq 0$$
$$H_o : \beta_1 = \beta_2 = \ldots = \beta_p = 0 \text{ vs. } H_1 : \text{existe } j : \beta_j \neq 0$$

Todas esta hipótesis son de la forma $\mathbf{c}'\boldsymbol{\beta} = 0$ o $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$.

Supongamos que tenemos q fucniones estimables: $\psi_1, \psi_2, \ldots, \psi_q$ donde:

$$\psi_i = \sum_{i=1}^{p} c_{ij}\beta_i \quad 1 \leq i \leq q$$

Sabemos que por ser estimables, aplicandoel Teorema de Gauss–Markov

$$\widehat{\psi}_i = \sum_{j=1}^{n} a^*_{ij} y_j \quad 1 \leq i \leq q$$

donde $\mathbf{a}^*_i \in \mathcal{V}_r \subset \Re^n$ de manera que

$$\begin{aligned} \Psi &= \mathbf{C}\boldsymbol{\beta} \qquad \mathbf{C} \in \Re^{q\times p} \\ \widehat{\Psi} &= \mathbf{A}^*\mathbf{Y} \qquad \mathbf{A}^* \in \Re^{q\times n} \end{aligned}$$

Más aún, sabemos que

$$\begin{aligned} \widehat{\Psi} &= \mathbf{C}\widehat{\boldsymbol{\beta}} \\ \Sigma_{\widehat{\Psi}} &= \sigma^2 \mathbf{A}^* \mathbf{A}^{*\prime} \end{aligned}$$

Estimamos a σ^2 por

$$s^2 = \frac{\|\mathbf{Y} - \widehat{\mathbf{Y}}\|^2}{n - r}$$

Bajo estas nuevas hipótesis obtenemos el siguiente resultado:

Teorema: Supongamos que se cumple Ω, es decir $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, $rg(\mathbf{X}) = r$, $\boldsymbol{\beta} \in \Re^p$ y que además que $\psi_1, \psi_2, \ldots, \psi_q$ son q funciones estimables l.i., de manera que $rg(\mathbf{C}) = q$. Entonces,

i) $\widehat{\Psi} \sim N_q(\Psi, \Sigma_{\widehat{\Psi}})$ (o lo que es igual $N_q(\Psi, \sigma^2\mathbf{A}^*\mathbf{A}^{*'})$)

ii) $\widehat{\Psi}$ y $\dfrac{s^2(n-r)}{\sigma^2}$ son independientes

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

En el caso de rango completo, es decir cuando $r = p$, obtenemos el siguiente resultado:

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

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

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

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

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

Si para un i fijo queremos testear

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

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

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

En el caso de regresión simple tendríamos

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad 1 \leq i \leq n, \quad \epsilon_i \sim N(0, \sigma^2)$$

Entonces:

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

y la inversa resulta

$$(\mathbf{X'X})^{-1} = \frac{1}{n\,\Sigma_{i=1}^{n} x_i^2 - n^2\overline{x}^2} \begin{pmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{pmatrix}$$

$$\widehat{\beta_0} = -\overline{x}b_1 + \overline{y}$$

y

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

Luego, el estadístico será

$$T = \left| \frac{\widehat{\beta_1}}{s\sqrt{d_{11}}} \right| = \left| \frac{\widehat{\beta_1}}{s/\sqrt{\Sigma_{i=1}^{n}(x_i - \overline{x})^2}} \right|$$

y rechazaremos H_o si $|T| > t_{n-2,\frac{\alpha}{2}}$

Veamos un ejemplo: Precio del papel.

```
> Ejemplo Precio del Papel
   precio  ganancia
        x         y
 1   1.83        28
 2   3.35        45
 3   0.64        12
 4   2.30        35
 5   2.39        45
 6   1.08        14
 7   2.92        39
 8   1.11        12
 9   2.57        43
10   1.22        23
```

```
> sal.lm
Coefficients:
 (Intercept)           x
    2.027775     14.20517

Degrees of freedom: 10 total; 8 residual
Residual standard error: 5.025083

> summary(sal.lm)

Call: lm(formula = y ~ x, x = T)
Residuals:
    Min         1Q     Median      3Q       Max
 -5.796   -4.222    0.1386    2.952     9.022
```

```
Coefficients:
              Value    Std. Error    t value    Pr(>|t|)
(Intercept)   2.0278     3.9383         0.5149     0.6206
          x  14.2052     1.8565         7.6516     0.0001

Residual standard error: 5.025 on 8 degrees of freedom
Multiple R-Squared: 0.8798
F-statistic: 58.55 on 1 and 8 degrees of freedom, the p-value is 0.

Correlation of Coefficients:
  (Intercept)
          x  -0.915

 X'X=
                (Intercept)       x
(Intercept)       10.00         19.4100
          x       19.41         45.0013
```

```
(X'X)^(-1)=

             (Intercept)          x
(Intercept)   0.6142273    -0.264929
          x  -0.2649290     0.136491

> matriz de covarianza de coeficientes

             (Intercept)          x
(Intercept)   15.510133    -6.689844
          x   -6.689844     3.446597
>
```

También podríamos interesarnos realizar in I. de C. para la esperanza de una nueva observación que cumpla el modelo en $\mathbf{x}_o = (\mathbf{x}_{o1}, \mathbf{x}_{o2}, \ldots, \mathbf{x}_{op})'$ en el modelo

$$y_i = \beta_1 x_1 + \beta_1 x_2 + \ldots + \beta_p x_p + \epsilon_i$$

donde $\epsilon_i \sim N(0, \sigma^2)$ independientes.

Como $E(y_o) = \mathbf{x}_o'\boldsymbol{\beta}$, podemos estimarlo por $E\widehat{(y_o)} = \mathbf{x}_o'\widehat{\boldsymbol{\beta}} = \widehat{y}_o$

Por lo tanto, de acuerdo con lo que hemos visto

$$\widehat{y}_o = \mathbf{x}_o'\widehat{\boldsymbol{\beta}} \sim N(\mathbf{x}_o'\boldsymbol{\beta}, \sigma^2\mathbf{x}_o'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_o)$$

y es independiente de

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

por lo tanto

$$T = \frac{\widehat{y}_o - \mathbf{x}_o'\boldsymbol{\beta}}{s\sqrt{\mathbf{x}_o'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_o}} \sim t_{n-p}$$

En consecuencia,

$$\widehat{y}_o \pm t_{n-2,\frac{\alpha}{2}}\ s\sqrt{\mathbf{x}_o'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_o}$$

es un intervalo de nivel exacto $1-\alpha$.

Asimismo, podríamos estar interesados en la predicción de y_o, una nueva observación que cumpla el modelo, y en un intervalo para ella, que llamaremos intervalo de predición.

Observemos que el predictor de y_o es $\widehat{y}_o = \mathbf{x}_o'\widehat{\boldsymbol{\beta}}$. En efecto, $E(\widehat{y}_o - y_o) = 0$. ¿Qué distribución tiene $\widehat{y}_o - y_o$?

Tenemos que

$$\begin{aligned} \widehat{y}_o &\sim N(\mathbf{x}_o'\boldsymbol{\beta}, \sigma^2\mathbf{x}_o'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_o) \\ y_o &\sim N(\mathbf{x}_o'\boldsymbol{\beta}, \sigma^2) \end{aligned}$$

y dado que y_o es independiente de las restantes y_i's con las que estimamos, entonces por la independencia entre estas dos normales queda que

$$\widehat{y}_o - y_o \sim N(0, \sigma^2(1 + \mathbf{x}_o'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_o))$$

Por lo tanto, el intervalo de predicción de nivel $1-\alpha$ estará dado por

$$\widehat{y}_o \pm t_{n-2,\frac{\alpha}{2}}\ s\sqrt{1 + \mathbf{x}_o'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_o}$$

Ejemplo Los siguientes son datos que corresponden a 10 porcentajes y_i de una sustancia que fueron medidos en experiencias de laboratorio y que se desean relacionar con la temperartura x_i a la que fueron realizados dichas experiencias.

i	x	y
1	100	45
2	110	52
3	120	54
4	130	63
5	140	62
6	150	68
7	160	75
8	170	76
9	180	92
10	190	88

La tabla con los estadísticos calculados es:

Coeficiente	Estimación	Error estandar	Valor de t
β_0	-4.47273	5.63433	-0.79
β_1	0.49636	0.03812	13.02
s	3.46213	g.l.=8	

Intervalos de Estimación y de Predicción

Ver lmintervalos.bmp

Tabla de Resultados

Ver Salida.bmp

- El valor estimado de $\widehat{\beta}_1 \simeq 0.5$, $\Rightarrow$ esperamos que el porcentaje aumente 0.5 por cada incremento de un grado en la temperatura.
- $s_{\widehat{\beta}_1} = 0.03812$
- Si testeamos $H_0 : \beta_1 = 0$ $t = \frac{0.49636}{0.038112} = 13.02$ y $t_{8,0.025} = 2.306004$

$\Rightarrow$ los datos nos dan evidencia suficiente como para concluir que la pendiente es no nula.

Observemos que en el gráfico la recta ajustada está encerrada entre 2 curvas interiores y 2 exteriores. Las externas corresponden al intervalo de predicción de nivel 0.95 y las internas a los intervalos de confianza de nivel 0.95 para la media.

Notemos que **el nivel de confianza 0.95 se aplica a cada punto y no es global**

Supongamos que queremos plantear un test de nivel α para

$$H_o : \mathbf{C}'\boldsymbol{\beta} = \boldsymbol{\delta} \text{ vs. } H_1 : \mathbf{C}'\boldsymbol{\beta} \neq \boldsymbol{\delta}$$

siendo $rg(\mathbf{C}) = q$, $\mathbf{C} \in \Re^{q \times p}$.

Sea $\Psi = \mathbf{C}'\boldsymbol{\beta}$. Sabemos que $\Psi \sim N_q(\Psi, \sigma^2 \mathbf{A}^* \mathbf{A}^{*'}) = N_q(\Psi, \sigma^2 \mathbf{B})$. Por lo tanto, tenemos que

$$(1): \quad Q = \frac{1}{q}(\Psi - \boldsymbol{\delta})' \mathbf{B}^{-1} (\Psi - \boldsymbol{\delta})$$

es independiente de

$$(2): \quad s^2 = \frac{\|\mathbf{Y} - \widehat{\mathbf{Y}}\|^2}{n - r}$$

Veremos que

$$E(Q) = \sigma^2 + \eta^2$$

y que $\eta^2 = 0$ sólo cuando H_o es cierta.

Bajo H_o ambos (1) y (2) son estimadores insesgados de σ^2, es decir que bajo H_o esperamos que

$$\frac{(1)}{(2)} \simeq 1,$$

pero si H_o no es cierta, esperamos que

$$\frac{(1)}{(2)} > 1.$$

Luego, el cociente $\dfrac{(\Psi - \boldsymbol{\delta})'\mathbf{B}^{-1}(\Psi - \boldsymbol{\delta})}{qs^2}$ nos dará una idea de la *veracidad de* H_O , de manera que rechazaremos H_o si el cociente es grande. ¿Cuán grande?

Bajo H_o

$$\frac{(\Psi - \boldsymbol{\delta})'\mathbf{B}^{-1}(\Psi - \boldsymbol{\delta})}{q\sigma^2} \sim \chi^2_{n-r}$$

independiente de

$$\frac{(n-r)s^2}{\sigma^2} \sim \chi^2_{n-r}$$

En consecuencia:

$$F = \frac{(\Psi - \boldsymbol{\delta})' \mathbf{B}^{-1} (\Psi - \boldsymbol{\delta})}{q s^2} \sim \mathcal{F}_{q,n-r}$$

Rechazaremos H_o si

$$F > \mathcal{F}_{q,n-r,\alpha}$$