Modelo Lineal

Graciela Boente[1]

[1]Universidad de Buenos Aires and CONICET, Argentina

Modelización y Predicción

Algunas de los métodos estadísticos más extendidos se ocupan de la modelización de datos y de la predicción.

Muchas de estas técnicas estadísticas se encuadran en lo que hoy se conoce como aprendizaje estadístico (AE).

Modelización y Predicción

Algunas de los métodos estadísticos más extendidos se ocupan de la modelización de datos y de la predicción.

Muchas de estas técnicas estadísticas se encuadran en lo que hoy se conoce como aprendizaje estadístico (AE).

El AE abarca una vasta cantidad de técnicas que ayudan a comprender los datos cuando se analizan varias variables al mismo tiempo, ya sea postulando modelos o encontrando relaciones entre las variables o estructuras que ayudan a su comprensión.

Modelización y Predicción

Algunas de los métodos estadísticos más extendidos se ocupan de la modelización de datos y de la predicción.

Muchas de estas técnicas estadísticas se encuadran en lo que hoy se conoce como aprendizaje estadístico (AE).

El AE abarca una vasta cantidad de técnicas que ayudan a comprender los datos cuando se analizan varias variables al mismo tiempo, ya sea postulando modelos o encontrando relaciones entre las variables o estructuras que ayudan a su comprensión.

Los métodos de AE pueden reunirse en dos grandes grupos:

- Aprendizaje Supervisado: Aquí una de las variables es identificada como una respuesta.

- Aprendizaje No Supervisado: todas las variables cumplen un rol análogo.

Aprendizaje Estadístico

Algunos ejemplos de aprendizaje estadístico:

Aprendizaje Estadístico

Algunos ejemplos de aprendizaje estadístico:

- Predecir si un paciente hospitalizado tendrá un segundo infarto de miocardio o no teniendo en cuenta mediciones clínicas, dietas y variables demográficas.
- Predecir los precios que tendrán en 6 meses las acciones de ciertas compañías a partir de mediciones del rendimiento de las compañías y datos macroeconómicos.
- Estimar la cantidad de glucosa en sangre que tendrá un individuo diabético a partir del espectro de adsorción infra-rojo de la sangre.
- Identificar los factores de riesgo de cáncer de próstata, usando mediciones clínicas y variables demográficas.

Vamos a considerar el último ejemplo:

En 97 pacientes que van a ser tratados con una prostatectomía radical se miden las siguientes variables:

- $x_1 =$ **lweight**: log del peso de la próstata
- $x_2 =$ **age**: edad
- $x_3 =$ **lbph**: log de la cantidad de hiperplasia prostática benigna
- $x_4 =$ **svi**: invasión seminal (si o no)
- $x_5 =$ **lcp**: logaritmo de la penetración capsular
- $x_6 =$ **gleason**: score de Gleason
- $x_7 =$ **pgg46**: porcentaje de scores de Gleason 4 or 5.
- $x_8 =$ **lpsa**: log de PSA

El objetivo es poder predecir el logaritmo del volumen del tumor ($y =$**lcavol**).

Diagramas de Dispersión

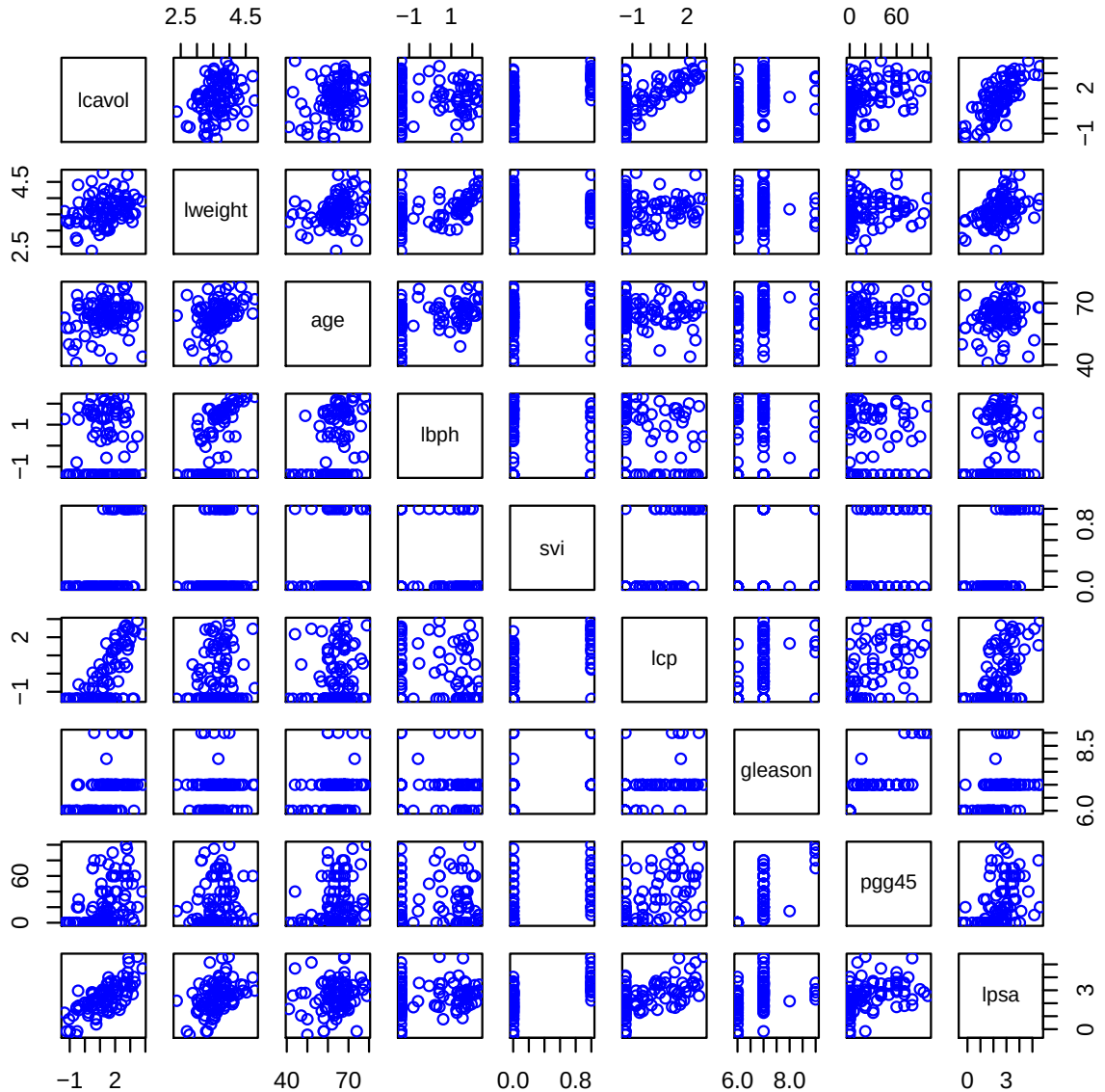
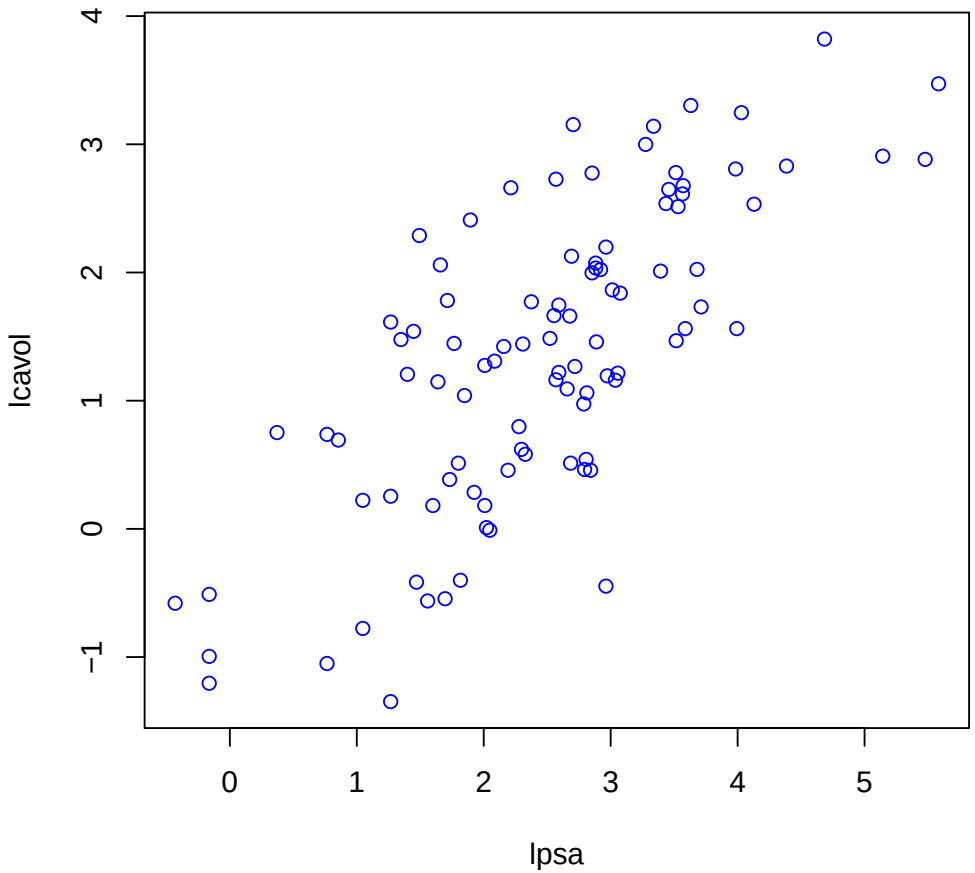


Diagrama de Dispersión



Modelos de regresión

Buscamos un modelo que exprese a la variable de respuesta en términos de las otras variables presentes (covariables).

Cuando hablamos de un modelo nos referimos a una expresión matemática

- que sea válida aproximadamente
- que describa en algún sentido el comportamiento de la variable de interés en función de las demás variables *predictoras*.

Modelos de regresión

En general, identificamos a

- la variable dependiente (o variable respuesta) con la letra y.
- las variables predictoras o covariables con la letra $\mathbf{x}$.

Modelos de regresión

En general, identificamos a

- la variable dependiente (o variable respuesta) con la letra y.
- las variables predictoras o covariables con la letra $\mathbf{x}$.

El modelo pretende describir cómo varía el comportamiento de $\mathbb{E}(y)$ bajo condiciones cambiantes del vector de p covariables $\mathbf{x}$.

Modelos de regresión

En general, identificamos a

- la variable dependiente (o variable respuesta) con la letra y.
- las variables predictoras o covariables con la letra $\mathbf{x}$.

El modelo pretende describir cómo varía el comportamiento de $\mathbb{E}(y)$ bajo condiciones cambiantes del vector de p covariables $\mathbf{x}$.

En un modelo de regresión se postularía:

$$y = \eta(x_1, x_2, x_3, \ldots, x_p) + \epsilon = \eta(\mathbf{x}) + \epsilon$$

Modelos de regresión

$$y = \eta(\mathbf{x}) + \epsilon$$

Las posibles funciones de regresión η 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 $\eta(\mathbf{x})$.

Modelos de regresión

$$y = \eta(\mathbf{x}) + \epsilon$$

Las posibles funciones de regresión η 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 $\eta(\mathbf{x})$.

Una forma de simplificar el problema es suponer 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.

$$\mathcal{G} = \{\eta(\mathbf{x}) = g(\mathbf{x}, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^s\}$$

En este caso, la función $g(\cdot, \boldsymbol{\theta})$ está completamente determinada por $\boldsymbol{\theta}$ y por eso, diremos que el **modelo de regresión es paramétrico**.

Modelos de regresión

Para simplificar, pensemos que tenemos dos covariables: x_1 y x_2.

Algunos ejemplos de modelos no paramétricos, semi-paramétricos y paramétricos cuando hay dos covariables son

Modelos de regresión

Para simplificar, pensemos que tenemos dos covariables: x_1 y x_2.

Algunos ejemplos de modelos no paramétricos, semi-paramétricos y paramétricos cuando hay dos covariables son

Modelos no paramétricos

(i) $y = \eta(x_1, x_2) + \varepsilon$ donde $\eta(x_1, x_2)$ es una función continua.

Modelos de regresión

Para simplificar, pensemos que tenemos dos covariables: x_1 y x_2.

Algunos ejemplos de modelos no paramétricos, semi-paramétricos y paramétricos cuando hay dos covariables son

Modelos no paramétricos

(i) $y = \eta(x_1, x_2) + \varepsilon$ donde $\eta(x_1, x_2)$ es una función continua.

(ii) $y = \eta(x_1, x_2) + \varepsilon$ donde $\eta(x_1, x_2)$ es una función continua y derivable.

Modelos de regresión

Para simplificar, pensemos que tenemos dos covariables: x_1 y x_2.

Algunos ejemplos de modelos no paramétricos, semi-paramétricos y paramétricos cuando hay dos covariables son

Modelos no paramétricos

(i) $y = \eta(x_1, x_2) + \varepsilon$ donde $\eta(x_1, x_2)$ es una función continua.

(ii) $y = \eta(x_1, x_2) + \varepsilon$ donde $\eta(x_1, x_2)$ es una función continua y derivable.

(iii) $y = \eta_1(x_1) + \eta_2(x_2) + \varepsilon$, con η_i funciones continuas. Modelo aditivo

Modelos de regresión

Para simplificar, pensemos que tenemos dos covariables: x_1 y x_2.

Algunos ejemplos de modelos no paramétricos, semi-paramétricos y paramétricos cuando hay dos covariables son

Modelos no paramétricos

(i) $y = \eta(x_1, x_2) + \varepsilon$ donde $\eta(x_1, x_2)$ es una función continua.

(ii) $y = \eta(x_1, x_2) + \varepsilon$ donde $\eta(x_1, x_2)$ es una función continua y derivable.

(iii) $y = \eta_1(x_1) + \eta_2(x_2) + \varepsilon$, con η_i funciones continuas. Modelo aditivo

(iv) $y = \eta(x_1, x_2) + \varepsilon$ donde $\eta(x_1, x_2)$ es monótona creciente en x_1 y x_2. Modelo isotónico

Modelos de regresión

Modelos semi-paramétricos

(i) $y = \eta(\boldsymbol{\theta}^{\mathrm{T}}\,\mathbf{x}) + \varepsilon$ donde $\eta : \mathbb{R} \to \mathbb{R}$ es una función continua y $\boldsymbol{\theta} = (\theta_1, \theta_2)^{\mathrm{T}}$ es tal que $\|\boldsymbol{\theta}\| = 1$. Modelo de índice simple

Modelos de regresión

Modelos semi-paramétricos

(i) $y = \eta(\boldsymbol{\theta}^{\mathrm{T}}\, \mathbf{x}) + \varepsilon$ donde $\eta : \mathbb{R} \to \mathbb{R}$ es una función continua y $\boldsymbol{\theta} = (\theta_1, \theta_2)^{\mathrm{T}}$ es tal que $\|\boldsymbol{\theta}\| = 1$. Modelo de índice simple

(ii) $y = \eta(x_1) + \theta_2\, x_2 + \varepsilon$ donde $\eta : \mathbb{R} \to \mathbb{R}$ es una función continua y $\theta_2 \in \mathbb{R}$. Modelo parcialmente lineal

Modelos de regresión

Modelos semi-paramétricos

(i) $y = \eta(\boldsymbol{\theta}^{\mathrm{T}}\,\mathbf{x}) + \varepsilon$ donde $\eta : \mathbb{R} \to \mathbb{R}$ es una función continua y $\boldsymbol{\theta} = (\theta_1, \theta_2)^{\mathrm{T}}$ es tal que $\|\boldsymbol{\theta}\| = 1$. Modelo de índice simple

(ii) $y = \eta(x_1) + \theta_2\, x_2 + \varepsilon$ donde $\eta : \mathbb{R} \to \mathbb{R}$ es una función continua y $\theta_2 \in \mathbb{R}$. Modelo parcialmente lineal

(iii) $y = \theta_1 + x_2\, \eta(x_1) + \varepsilon$, donde $\eta : \mathbb{R} \to \mathbb{R}$ es una función continua y $\theta_1 \in \mathbb{R}$. Modelo de coeficientes variables

Modelos de regresión

Modelos paramétricos

(i) $y = \alpha + \beta x_1 + \gamma x_2 + \varepsilon$

Modelos de regresión

Modelos paramétricos

(i) $y = \alpha + \beta x_1 + \gamma x_2 + \varepsilon$

(ii) $y = \alpha \log(x_1) + \beta \log(x_2) + \gamma x_1^3 + \delta \text{ sen}(x_2) + \varepsilon$

Modelos de regresión

Modelos paramétricos

(i) $y = \alpha + \beta x_1 + \gamma x_2 + \varepsilon$

(ii) $y = \alpha \log(x_1) + \beta \log(x_2) + \gamma x_1^3 + \delta \text{ sen}(x_2) + \varepsilon$

(iii) $y = \alpha x_1^{\beta} x_2^{\delta} + \varepsilon$

Modelos de regresión

Modelos paramétricos

(i) $y = \alpha + \beta x_1 + \gamma x_2 + \varepsilon$

(ii) $y = \alpha \log(x_1) + \beta \log(x_2) + \gamma x_1^3 + \delta \text{ sen}(x_2) + \varepsilon$

(iii) $y = \alpha x_1^{\beta} x_2^{\delta} + \varepsilon$

(iv) $y = \alpha e^{\beta x_1} + \gamma e^{\delta x_2} + \varepsilon$

Modelo Lineal

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

Modelo Lineal

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

Es el caso de:

(i) $y = \alpha + \beta x_1 + \gamma x_2 + \varepsilon$

(ii) $y = \alpha \log(x_1) + \beta \log(x_2) + \gamma x_1^3 + \delta \sin(x_2) + \varepsilon$

Modelo Lineal

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

Es el caso de:

(i) $y = \alpha + \beta x_1 + \gamma x_2 + \varepsilon$

(ii) $y = \alpha \log(x_1) + \beta \log(x_2) + \gamma x_1^3 + \delta \sin(x_2) + \varepsilon$

Un modelo se dirá un **modelo de regresión lineal**, si es lineal en los parámetros, o sea, si

$$g(\mathbf{x}, a\,\boldsymbol{\theta}_1 + b\,\boldsymbol{\theta}_2) = a\,g(\mathbf{x}, \boldsymbol{\theta}_1) + b\,g(\mathbf{x}, \boldsymbol{\theta}_2)$$

Modelo Lineal

En los casos

(i) $y = \alpha + \beta x_1 + \gamma x_2 + \varepsilon$

(ii) $y = \alpha \log(x_1) + \beta \log(x_2) + \gamma x_1^3 + \delta \sin(x_2) + \varepsilon$

el modelo es **lineal** en los **parámetros**.

Modelo Lineal

En los casos

(i) $y = \alpha + \beta x_1 + \gamma x_2 + \varepsilon$

(ii) $y = \alpha \log(x_1) + \beta \log(x_2) + \gamma x_1^3 + \delta \sin(x_2) + \varepsilon$

el modelo es **lineal** en los **parámetros**.

No es el caso, por ejemplo, del modelo

$$g(x, \alpha, \beta) = \alpha \, e^{\beta x} \, ,$$

conocido como crecimiento exponencial, ya que no es lineal en el parámetro β.

Modelo Lineal

Algunos ejemplos sencillos de modelos lineales dependientes de una sola variable son:

$$
\begin{aligned}
g(x, \boldsymbol{\theta}) &= \alpha + \beta x & \boldsymbol{\theta} &= (\alpha, \beta)^{\mathrm{T}} \\
g(x, \boldsymbol{\theta}) &= \alpha + \beta x + \gamma x^2 & \boldsymbol{\theta} &= (\alpha, \beta, \gamma)^{\mathrm{T}} \\
g(x, \boldsymbol{\theta}) &= \alpha + \beta \log(x) & \boldsymbol{\theta} &= (\alpha, \beta)^{\mathrm{T}}
\end{aligned}
$$

Modelo Lineal

En general, en un modelo lineal tendremos que la $i-$ésima respuesta y_i está asociada a un vector de covariables $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$ de la siguiente forma

$$y_i = \theta_1 x_{i1} + \cdots + \theta_p x_{ip} + \epsilon_i \, .$$

es decir

$$g(\mathbf{x}_i, \boldsymbol{\theta}) = \theta_1 x_{i1} + \cdots + \theta_p x_{ip} = \boldsymbol{\theta}^{\mathrm{T}} \mathbf{x}_i$$

siendo $\boldsymbol{\theta}^{\mathrm{T}} = (\theta_1, \ldots, \theta_p)$.

Modelo Lineal

En general, en un modelo lineal tendremos que la $i-$ésima respuesta y_i está asociada a un vector de covariables $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$ de la siguiente forma

$$y_i = \theta_1 x_{i1} + \cdots + \theta_p x_{ip} + \epsilon_i \, .$$

es decir

$$g(\mathbf{x}_i, \boldsymbol{\theta}) = \theta_1 x_{i1} + \cdots + \theta_p x_{ip} = \boldsymbol{\theta}^{\mathrm{T}} \mathbf{x}_i$$

siendo $\boldsymbol{\theta}^{\mathrm{T}} = (\theta_1, \ldots, \theta_p)$.

- Eventualmente, las covariables podrían ser funciones de otras variables, tal como ocurre en el caso ii).
- Cuando $x_{i1} = 1$ para todo i, decimos que el modelo tiene intercept u ordenada al origen.

Hipótesis del modelo

Consideremos una muesta $(y_i, \mathbf{x}_i)$ con $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$, $1 \leq i \leq n$ que cumplen el modelo lineal

$$y_i = \theta_1 x_{i1} + \cdots + \theta_p x_{ip} + \epsilon_i = \boldsymbol{\theta}^{\mathrm{T}} \mathbf{x}_i + \epsilon_i \, .$$

Supondremos x_{ij}, $1 \leq j \leq p$ son determinísticas, es decir, no son variables aleatorias.

Hipótesis del modelo

Consideremos una muesta $(y_i, \mathbf{x}_i)$ con $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$, $1 \leq i \leq n$ que cumplen el modelo lineal

$$y_i = \theta_1 x_{i1} + \cdots + \theta_p x_{ip} + \epsilon_i = \boldsymbol{\theta}^{\mathrm{T}} \mathbf{x}_i + \epsilon_i \, .$$

Supondremos x_{ij}, $1 \leq j \leq p$ son determinísticas, es decir, no son variables aleatorias.

$\theta_1, \ldots, \theta_p$ son los p parámetros desconocidos a estimar.

Hipótesis del modelo

Las hipótesis que usaremos son

H1 $\mathbb{E}(\epsilon_i) = 0$, para $1 \leq i \leq n$

Hipótesis del modelo

Las hipótesis que usaremos son

H1 $\mathbb{E}(\epsilon_i) = 0$, para $1 \leq i \leq n$

H2 $\textsc{Var}(\epsilon_i) = \sigma^2$, para $1 \leq i \leq n$

$\textsc{Cov}(\epsilon_i, \epsilon_\ell) = 0$, para $1 \leq i \neq \ell \leq n$

Hipótesis del modelo

Las hipótesis que usaremos son

H1 $\mathbb{E}(\epsilon_i) = 0$, para $1 \leq i \leq n$

H2 $\textsc{Var}(\epsilon_i) = \sigma^2$, para $1 \leq i \leq n$

$\textsc{Cov}(\epsilon_i, \epsilon_\ell) = 0$, para $1 \leq i \neq \ell \leq n$

H3 $\epsilon_1, \ldots, \epsilon_n$ son i.i.d. $N(0, \sigma^2)$

Enfoque matricial

Queremos escribir el modelo

$$y_i = \theta_1 x_{i1} + \cdots + \theta_p x_{ip} + \epsilon_i = \boldsymbol{\theta}^{\mathrm{T}} \mathbf{x}_i + \epsilon_i \, .$$

en forma matricial.

Enfoque matricial

Queremos escribir el modelo

$$y_i = \theta_1 x_{i1} + \cdots + \theta_p x_{ip} + \epsilon_i = \boldsymbol{\theta}^{\mathrm{T}} \mathbf{x}_i + \epsilon_i \,.$$

en forma matricial. Sean

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} \mathbf{x}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_n^{\mathrm{T}} \end{pmatrix} \qquad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$\Downarrow$$

$$\Omega: \quad \underset{n\times 1}{\mathbf{y}} \quad = \quad \underset{n\times p}{\mathbf{X}} \quad \underset{p\times 1}{\boldsymbol{\theta}} \quad + \quad \underset{n\times 1}{\boldsymbol{\epsilon}}$$

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

En general, se elige de tal forma que tenga rango máximo, es decir, rango$(\mathbf{X}) = p$.

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

En general, se elige de tal forma que tenga rango máximo, es decir, rango$(\mathbf{X}) = p$.

Sin embargo esto no siempre es posible, como en el caso de algunos diseños tratados en análisis de la varianza (ANOVA).

Trataremos sólo el caso de rango completo.

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

En general, se elige de tal forma que tenga rango máximo, es decir, rango$(\mathbf{X}) = p$.

Sin embargo esto no siempre es posible, como en el caso de algunos diseños tratados en análisis de la varianza (ANOVA).

Trataremos sólo el caso de rango completo.

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

Algunos ejemplos: Modelo de regresión simple

En el caso más sencillo de regresión simple, tendríamos

$$y_i = \theta_1 + \theta_2 x_i + \epsilon_i \quad 1 \leq i \leq n$$

$$p = 2 \qquad \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \qquad \boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$$

Modelo de regresión a través del origen

Es un modelo de regresión simple donde se supone que la ordenada al origen es 0

$$y_i = \theta_1 x_i + \epsilon_i \quad 1 \leq i \leq n$$

$$p = 1 \qquad \mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad \boldsymbol{\theta} = \theta_1$$

Modelo de posición

$$y_i = \mu + \epsilon_i \quad 1 \leq i \leq n$$

$$p = 1 \qquad \mathbf{X} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \qquad \theta = \mu$$

Modelo de 2 muestras

$$y_{i1} = \mu_1 + \epsilon_{i1} \quad 1 \leq i \leq n_1$$

$$y_{i2} = \mu_2 + \epsilon_{i2} \quad 1 \leq i \leq n_2$$

$$p = 2 \qquad \mathbf{X} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \qquad \boldsymbol{\theta} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

Modelo de regresión cuadrática

$$y_i = \theta_1 + \theta_2 x_i + \theta_3 x_i^2 + \epsilon_i \quad 1 \leq i \leq n$$

$$p = 3 \qquad \mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix} \qquad \boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix}$$

Definiciones

- Sea $\mathbf{V} = (V_{ij})$ una matriz de variables aleatorias, se define

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

- Sean $\mathbf{x} \in \mathbb{R}^p$, $\mathbf{y} \in \mathbb{R}^q$ dos vectores aleatorios, se define la matriz de covarianzas de $\mathbf{x}$ e $\mathbf{y}$ como

$$\begin{aligned} \mathbf{\Sigma}_{\mathbf{x},\mathbf{y}} &= \operatorname{Cov}(\mathbf{x}, \mathbf{y}) = (\operatorname{Cov}(x_\ell, y_s))_{1\leq \ell \leq p, 1 \leq s \leq q} \\ &= \mathbb{E}(\mathbf{x}\,\mathbf{y}^{\mathrm{T}}) - \mathbb{E}(\mathbf{x})\,\mathbb{E}(\mathbf{y})^{\mathrm{T}}. \end{aligned}$$

- Dado $\mathbf{x} \in \mathbb{R}^p$, definimos la matriz de covarianza de $\mathbf{x}$ como

$$\mathbf{\Sigma}_{\mathbf{x}} = \operatorname{Cov}(\mathbf{x}, \mathbf{x})$$

Definiciones

- Sea $\mathbf{V} = (V_{ij})$ una matriz de variables aleatorias, se define

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

- Sean $\mathbf{x} \in \mathbb{R}^p$, $\mathbf{y} \in \mathbb{R}^q$ dos vectores aleatorios, se define la matriz de covarianzas de $\mathbf{x}$ e $\mathbf{y}$ como

$$\begin{aligned} \mathbf{\Sigma}_{\mathbf{x},\mathbf{y}} &= \operatorname{Cov}(\mathbf{x},\mathbf{y}) = (\operatorname{Cov}(x_\ell, y_s))_{1\le\ell\le p, 1\le s\le q} \\ &= \mathbb{E}(\mathbf{x}\,\mathbf{y}^{\mathrm{T}}) - \mathbb{E}(\mathbf{x})\,\mathbb{E}(\mathbf{y})^{\mathrm{T}}. \end{aligned}$$

- Dado $\mathbf{x} \in \mathbb{R}^p$, definimos la matriz de covarianza de $\mathbf{x}$ como

$$\mathbf{\Sigma}_{\mathbf{x}} = \operatorname{Cov}(\mathbf{x},\mathbf{x}) = \mathbb{E}(\mathbf{x}\,\mathbf{x}^{\mathrm{T}}) - \mathbb{E}(\mathbf{x})\,\mathbb{E}(\mathbf{x})^{\mathrm{T}} = (\operatorname{Cov}(x_\ell, x_s))_{\substack{1\le\ell\le p \\ 1\le s\le p}}$$

- Escribiremos $\operatorname{Var}(\mathbf{x}) = \mathbf{\Sigma}_{\mathbf{x}}$
- $\mathbf{\Sigma}_{\mathbf{x}}$ es una matriz simétrica definida no–negativa.

Propiedades

Sean $\mathbf{V} = (V_{ij}) \in \mathbb{R}^{\ell \times m}$ una matriz de variables aleatorias, $\mathbf{x} \in \mathbb{R}^p$ e $\mathbf{y} \in \mathbb{R}^q$ dos vectores aleatorios

- $\mathbb{E}(\mathbf{AVB} + \mathbf{C}) = \mathbf{A}\,\mathbb{E}(\mathbf{V})\,\mathbf{B} + \mathbf{C}$, con $\mathbf{A} \in \mathbb{R}^{s \times \ell}$, $\mathbf{B} \in \mathbb{R}^{m \times k}$ y $\mathbf{C} \in \mathbb{R}^{s \times k}$ matrices fijas.

- Si $\mathbf{x}$ e $\mathbf{y}$ son independientes, entonces $\mathrm{Cov}\,(\mathbf{x}, \mathbf{y}) = \mathbf{0}$.

- $\mathrm{Cov}\,(\mathbf{A}\,\mathbf{x}, \mathbf{B}\,\mathbf{y}) = \mathbf{A}\,\mathrm{Cov}\,(\mathbf{x}, \mathbf{y})\;\mathbf{B}^{\mathrm{T}}$, siendo $\mathbf{A} \in \mathbb{R}^{s \times p}$ y $\mathbf{B} \in \mathbb{R}^{\ell \times q}$ matrices fijas.

- $\mathrm{Var}(\mathbf{A}\,\mathbf{x} + \mathbf{d}) = \mathbf{A}\,\mathrm{Var}(\mathbf{x})\,\mathbf{A}^{\mathrm{T}}$ siendo $\mathbf{A} \in \mathbb{R}^{s \times p}$ una matriz fija y $\mathbf{d} \in \mathbb{R}^s$ un vector fijo.

Propiedades

Lema 1. Sea $\mathbf{x}$ un vector aleatorio $n-$dimensional y sea $\mathbf{A} \in \mathbb{R}^{n \times n}$ una matriz simétrica.

Si $\mathbb{E}(\mathbf{x}) = \boldsymbol{\mu}$ y su matriz de covarianza es $\boldsymbol{\Sigma}_{\mathbf{x}}$ entonces

$$\mathbb{E}(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}) = \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^{\mathrm{T}}\mathbf{A}\ \boldsymbol{\mu}$$

Hipótesis del modelo

En el caso del modelo Ω,la hipótesis H1 se escribe

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

Por otra parte, H2 puede escribirse como

$$\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = \mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathrm{T}}) = \sigma^2\mathbf{I}$$

El modelo lineal

El modelo lineal puede entonces escribirse como:

$$\Omega : \mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon} \qquad \text{con} \qquad \mathbb{E}(\boldsymbol{\epsilon}) = \mathbf{0} \qquad \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = \sigma^2 \mathbf{I}$$

El modelo lineal

El modelo lineal puede entonces escribirse como:

$$\Omega : \mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon} \quad \text{con} \quad \mathbb{E}(\boldsymbol{\epsilon}) = \mathbf{0} \quad \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = \sigma^2\mathbf{I}$$

o equivalentemente

$$\Omega : \mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\theta} \quad \boldsymbol{\Sigma}_{\mathbf{y}} = \sigma^2\mathbf{I}$$

El modelo lineal: Interpretación

Supondremos que $n \geq p$,

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_n^{\mathrm{T}} \end{pmatrix} = \left(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(p)} \right)$$

Sea $\mathcal{V}$ el subespacio de $\mathbb{R}^n$ generado por las columnas de $\mathbf{X}$.

$$\dim(\mathcal{V}) = \text{rango}(\mathbf{X})$$

El modelo lineal: Interpretación

Supondremos que $n \geq p$,

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_n^{\mathrm{T}} \end{pmatrix} = \left(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(p)} \right)$$

Sea $\mathcal{V}$ el subespacio de $\mathbb{R}^n$ generado por las columnas de $\mathbf{X}$.

$$\dim(\mathcal{V}) = \text{rango}(\mathbf{X})$$

La hipótesis H1 dice que

$$\mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\theta} = \sum_{j=1}^{p} \theta_j \mathbf{x}^{(j)} \in \mathcal{V}$$

El modelo lineal: Interpretación

Supondremos que $n \geq p$,

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_n^{\mathrm{T}} \end{pmatrix} = \left(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(p)} \right)$$

Sea $\mathcal{V}$ el subespacio de $\mathbb{R}^n$ generado por las columnas de $\mathbf{X}$.

$$\dim(\mathcal{V}) = \text{rango}(\mathbf{X})$$

La hipótesis H1 dice que

$$\mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\theta} = \sum_{j=1}^{p} \theta_j \mathbf{x}^{(j)} \in \mathcal{V}$$

Luego, H1 es equivalente a decir que $\mathbb{E}(\mathbf{y}) \in \mathcal{V}$

El modelo lineal: Interpretación

H1 es equivalente a decir que $\mathbb{E}(\mathbf{y}) \in \mathcal{V}$

El modelo lineal: Interpretación

H1 es equivalente a decir que $\mathbb{E}(\mathbf{y}) \in \mathcal{V}$

- Si $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(p)}$ son linealmente independientes o sea, si $\dim(\mathcal{V}) = \text{rango}(\mathbf{X}) = p$, existe un **único** $\boldsymbol{\theta}$ tal que $\mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\theta}$.

El modelo lineal: Interpretación

H1 es equivalente a decir que $\mathbb{E}(\mathbf{y}) \in \mathcal{V}$

- Si $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(p)}$ son linealmente independientes o sea, si $\dim(\mathcal{V}) = \text{rango}(\mathbf{X}) = p$, existe un **único** $\boldsymbol{\theta}$ tal que $\mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\theta}$.

- Si $\dim(\mathcal{V}) = q < p$, $\boldsymbol{\theta}$ no está determinado aunque $\mathbb{E}(\mathbf{y})$ sí lo está. En este caso, tenemos dos opciones

El modelo lineal: Interpretación

H1 es equivalente a decir que $\mathbb{E}(\mathbf{y}) \in \mathcal{V}$

- Si $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(p)}$ son linealmente independientes o sea, si $\dim(\mathcal{V}) = \text{rango}(\mathbf{X}) = p$, existe un **único** $\boldsymbol{\theta}$ tal que $\mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\theta}$.

- Si $\dim(\mathcal{V}) = q < p$, $\boldsymbol{\theta}$ no está determinado aunque $\mathbb{E}(\mathbf{y})$ sí lo está. En este caso, tenemos dos opciones
  - Si $\dim(\mathcal{V}) = q$ y $\mathbf{x}^{(j_1)}, \ldots, \mathbf{x}^{(j_q)}$ generan $\mathcal{V}$, tenemos que

$$\mathbb{E}(\mathbf{y}) = \sum_{s=1}^{q} \widetilde{\theta}_s \, \mathbf{x}^{(j_s)}$$

y el vector $\widetilde{\boldsymbol{\theta}} = (\widetilde{\theta}_1, \ldots, \widetilde{\theta}_q)^{\mathrm{T}}$ está bien determinado.

El modelo lineal: Interpretación

H1 es equivalente a decir que $\mathbb{E}(\mathbf{y}) \in \mathcal{V}$

- Si $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(p)}$ son linealmente independientes o sea, si $\dim(\mathcal{V}) = \text{rango}(\mathbf{X}) = p$, existe un **único** $\boldsymbol{\theta}$ tal que $\mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\theta}$.

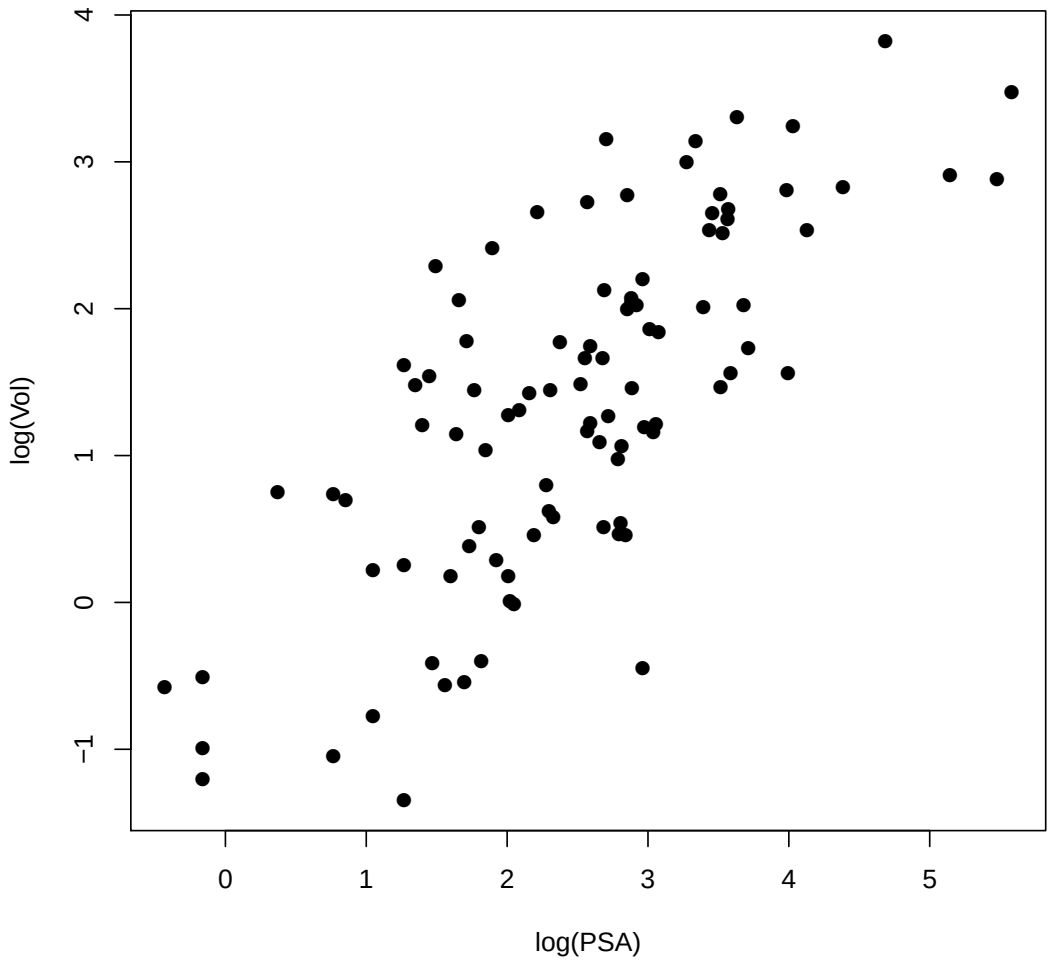
- Si $\dim(\mathcal{V}) = q < p$, $\boldsymbol{\theta}$ no está determinado aunque $\mathbb{E}(\mathbf{y})$ sí lo está. En este caso, tenemos dos opciones
  - Si $\dim(\mathcal{V}) = q$ y $\mathbf{x}^{(j_1)}, \ldots, \mathbf{x}^{(j_q)}$ generan $\mathcal{V}$, tenemos que

$$\mathbb{E}(\mathbf{y}) = \sum_{s=1}^{q} \widetilde{\theta}_s \, \mathbf{x}^{(j_s)}$$

    y el vector $\widetilde{\boldsymbol{\theta}} = (\widetilde{\theta}_1, \ldots, \widetilde{\theta}_q)^{\mathrm{T}}$ está bien determinado.
  - La otra opción es imponer restricciones lineales a $\boldsymbol{\theta}$ mediante

$$\mathbf{A}\boldsymbol{\theta} = 0$$

    con $\mathbf{A}$ una matriz elegida para que el vector quede unívocamente determinado.


log(Vol)
log(PSA)
−1
0
1
2
3
4
0
1
2
3
4
5

¿Cómo estimamos los parámetros?

Mínimos Cuadrados

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

¿Cómo estimamos los parámetros?

Mínimos Cuadrados

Si los puntos en un gráfico parecen seguir una recta como en la figura anterior, 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

¿Cómo estimamos los parámetros?

Mínimos Cuadrados

Si los puntos en un gráfico parecen seguir una recta como en la figura anterior, 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.

¿Cómo estimamos los parámetros?

Mínimos Cuadrados

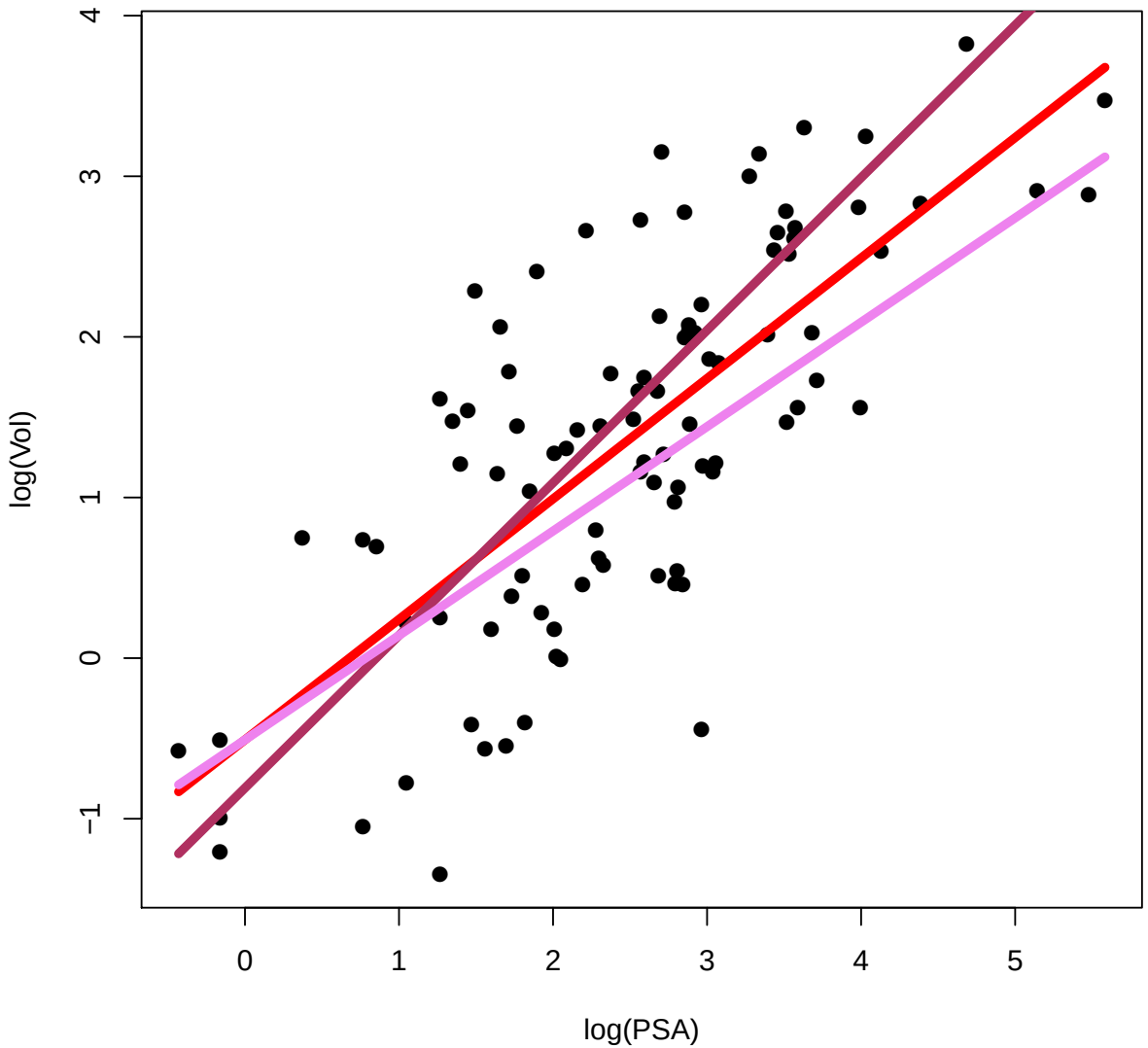
Si los puntos en un gráfico parecen seguir una recta como en la figura anterior, el problema es elegir la recta que mejor ajusta los puntos.

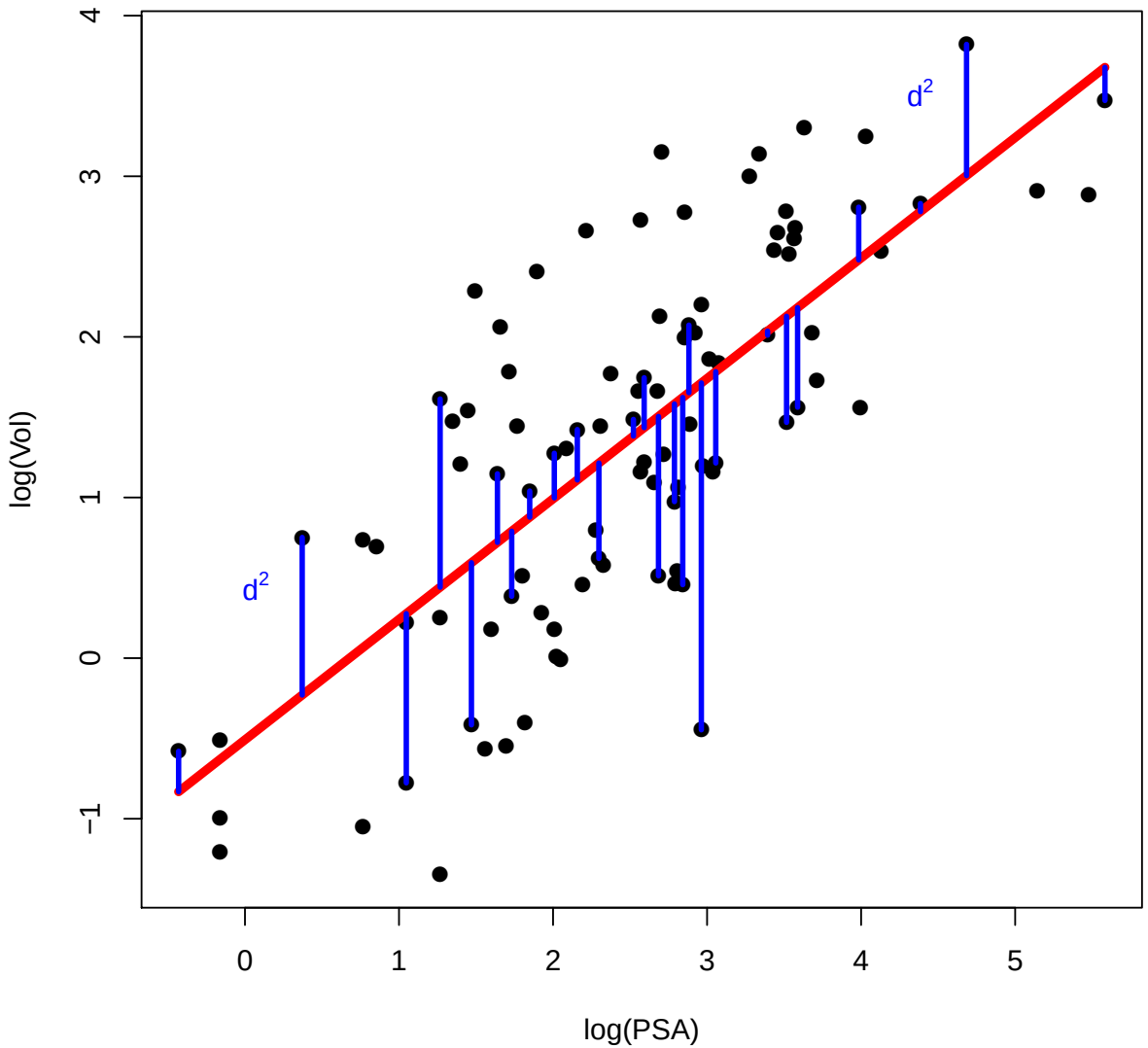
Tendremos en cuenta:

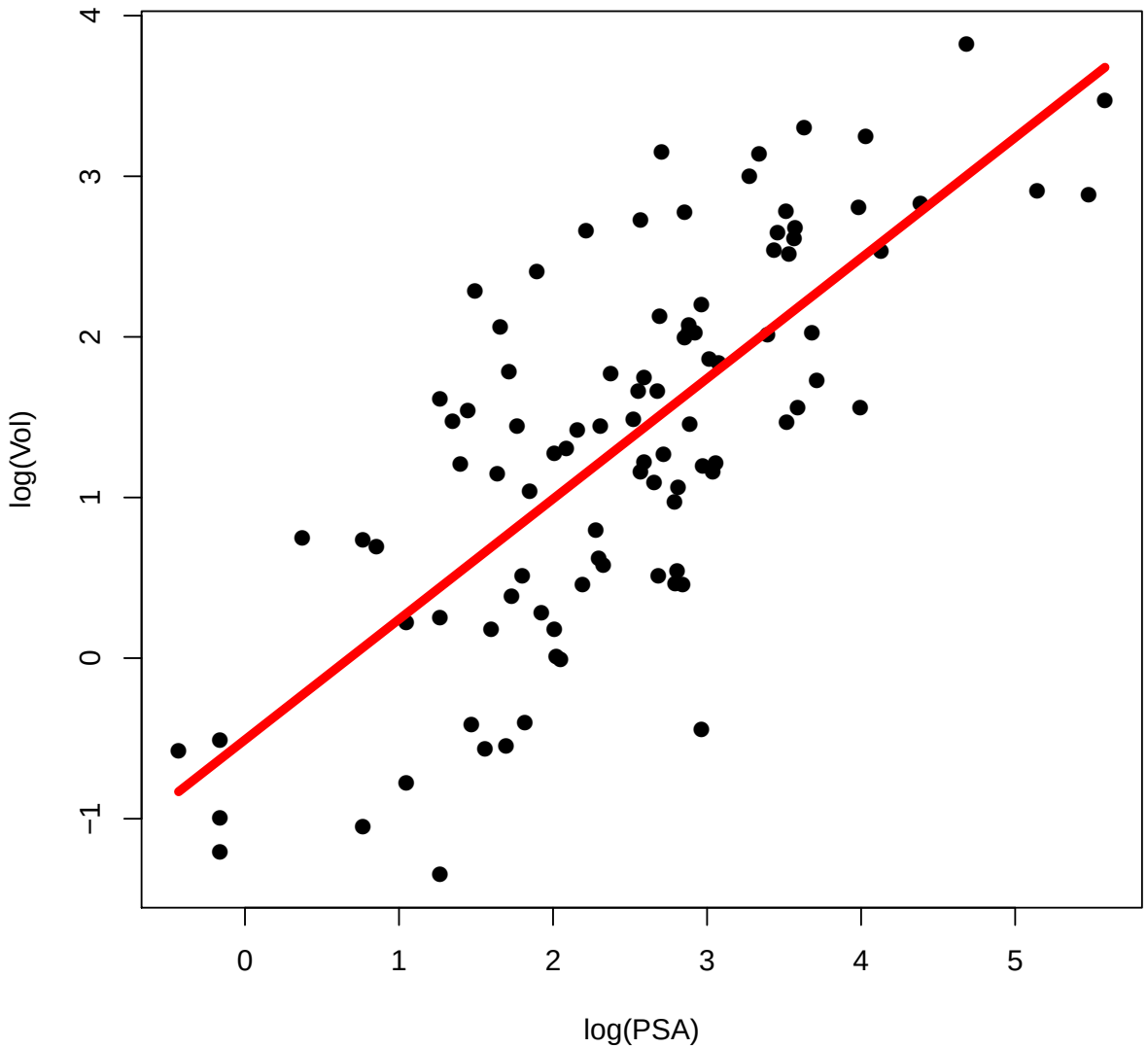
a) tomar una distancia promedio de la recta a todos los puntos

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.


log(Vol)
log(PSA)
−1
0
1
2
3
4
0
1
2
3
4
5


log(Vol)
log(PSA)
d2
d2
4
3
2
1
0
−1
0
1
2
3
4
5


log(Vol)
log(PSA)
−1
0
1
2
3
4
0
1
2
3
4
5

Repaso: Mínimos Cuadrados

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

$$y_i = g(\mathbf{x}_i, \theta_1 \dots \theta_p) + \varepsilon_i$$

donde

- $\mathbb{E}(\varepsilon_i) = 0$, $\textsc{Var}(\varepsilon_i) = \sigma^2$, $\textsc{Cov}(\varepsilon_i, \varepsilon_j) = 0$

Repaso: Mínimos Cuadrados

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

$$y_i = g(\mathbf{x}_i, \theta_1 \dots \theta_p) + \varepsilon_i$$

donde

- $\mathbb{E}(\varepsilon_i) = 0, \;\; \textsc{Var}(\varepsilon_i) = \sigma^2, \; \textsc{Cov}(\varepsilon_i, \varepsilon_j) = 0$

La función g es conocida salvo por los parámetros $\theta_1 \dots \theta_p$.

Estimamos $\theta_1 \dots \theta_p$ minimizando la suma de cuadrados residual

Repaso: Mínimos Cuadrados

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

$$y_i = g(\mathbf{x}_i, \theta_1 \dots \theta_p) + \varepsilon_i$$

donde

- $\mathbb{E}(\varepsilon_i) = 0, \;\; \text{VAR}(\varepsilon_i) = \sigma^2, \; \text{COV}(\varepsilon_i, \varepsilon_j) = 0$

La función g es conocida salvo por los parámetros $\theta_1 \dots \theta_p$.

Estimamos $\theta_1 \dots \theta_p$ minimizando la suma de cuadrados residual

El EMC $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \dots, \widehat{\theta}_p)$ cumple

$$\widehat{\boldsymbol{\theta}} = \underset{\mathbf{b} \in \mathbb{R}^p}{\text{argmin}} \sum_{i=1}^{n} \{y_i - g(\mathbf{x}_i, b_1 \dots, b_p)\}^2$$

Repaso: Mínimos Cuadrados

En el caso de la regresión simple en el cual $g(x, \theta_1, \theta_2) = \theta_1 + \theta_2\, x$,

$$\widehat{\theta} = \underset{\mathbf{b}\in\mathbb{R}^2}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \left[y_i - (b_1 + b_2 x_i)\right]^2 .$$

Esta suma se llama la suma de cuadrados residual para la recta

La recta resultante recta de cuadrados mínimos.

Caso Modelo Lineal

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

$$\mathbf{r}(\mathbf{b}) = \mathbf{y} - \mathbf{X}\mathbf{b}\,.$$

El estimador de mínimos cuadrados de $\theta_1 \dots \theta_p$ 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 $\|\mathbf{u}\|^2 = \mathbf{u}^{\mathrm{T}}\mathbf{u} = \sum_{i=1}^{n} u_i^2$.

Llamemos

$$\mathcal{S}(\mathbf{b}) = \|\mathbf{r}(\mathbf{b})\|^2 = \|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2 = (\mathbf{y} - \mathbf{X}\mathbf{b})^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\mathbf{b})$$

Caso Modelo Lineal

El estimador de mínimos cuadrados de $\boldsymbol{\theta}$, es un conjunto de funciones de $\mathbf{y}$,

$$\widehat{\theta}_1 = \widehat{\theta}_1(\mathbf{y}), \quad \widehat{\theta}_2 = \widehat{\theta}_2(\mathbf{y}), \ldots \quad \widehat{\theta}_p = \widehat{\theta}_p(\mathbf{y})$$

que minimiza $\mathcal{S}(\mathbf{b})$.

$$\mathcal{S}(\widehat{\boldsymbol{\theta}}) = \underset{\mathbf{b}\in\mathbb{R}^p}{\operatorname{argmin}}\, \mathcal{S}(\mathbf{b})$$

Caso Modelo Lineal

El estimador de mínimos cuadrados de $\boldsymbol{\theta}$, es un conjunto de funciones de $\mathbf{y}$,

$$\widehat{\theta}_1 = \widehat{\theta}_1(\mathbf{y}), \quad \widehat{\theta}_2 = \widehat{\theta}_2(\mathbf{y}), \ldots \quad \widehat{\theta}_p = \widehat{\theta}_p(\mathbf{y})$$

que minimiza $\mathcal{S}(\mathbf{b})$.

$$\mathcal{S}(\widehat{\boldsymbol{\theta}}) = \underset{\mathbf{b}\in\mathbb{R}^p}{\operatorname{argmin}}\, \mathcal{S}(\mathbf{b})$$

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

Caso de errores normales

Bajo H3, o sea si $\boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{I})$ se tiene que

- $\widehat{\boldsymbol{\theta}}$ es el EMV de $\boldsymbol{\theta}$ si y sólo si

$$\mathcal{S}(\widehat{\boldsymbol{\theta}}) = \underset{\mathbf{b} \in \mathbb{R}^p}{\operatorname{argmin}} \mathcal{S}(\mathbf{b})$$

- El EMV de σ^2 es

$$\widehat{\sigma}^2 = \frac{1}{n} \mathcal{S}(\widehat{\boldsymbol{\theta}})$$

Ecuaciones Normales

$\widehat{\boldsymbol{\theta}}$ minimiza $\mathcal{S}(\mathbf{b})$, por lo tanto, derivando e igualando a 0 obtenemos las **ecuaciones normales**.

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

$$\frac{\partial \mathcal{S}(\mathbf{b})}{\partial b_k}\Big|_{\mathbf{b}=\widehat{\boldsymbol{\theta}}} = -2\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij} b_j\right) x_{ik}\Big|_{\mathbf{b}=\widehat{\boldsymbol{\theta}}} = 0$$

Ecuaciones Normales

$\widehat{\boldsymbol{\theta}}$ minimiza $\mathcal{S}(\mathbf{b})$, por lo tanto, derivando e igualando a 0 obtenemos las **ecuaciones normales**.

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

$$\frac{\partial \mathcal{S}(\mathbf{b})}{\partial b_k}\Big|_{\mathbf{b}=\widehat{\boldsymbol{\theta}}} = -2\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij} b_j\right) x_{ik}\Big|_{\mathbf{b}=\widehat{\boldsymbol{\theta}}} = 0$$

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

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

Ecuaciones Normales

$\widehat{\boldsymbol{\theta}}$ minimiza $\mathcal{S}(\mathbf{b})$, por lo tanto, derivando e igualando a 0 obtenemos las **ecuaciones normales**.

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

$$\frac{\partial \mathcal{S}(\mathbf{b})}{\partial b_k}\Big|_{\mathbf{b}=\widehat{\boldsymbol{\theta}}} = -2\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij} b_j\right) x_{ik}\Big|_{\mathbf{b}=\widehat{\boldsymbol{\theta}}} = 0$$

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

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

reordenando la suma

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

Ecuaciones Normales

Como

$$\sum_{i=1}^{n} y_i x_{ik} = \sum_{j=1}^{p} \widehat{\theta}_j \sum_{i=1}^{n} x_{ij} x_{ik} \quad 1 \leq k \leq p$$

Ecuaciones Normales

Como

$$\sum_{i=1}^{n} y_i x_{ik} \quad = \quad \sum_{j=1}^{p} \widehat{\theta}_j \sum_{i=1}^{n} x_{ij} x_{ik} \quad 1 \leq k \leq p$$

El estimador de mínimos cuadrados cumple estas p ecuaciones, entonces

$$\mathbf{X}^{\mathrm{T}} \mathbf{X} \widehat{\theta} = \mathbf{X}^{\mathrm{T}} \mathbf{y} \,,$$

que se conocen como ecuaciones normales.

Ecuaciones Normales

Como

$$\sum_{i=1}^{n} y_i x_{ik} \quad = \quad \sum_{j=1}^{p} \widehat{\theta}_j \sum_{i=1}^{n} x_{ij} x_{ik} \quad 1 \leq k \leq p$$

El estimador de mínimos cuadrados cumple estas p ecuaciones, entonces

$$\mathbf{X}^{\mathrm{T}} \mathbf{X} \widehat{\theta} = \mathbf{X}^{\mathrm{T}} \mathbf{y} \,,$$

que se conocen como ecuaciones normales.

Si rango$(\mathbf{X}) = p$, $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ es no singular, la solución es única y resulta

$$\widehat{\theta} = \left(\mathbf{X}^{\mathrm{T}} \mathbf{X}\right)^{-1} \mathbf{X}^{\mathrm{T}} \mathbf{y} \,.$$

Ejemplo

En el caso de regresión simple tendríamos

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

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

El sistema sería

$$\begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix} \begin{pmatrix} \widehat{\theta}_1 \\ \widehat{\theta}_2 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix}$$

Ejemplo

La inversa resulta

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

y además

$$\mathbf{X}^{\mathrm{T}}\mathbf{y} = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix}$$

y por lo tanto

$$\widehat{\boldsymbol{\theta}} = \begin{pmatrix} \widehat{\theta}_1 \\ \widehat{\theta}_2 \end{pmatrix} = \frac{1}{n\sum_{i=1}^{n}(x_i - \overline{x})^2} \begin{pmatrix} (\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} x_i y_i) \\ n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} x_i) \end{pmatrix}$$

Ejemplo

Entonces

$$\widehat{\theta}_1 = \overline{y} - \overline{x}\,\widehat{\theta}_2$$

y por otro lado

$$\widehat{\theta}_2 = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Interpretación Geométrica del EMC

Nuestro modelo plantea

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

Luego, si

$$\boldsymbol{\eta} = \mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\theta}$$

Tenemos

$$\boldsymbol{\eta} = \theta_1 \mathbf{x}^{(1)} + \cdots + \theta_p \mathbf{x}^{(p)} \in \mathcal{V}$$

Interpretación Geométrica del EMC

$$\mathcal{S}(\widehat{\boldsymbol{\theta}}) = \|\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\theta}}\|^2 = \min_{\mathbf{b}\in\mathbb{R}^p} \|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2 = \min_{\mathbf{z}\in\mathcal{V}} \|\mathbf{y} - \mathbf{z}\|^2$$

Interpretación Geométrica del EMC

$$\mathcal{S}(\widehat{\boldsymbol{\theta}}) = \|\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\theta}}\|^2 = \min_{\mathbf{b}\in\mathbb{R}^p} \|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2 = \min_{\mathbf{z}\in\mathcal{V}} \|\mathbf{y} - \mathbf{z}\|^2$$

Es decir, que

$$\widehat{\boldsymbol{\eta}} = \mathbf{X}\widehat{\boldsymbol{\theta}}$$

es la proyección ortogonal de $\mathbf{y}$ sobre $\mathcal{V}$.

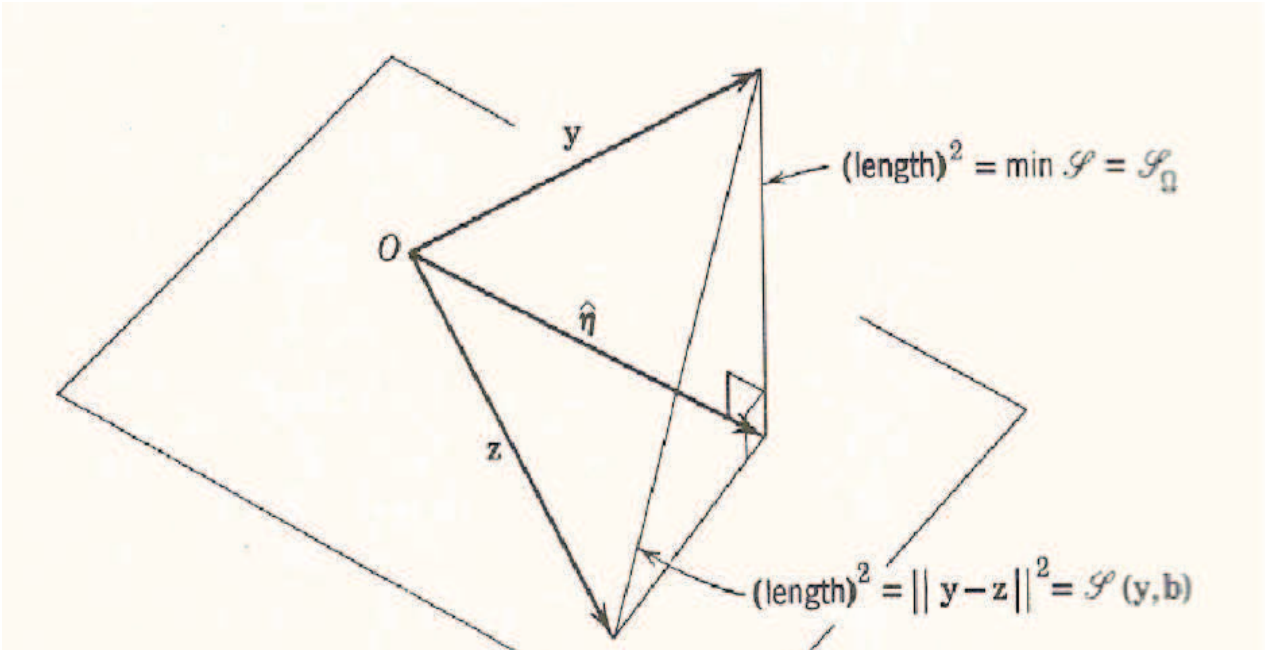
Interpretación Geométrica del EMC

$$\mathcal{S}(\widehat{\boldsymbol{\theta}}) = \|\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\theta}}\|^2 = \min_{\mathbf{b}\in\mathbb{R}^p} \|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2 = \min_{\mathbf{z}\in\mathcal{V}} \|\mathbf{y} - \mathbf{z}\|^2$$

Es decir, que

$$\widehat{\boldsymbol{\eta}} = \mathbf{X}\widehat{\boldsymbol{\theta}}$$

es la proyección ortogonal de $\mathbf{y}$ sobre $\mathcal{V}$.



Interpretación Geométrica del EMC

Por lo tanto, para todo $\mathbf{z} \in \mathcal{V}$

$$\mathbf{z}^{\mathrm{T}}(\mathbf{y} - \widehat{\boldsymbol{\eta}}) = 0$$

Interpretación Geométrica del EMC

Por lo tanto, para todo $\mathbf{z} \in \mathcal{V}$

$$\mathbf{z}^{\mathrm{T}}(\mathbf{y} - \widehat{\boldsymbol{\eta}}) = 0$$

en particular,

$$(\mathbf{y} - \widehat{\boldsymbol{\eta}})^{\mathrm{T}}\mathbf{x}^{(j)} = 0 \quad 1 \leq j \leq p \quad \Leftrightarrow \quad \mathbf{X}^{\mathrm{T}}(\mathbf{y} - \widehat{\boldsymbol{\eta}}) = 0$$

Interpretación Geométrica del EMC

Por lo tanto, para todo $\mathbf{z} \in \mathcal{V}$

$$\mathbf{z}^{\mathrm{T}}(\mathbf{y} - \widehat{\boldsymbol{\eta}}) = 0$$

en particular,

$$(\mathbf{y} - \widehat{\boldsymbol{\eta}})^{\mathrm{T}}\mathbf{x}^{(j)} = 0 \quad 1 \leq j \leq p \quad \Leftrightarrow \quad \mathbf{X}^{\mathrm{T}}(\mathbf{y} - \widehat{\boldsymbol{\eta}}) = 0$$

En términos de las ecuaciones normales tenemos que:

$$\mathbf{X}^{\mathrm{T}}\widehat{\boldsymbol{\eta}} \;=\; \mathbf{X}^{\mathrm{T}}\mathbf{y}$$

Interpretación Geométrica del EMC

Por lo tanto, para todo $\mathbf{z} \in \mathcal{V}$

$$\mathbf{z}^{\mathrm{T}}(\mathbf{y} - \widehat{\boldsymbol{\eta}}) = 0$$

en particular,

$$(\mathbf{y} - \widehat{\boldsymbol{\eta}})^{\mathrm{T}}\mathbf{x}^{(j)} = 0 \quad 1 \leq j \leq p \quad \Leftrightarrow \quad \mathbf{X}^{\mathrm{T}}(\mathbf{y} - \widehat{\boldsymbol{\eta}}) = 0$$

En términos de las ecuaciones normales tenemos que:

$$\mathbf{X}^{\mathrm{T}}\widehat{\boldsymbol{\eta}} = \mathbf{X}^{\mathrm{T}}\mathbf{y}$$

$$\mathbf{X}^{\mathrm{T}}\mathbf{X}\widehat{\boldsymbol{\theta}} = \mathbf{X}^{\mathrm{T}}\mathbf{y}$$

Interpretación Geométrica del EMC

Es decir, tenemos que

Proposición 1.

- El EMC de $\boldsymbol{\theta}$ existe y es cualquier solución de

$$\widehat{\boldsymbol{\eta}} = \mathbf{X}\widehat{\boldsymbol{\theta}}$$

- $\widehat{\boldsymbol{\theta}}$ es EMC si y sólo si

$$\mathbf{X}^{\mathrm{T}}\mathbf{X}\widehat{\boldsymbol{\theta}} = \mathbf{X}^{\mathrm{T}}\mathbf{y}$$

- El EMC es único si rango$(\mathbf{X}) = p$. Si rango$(\mathbf{X}) = p$

$$\widehat{\boldsymbol{\theta}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

Interpretación Geométrica del EMC

Proposición 2.

- El valor mínimo de $\mathcal{S}(\mathbf{b})$ es igual a

$$\mathcal{S}(\widehat{\boldsymbol{\theta}}) = \|\mathbf{y}\|^2 - \|\mathbf{X}\widehat{\boldsymbol{\theta}}\|^2 = \mathbf{y}^{\mathrm{T}}\mathbf{y} - \mathbf{y}^{\mathrm{T}}\mathbf{X}\widehat{\boldsymbol{\theta}}$$

- Si rango$(\mathbf{X}) = p$

$$\mathcal{S}(\widehat{\boldsymbol{\theta}}) = \mathbf{y}^{\mathrm{T}}\mathbf{y} - \mathbf{y}^{\mathrm{T}}\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

De ahora en más rango($\mathbf{X}$) = p

Por lo tanto,

$$\widehat{\boldsymbol{\theta}} \quad = \quad (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

De ahora en más rango($\mathbf{X}$) = p

Por lo tanto,

$$\widehat{\boldsymbol{\theta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

de donde

$$\widehat{\boldsymbol{\eta}} = \mathbf{X}\,\widehat{\boldsymbol{\theta}} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} = \mathbf{P}\mathbf{y}$$

De ahora en más rango($\mathbf{X}$) = p

Por lo tanto,

$$\widehat{\boldsymbol{\theta}} \quad = \quad (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

de donde

$$\widehat{\boldsymbol{\eta}} = \mathbf{X}\,\widehat{\boldsymbol{\theta}} \quad = \quad \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} = \mathbf{P}\mathbf{y}$$

Observemos que

$$\widehat{\mathbf{y}} = \widehat{\boldsymbol{\eta}} = \mathbf{X}\,\widehat{\boldsymbol{\theta}}$$

corresponde al valor predicho de $\mathbf{y}$.

$$\text{rango}(\mathbf{X}) = p$$

El vector de residuos se escribe como

$$\mathbf{r}(\widehat{\boldsymbol{\theta}}) \quad = \quad \mathbf{y} - \widehat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\theta}}$$

$$\text{rango}(\mathbf{X}) = p$$

El vector de residuos se escribe como

$$\begin{aligned} \mathbf{r}(\widehat{\boldsymbol{\theta}}) &= \mathbf{y} - \widehat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\theta}} \\ &= \mathbf{y} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} \end{aligned}$$

$$\text{rango}(\mathbf{X}) = p$$

El vector de residuos se escribe como

$$\begin{aligned} \mathbf{r}(\widehat{\theta}) &= \mathbf{y} - \widehat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\widehat{\theta} \\ &= \mathbf{y} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} \\ &= \mathbf{y} - \mathbf{P}\mathbf{y} = (\mathbf{I} - \mathbf{P})\mathbf{y} \end{aligned}$$

donde

$$\mathbf{P} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}} \in \mathbb{R}^{n \times n}$$

Propiedades de $\mathbf{P}$

$\mathbf{P} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}} \in \mathbb{R}^{n\times n}$ es la matriz de **proyección sobre el espacio generado por las columnas de $\mathbf{X}$**.

Propiedades de $\mathbf{P}$

$\mathbf{P} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}} \in \mathbb{R}^{n\times n}$ es la matriz de **proyección sobre el espacio generado por las columnas de X**.

Suele llamarse a esta matriz $\mathbf{P}$ por ser de proyección $\mathbf{P}$ o $\mathbf{H}$ de hat matrix.

$\mathbf{P}$ es una matriz simétrica e idempotente, es decir:

$$\mathbf{P} = \mathbf{P}^{\mathrm{T}} = \mathbf{P}^2$$

Propiedades de $\mathbf{P}$

$\mathbf{P} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}} \in \mathbb{R}^{n\times n}$ es la matriz de **proyección sobre el espacio generado por las columnas de X**.

Suele llamarse a esta matriz $\mathbf{P}$ por ser de proyección $\mathbf{P}$ o $\mathbf{H}$ de hat matrix.

$\mathbf{P}$ es una matriz simétrica e idempotente, es decir:

$$\mathbf{P} = \mathbf{P}^{\mathrm{T}} = \mathbf{P}^2$$

Luego, $\mathbf{I} - \mathbf{P}$ también es simétrica es idempotente, es decir también es una matriz de proyección y proyecta sobre el ortogonal de $\mathcal{V}$.

Propiedades de $\mathbf{P}$

$\mathbf{P} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}} \in \mathbb{R}^{n\times n}$ es la matriz de **proyección sobre el espacio generado por las columnas de X**.

Suele llamarse a esta matriz $\mathbf{P}$ por ser de proyección $\mathbf{P}$ o $\mathbf{H}$ de hat matrix.

$\mathbf{P}$ es una matriz simétrica e idempotente, es decir:

$$\mathbf{P} = \mathbf{P}^{\mathrm{T}} = \mathbf{P}^2$$

Luego, $\mathbf{I} - \mathbf{P}$ también es simétrica es idempotente, es decir también es una matriz de proyección y proyecta sobre el ortogonal de $\mathcal{V}$.

Lema 2.

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

ii) $\mathrm{rango}(\mathbf{P}) = \mathrm{tr}(\mathbf{P}) = p$

$\mathrm{rango}(\mathbf{I} - \mathbf{P}) = \mathrm{tr}(\mathbf{I} - \mathbf{P}) = n - p$

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

Suma de Cuadrados

Tenemos que

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

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,

$$\|\mathbf{y} - \widehat{\mathbf{y}}\|^2 = \|\mathbf{y}\|^2 - \|\widehat{\mathbf{y}}\|^2 = \|\mathbf{y}\|^2 - \|\mathbf{P}\mathbf{y}\|^2$$

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,

$$\|\mathbf{y} - \widehat{\mathbf{y}}\|^2 = \|\mathbf{y}\|^2 - \|\widehat{\mathbf{y}}\|^2 = \|\mathbf{y}\|^2 - \|\mathbf{P}\mathbf{y}\|^2$$

Por lo tanto,

$$\|\mathbf{y}\|^2 = \|\mathbf{y} - \widehat{\mathbf{y}}\|^2 + \|\widehat{\mathbf{y}}\|^2$$

Propiedades del Estimador de Mínimos Cuadrados

Usando la notación matricial escribimos el modelo como

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

Sea $\mathbf{a} \in \mathbb{R}^p$ y $\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$.

Como rango$(\mathbf{X}) = p$, definimos el EMC de φ como

$$\widehat{\varphi} = \mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}$$

Estimación de σ^2: rango($\mathbf{X}$) = p

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

Estimación de σ^2: rango$(\mathbf{X}) = p$

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

Dado que $\sigma^2 = \mathbb{E}(\epsilon^2)$, parece natural estimarla mediante el promedio de los cuadrados de los residuos $\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$.

Estimación de σ^2: rango($\mathbf{X}$) $= p$

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

Dado que $\sigma^2 = \mathbb{E}(\epsilon^2)$, parece natural estimarla mediante el promedio de los cuadrados de los residuos $\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$.

El vector de residuos es

$$\mathbf{r} = \mathbf{r}(\widehat{\boldsymbol{\theta}}) = \mathbf{y} - \widehat{\mathbf{y}} = \mathbf{y} - \mathbf{P}\mathbf{y}\,,$$

Como $\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 = \|\mathbf{y} - \mathbf{P}\mathbf{y}\|^2$, podemos estimar a σ^2 por

$$s^2 = \frac{\|\mathbf{y} - \widehat{\mathbf{y}}\|^2}{n-p} = \frac{\|\mathbf{y} - \mathbf{P}\mathbf{y}\|^2}{n-p} = \frac{\mathbf{y}^{\mathrm{T}}(\mathbf{I} - \mathbf{P})\mathbf{y}}{n-p} = \frac{\mathcal{S}(\widehat{\boldsymbol{\theta}})}{n-p}$$

Propiedades

Proposición 3. Supongamos que rango($\mathbf{X}$) $= p$.

- Si se cumple H1, $\widehat{\boldsymbol{\theta}}$ es un estimador insesgado de $\boldsymbol{\theta}$ y $\widehat{\varphi} = \mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}$ es un estimador insesgado de $\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$.
- Si se cumplen H1 y H2,

$$s^2 = \frac{\mathcal{S}(\widehat{\boldsymbol{\theta}})}{n-p}$$

es un estimador insesgado de σ^2.

- Si se cumplen H1 y H2,

$$\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}} = \mathrm{VAR}(\widehat{\boldsymbol{\theta}}) = \sigma^2 \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}$$

$$\sigma^2_{\widehat{\varphi}} = \mathrm{VAR}(\widehat{\varphi}) = \sigma^2 \, \mathbf{a}^{\mathrm{T}} (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} \, \mathbf{a}$$

Propiedades

Proposición 3 (continuación).

- Si se cumplen H1 y H2,

$$\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}} = \mathrm{VAR}(\widehat{\boldsymbol{\theta}}) = \sigma^2 \, (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$$

$$\sigma^2_{\widehat{\varphi}} = \mathrm{VAR}(\widehat{\varphi}) = \sigma^2 \, \mathbf{a}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} \, \mathbf{a}$$

Además, podemos obtener estimadores insesgados de $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}$ y de $\sigma^2_{\widehat{\varphi}}$ como

$$\widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\theta}}} = \mathrm{VAR}(\widehat{\boldsymbol{\theta}}) = s^2 \, (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$$

$$\widehat{\sigma}^2_{\widehat{\varphi}} = s^2 \, \mathbf{a}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} \, \mathbf{a}$$

Propiedades

Sea

- $\mathbf{v}_1, \ldots, \mathbf{v}_p \in \mathbb{R}^n$ una base ortonormal de $\mathcal{V}$

Propiedades

Sea

- $\mathbf{v}_1, \ldots, \mathbf{v}_p \in \mathbb{R}^n$ una base ortonormal de $\mathcal{V}$
- $\mathbf{v}_{p+1}, \ldots, \mathbf{v}_n \in \mathbb{R}^n$ tales que $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ es b.o. de $\mathbb{R}^n$
- Definamos $\mathbf{B} = (\mathbf{v}_1, \ldots, \mathbf{v}_n) \qquad \mathbf{z} = \mathbf{B}^{\mathrm{T}}\mathbf{y}$.

Propiedades

Luego,

- $\mathcal{S}(\widehat{\boldsymbol{\theta}}) = \sum_{i=p+1}^{n} z_i^2$

Propiedades

Luego,

- $\mathcal{S}(\widehat{\boldsymbol{\theta}}) = \sum_{i=p+1}^{n} z_i^2$
- $\mathbb{E}(\mathbf{z}) = \boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p, 0, \ldots, 0)^{\mathrm{T}}$ con $\gamma_j = \mathbf{v}_j^{\mathrm{T}} \boldsymbol{\eta}$.

Propiedades

Luego,

- $\mathcal{S}(\widehat{\boldsymbol{\theta}}) = \sum_{i=p+1}^{n} z_i^2$
- $\mathbb{E}(\mathbf{z}) = \boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p, 0, \ldots, 0)^{\mathrm{T}}$ con $\gamma_j = \mathbf{v}_j^{\mathrm{T}} \boldsymbol{\eta}$.
- $\mathrm{VAR}(\mathbf{z}) = \sigma^2 \mathbf{I}$

Propiedades

Luego,

- $\mathcal{S}(\widehat{\boldsymbol{\theta}}) = \sum_{i=p+1}^{n} z_i^2$
- $\mathbb{E}(\mathbf{z}) = \boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p, 0, \ldots, 0)^{\mathrm{T}}$ con $\gamma_j = \mathbf{v}_j^{\mathrm{T}} \boldsymbol{\eta}$.
- $\textsc{Var}(\mathbf{z}) = \sigma^2 \mathbf{I}$
- Bajo H3, $\mathbf{z} \sim N\left(\boldsymbol{\gamma}, \sigma^2 \mathbf{I}\right)$, con lo cual
  - $z_1, \ldots, z_n$ son independientes y
  - $z_j \sim N(0, \sigma^2)$ si $j \geq p + 1$.

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:

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

Vimos que el estimador de máxima verosimilitud de $\boldsymbol{\theta}$ coincide con el de mínimos cuadrados.

Propiedad de EMC

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

Teorema 1. Supongamos que rango$(\mathbf{X}) = p$, $\boldsymbol{\theta} \in \mathbb{R}^p$. Entonces,

i) Bajo H3, $\widehat{\boldsymbol{\theta}}$ y s^2 son estadísticos suficientes y completos para $(\boldsymbol{\theta}, \sigma^2)$.

Por lo tanto, $\widehat{\boldsymbol{\theta}}$, $\widehat{\varphi} = \mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}$ y s^2 son estimadores IMVU para $\boldsymbol{\theta}$, $\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$ y σ^2, respectivamente.

Propiedad de EMC

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

Teorema 1. Supongamos que rango$(\mathbf{X}) = p$, $\boldsymbol{\theta} \in \mathbb{R}^p$. Entonces,

i) Bajo H3, $\widehat{\boldsymbol{\theta}}$ y s^2 son estadísticos suficientes y completos para $(\boldsymbol{\theta}, \sigma^2)$.

Por lo tanto, $\widehat{\boldsymbol{\theta}}$, $\widehat{\varphi} = \mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}$ y s^2 son estimadores IMVU para $\boldsymbol{\theta}$, $\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$ y σ^2, respectivamente.

ii) Bajo H1 y H2, $\widehat{\varphi} = \mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}$ es el estimador insesgado de menor varianza entre todos los estimadores lineales e insesgados de φ, o sea,

$$\textsc{Var}(\widehat{\varphi}) = \min_{\widetilde{\varphi} \in \mathcal{C}} \textsc{Var}(\widetilde{\varphi}),$$

donde $\mathcal{C} = \{\widetilde{\varphi} = \mathbf{c}^{\mathrm{T}}\mathbf{y}, \quad \mathbf{c} \in \mathbb{R}^n \quad \text{con} \mathbb{E}(\widetilde{\varphi}) = \varphi\}$

Propiedad de EMC

Teorema 2. Supongamos que rango$(\mathbf{X}) = p$, $\boldsymbol{\theta} \in \mathbb{R}^p$ y vale H3. Entonces,

i) $\widehat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1})$

Propiedad de EMC

Teorema 2. Supongamos que rango$(\mathbf{X}) = p$, $\boldsymbol{\theta} \in \mathbb{R}^p$ y vale H3. Entonces,

i) $\widehat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1})$

ii) $$\frac{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})}{\sigma^2} \sim \chi^2_p$$

Propiedad de EMC

Teorema 2. Supongamos que rango$(\mathbf{X}) = p$, $\boldsymbol{\theta} \in \mathbb{R}^p$ y vale H3. Entonces,

i) $\widehat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1})$

ii) $$\frac{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})}{\sigma^2} \sim \chi^2_p$$

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

Propiedad de EMC

Teorema 2. Supongamos que rango$(\mathbf{X}) = p$, $\boldsymbol{\theta} \in \mathbb{R}^p$ y vale H3. Entonces,

i) $\widehat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1})$

ii) $$\frac{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})}{\sigma^2} \sim \chi^2_p$$

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

iv) $\widehat{\boldsymbol{\theta}}$ y s^2 son independientes

v) $$\frac{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})}{p\, s^2} \sim \mathcal{F}_{p,\, n-p}$$

Propiedad de EMC

Teorema 3. Supongamos que rango($\mathbf{X}$) $= p$, $\boldsymbol{\theta} \in \mathbb{R}^p$ y vale H3. Entonces,

i) $\widehat{\varphi} = \mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}$ es un estimador IMVU para $\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$

ii) $\mathrm{VAR}(\widehat{\varphi}) = \sigma^2\mathbf{a}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{a} = \sigma^2_{\widehat{\varphi}}$

Propiedad de EMC

Teorema 3. Supongamos que rango($\mathbf{X}$) $= p$, $\boldsymbol{\theta} \in \mathbb{R}^p$ y vale H3. Entonces,

i) $\widehat{\varphi} = \mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}$ es un estimador IMVU para $\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$

ii) $\mathrm{VAR}(\widehat{\varphi}) = \sigma^2\mathbf{a}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{a} = \sigma^2_{\widehat{\varphi}}$

iii) $\widehat{\varphi} \sim N(\varphi, \sigma^2\mathbf{a}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{a}) = N(\varphi, \sigma^2_{\widehat{\varphi}})$

Propiedad de EMC

Teorema 3. Supongamos que rango($\mathbf{X}$) $= p$, $\boldsymbol{\theta} \in \mathbb{R}^p$ y vale H3. Entonces,

i) $\widehat{\varphi} = \mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}$ es un estimador IMVU para $\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$

ii) $\mathrm{VAR}(\widehat{\varphi}) = \sigma^2\mathbf{a}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{a} = \sigma^2_{\widehat{\varphi}}$

iii) $\widehat{\varphi} \sim N(\varphi, \sigma^2\mathbf{a}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{a}) = N(\varphi, \sigma^2_{\widehat{\varphi}})$

iv) Sea $\widehat{\sigma}^2_{\widehat{\varphi}} = s^2\,\mathbf{a}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{a}$ entonces

$$\frac{(\widehat{\varphi} - \varphi)}{\widehat{\sigma}_{\widehat{\varphi}}} \sim \mathcal{T}_{n-p}$$

Propiedad de EMC

Teorema 4. Supongamos que rango($\mathbf{X}$) $= p$, $\boldsymbol{\theta} \in \mathbb{R}^p$ y vale H3.

Sea $\mathbf{A} \in \mathbb{R}^{k \times p}$ con rango($\mathbf{A}$) $= k \leq p$ y definamos

$$\boldsymbol{\psi} = \mathbf{A}\boldsymbol{\theta} \qquad \widehat{\boldsymbol{\psi}} = \mathbf{A}\widehat{\boldsymbol{\theta}}$$

Entonces,

i) $\widehat{\boldsymbol{\psi}}$ es un estimador IMVU para $\boldsymbol{\psi}$

Propiedad de EMC

Teorema 4. Supongamos que rango($\mathbf{X}$) $= p$, $\boldsymbol{\theta} \in \mathbb{R}^p$ y vale H3.

Sea $\mathbf{A} \in \mathbb{R}^{k \times p}$ con rango($\mathbf{A}$) $= k \leq p$ y definamos

$$\boldsymbol{\psi} = \mathbf{A}\boldsymbol{\theta} \qquad \widehat{\boldsymbol{\psi}} = \mathbf{A}\widehat{\boldsymbol{\theta}}$$

Entonces,

i) $\widehat{\boldsymbol{\psi}}$ es un estimador IMVU para $\boldsymbol{\psi}$

ii) $\textsc{Var}(\widehat{\boldsymbol{\psi}}) = \sigma^2 \mathbf{A}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{A} = \boldsymbol{\Sigma}_{\widehat{\boldsymbol{\psi}}}$

Propiedad de EMC

Teorema 4. Supongamos que rango$(\mathbf{X}) = p$, $\boldsymbol{\theta} \in \mathbb{R}^p$ y vale H3.

Sea $\mathbf{A} \in \mathbb{R}^{k \times p}$ con rango$(\mathbf{A}) = k \leq p$ y definamos

$$\boldsymbol{\psi} = \mathbf{A}\boldsymbol{\theta} \qquad \widehat{\boldsymbol{\psi}} = \mathbf{A}\widehat{\boldsymbol{\theta}}$$

Entonces,

i) $\widehat{\boldsymbol{\psi}}$ es un estimador IMVU para $\boldsymbol{\psi}$

ii) $\text{VAR}(\widehat{\boldsymbol{\psi}}) = \sigma^2 \mathbf{A}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{A} = \boldsymbol{\Sigma}_{\widehat{\boldsymbol{\psi}}}$

iii) $\widehat{\boldsymbol{\psi}} \sim N\left(\boldsymbol{\psi}, \sigma^2 \mathbf{A}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{A}\right) = N\left(\boldsymbol{\psi}, \boldsymbol{\Sigma}_{\widehat{\boldsymbol{\psi}}}\right)$

Propiedad de EMC

Teorema 4. Supongamos que rango($\mathbf{X}$) $= p$, $\boldsymbol{\theta} \in \mathbb{R}^p$ y vale H3.

Sea $\mathbf{A} \in \mathbb{R}^{k \times p}$ con rango($\mathbf{A}$) $= k \leq p$ y definamos

$$\psi = \mathbf{A}\boldsymbol{\theta} \qquad \widehat{\psi} = \mathbf{A}\widehat{\boldsymbol{\theta}}$$

Entonces,

i) $\widehat{\psi}$ es un estimador IMVU para ψ

ii) $\text{VAR}(\widehat{\psi}) = \sigma^2 \mathbf{A}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{A} = \boldsymbol{\Sigma}_{\widehat{\psi}}$

iii) $\widehat{\psi} \sim N\left(\psi, \sigma^2 \mathbf{A}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{A}\right) = N\left(\psi, \boldsymbol{\Sigma}_{\widehat{\psi}}\right)$

iv) Sea $\widehat{\boldsymbol{\Sigma}}_{\widehat{\psi}} = s^2\, \mathbf{A}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{A}$ entonces

$$\frac{(\widehat{\psi} - \psi)^{\mathrm{T}}\, \widehat{\boldsymbol{\Sigma}}_{\widehat{\psi}}^{-1}\, (\widehat{\psi} - \psi)}{k} \sim \mathcal{F}_{k,n-p}$$

Test y Regiones de Confianza

Sea $\mathbf{A} \in \mathbb{R}^{k \times p}$ con rango$(\mathbf{A}) = k \leq p$ y

$$\psi = \mathbf{A}\theta \qquad \widehat{\psi} = \mathbf{A}\widehat{\theta}$$

Test y Regiones de Confianza

Sea $\mathbf{A} \in \mathbb{R}^{k \times p}$ con rango$(\mathbf{A}) = k \leq p$ y

$$\psi = \mathbf{A}\theta \qquad \widehat{\psi} = \mathbf{A}\widehat{\theta}$$

Del Teorema 4 deducimos

Corolario 1. Supongamos que rango$(\mathbf{X}) = p$, $\theta \in \mathbb{R}^p$ y vale H3. Entonces,

i) Una región de confianza de nivel $1 - \alpha$ para $\psi = \mathbf{A}\theta$ es

$$\{\psi \in \mathbb{R}^k : (\widehat{\psi} - \psi)^{\mathrm{T}} \, \widehat{\mathbf{\Sigma}}_{\widehat{\psi}}^{-1} \, (\widehat{\psi} - \psi) \leq k \, f_{k,n-p,\alpha}\}$$

Test y Regiones de Confianza

Sea $\mathbf{A} \in \mathbb{R}^{k \times p}$ con rango$(\mathbf{A}) = k \leq p$ y

$$\psi = \mathbf{A}\boldsymbol{\theta} \qquad \widehat{\psi} = \mathbf{A}\widehat{\boldsymbol{\theta}}$$

Del Teorema 4 deducimos

Corolario 1. Supongamos que rango$(\mathbf{X}) = p$, $\boldsymbol{\theta} \in \mathbb{R}^p$ y vale H3. Entonces,

i) Una región de confianza de nivel $1 - \alpha$ para $\psi = \mathbf{A}\boldsymbol{\theta}$ es

$$\{\psi \in \mathbb{R}^k : (\widehat{\psi} - \psi)^{\mathrm{T}} \, \widehat{\mathbf{\Sigma}}_{\widehat{\psi}}^{-1} \, (\widehat{\psi} - \psi) \leq k \, f_{k,n-p,\alpha}\}$$

ii) Un test de hipótesis para $H_0 : \psi = \psi_0$ versus $H_1 : \psi \neq \psi_0$ está dado por

$$\phi(\mathbf{y}) = \begin{cases} 1 & (\widehat{\psi} - \psi_0)^{\mathrm{T}} \, \widehat{\mathbf{\Sigma}}_{\widehat{\psi}}^{-1} \, (\widehat{\psi} - \psi_0) > k \, f_{k,n-p,\alpha} \\ 0 & (\widehat{\psi} - \psi_0)^{\mathrm{T}} \, \widehat{\mathbf{\Sigma}}_{\widehat{\psi}}^{-1} \, (\widehat{\psi} - \psi_0) \leq k \, f_{k,n-p,\alpha} \end{cases}$$

Test e Intervalos de Confianza

En particular, si $\mathbf{a} \in \mathbb{R}^p$ y

$$\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta} \qquad \widehat{\varphi} = \mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}$$

Test e Intervalos de Confianza

En particular, si $\mathbf{a} \in \mathbb{R}^p$ y

$$\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta} \qquad \widehat{\varphi} = \mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}$$

Del Teorema 3 (o del Corolario 1) deducimos

Corolario 2. Supongamos que rango$(\mathbf{X}) = p$, $\boldsymbol{\theta} \in \mathbb{R}^p$ y vale H3. Sea $\widehat{\sigma}^2_{\widehat{\varphi}} = s^2\, \mathbf{a}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{a}$ Entonces,

i) Un intervalo de confianza de nivel $1-\alpha$ para $\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$ es

$$\left[\widehat{\varphi} - t_{n-p,\frac{\alpha}{2}}\ \widehat{\sigma}_{\widehat{\varphi}}; \widehat{\varphi} + t_{n-p,\frac{\alpha}{2}}\ \widehat{\sigma}_{\widehat{\varphi}}\right]$$

Test e Intervalos de Confianza

En particular, si $\mathbf{a} \in \mathbb{R}^p$ y

$$\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta} \qquad \widehat{\varphi} = \mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}$$

Del Teorema 3 (o del Corolario 1) deducimos

Corolario 2. Supongamos que rango$(\mathbf{X}) = p$, $\boldsymbol{\theta} \in \mathbb{R}^p$ y vale H3. Sea $\widehat{\sigma}^2_{\widehat{\varphi}} = s^2\, \mathbf{a}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{a}$ Entonces,

i) Un intervalo de confianza de nivel $1-\alpha$ para $\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$ es

$$\left[\widehat{\varphi} - t_{n-p,\frac{\alpha}{2}}\ \widehat{\sigma}_{\widehat{\varphi}}; \widehat{\varphi} + t_{n-p,\frac{\alpha}{2}}\ \widehat{\sigma}_{\widehat{\varphi}}\right]$$

ii) Un test de hipótesis para $H_0 : \varphi = \varphi_0$ versus $H_1 : \varphi \neq \varphi_0$ está dado

$$\phi(\mathbf{y}) = \begin{cases} 1 & \dfrac{|\widehat{\varphi} - \varphi_0|}{\widehat{\sigma}_{\widehat{\varphi}}} > t_{n-p,\frac{\alpha}{2}} \\ 0 & \dfrac{|\widehat{\varphi} - \varphi_0|}{\widehat{\sigma}_{\widehat{\varphi}}} \leq t_{n-p,\frac{\alpha}{2}} \end{cases}$$

Test e Intervalos de Confianza para θ_i

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

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

$$\widehat{\theta}_i = \mathbf{e}_i^{\mathrm{T}}\widehat{\boldsymbol{\theta}} \sim N(\theta_i, \sigma^2\mathbf{e}_i^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{e}_i)$$

Si llamamos $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}} = \sigma^2\mathbf{D}$

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

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

Test e Intervalos de Confianza para θ_i

Si para un i fijo quisiéramos testear

$$H_0 : \theta_i = 0 \quad \text{vs.} \quad H_1 : \theta_i \neq 0$$

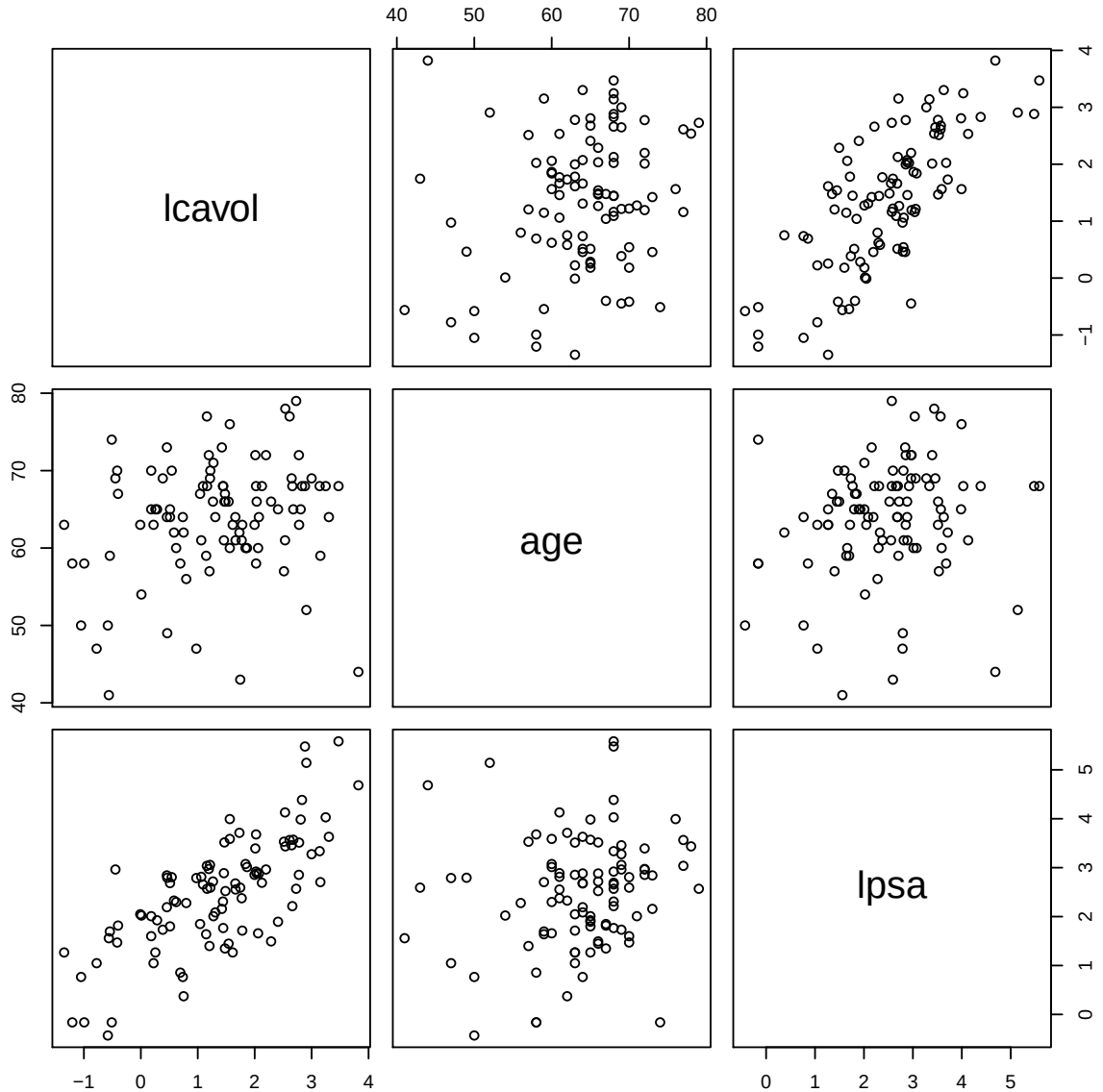
tenemos que bajo H_0

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

Por lo tanto, rechazaremos H_0 con nivel α si

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

Ejemplo



$$lcavol_i = \theta_0 + \theta_1\, age_i + \theta_2\, lpsa_i + \varepsilon_i$$

```
salida.2< −lm(lcavol ~age+lpsa)
summary(salida.2)
```

$$lcavol_i = \theta_0 + \theta_1\, age_i + \theta_2\, lpsa_i + \varepsilon_i$$

```
salida.2< −lm(lcavol ~age+lpsa)
summary(salida.2)
```

```
Call:
lm(formula = lcavol ~age + lpsa)  Residuals:
     Min       1Q   Median       3Q      Max
-2.23487 -0.62467  0.02114  0.54422  1.71757

Coefficients:
             Estimate  Std. Error  t value  Pr(>|t|)
(Intercept) -1.50978   0.70670    -2.136   0.0352     *
age          0.01637   0.01112     1.473   0.1442
lpsa         0.73201   0.07170    10.210   <2e-16     ***

Residual standard error: 0.7992 on 94 degrees of freedom
Multiple R-squared: 0.5498, Adjusted R-squared: 0.5402
F-statistic: 57.4 on 2 and 94 DF, p-value: < 2.2e-16
```

Ejemplo

Vamos a reproducir el cálculo del estadístico F reportado en el summary en el que se testea la significación de la regresión, o sea

$$H_0 : \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Ejemplo

Vamos a reproducir el cálculo del estadístico F reportado en el summary en el que se testea la significación de la regresión, o sea

$$H_0 : \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

El vector reducido (es decir sin la primera componente) $\boldsymbol{\theta}^* = (\theta_1, \theta_2)^{\mathrm{T}}$ es el resultado de multiplicar a $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)^{\mathrm{T}}$ por la matriz

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Ejemplo

Usando las propiedades de vectores aleatorios y de la distribución normal tenemos que

$$\widehat{\boldsymbol{\theta}}^* = \begin{pmatrix} \widehat{\theta}_1 \\ \widehat{\theta}_2 \end{pmatrix} \sim N_2\left(\boldsymbol{\theta}^*, \sigma^2 \, \mathbf{A} \, \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1} \, \mathbf{A}^{\mathrm{T}}\right)$$

Ejemplo

Usando las propiedades de vectores aleatorios y de la distribución normal tenemos que

$$\widehat{\boldsymbol{\theta}}^* = \begin{pmatrix} \widehat{\theta}_1 \\ \widehat{\theta}_2 \end{pmatrix} \sim N_2\left(\boldsymbol{\theta}^*, \sigma^2 \, \mathbf{A} \, \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1} \, \mathbf{A}^{\mathrm{T}}\right)$$

que se puede estimar mediante los siguientes comandos

Ejemplo

```
tita.est <- salida.2$coefficients

ese<- summary(salida.2)$sigma

XtX = t(model.matrix(salida.2))%*%model.matrix(salida.2)
XtX.inv = solve(XtX)
XtX.inv
```

Ejemplo

tita.est $<-$ salida.2$coefficients

ese$<-$ summary(salida.2)$sigma

XtX = t(model.matrix(salida.2))%*%model.matrix(salida.2)
XtX.inv = solve(XtX)
XtX.inv

	(Intercept)	age	lpsa
(Intercept)	0.781970977	-0.0118329283	-0.0064316618
age	-0.011832928	0.0001934906	-0.0002116456
lpsa	-0.006431662	-0.0002116456	0.0080490331

Ejemplo

```
tita.est <- salida.2$coefficients

ese<- summary(salida.2)$sigma

XtX = t(model.matrix(salida.2))%*%model.matrix(salida.2)
XtX.inv = solve(XtX)
XtX.inv
```

	(Intercept)	age	lpsa
(Intercept)	0.781970977	-0.0118329283	-0.0064316618
age	-0.011832928	0.0001934906	-0.0002116456
lpsa	-0.006431662	-0.0002116456	0.0080490331

```
A<-matrix(rep(0,6),ncol = 3,nrow = 2)
A[1,2]=A[2,3]=1
A
```

	[,1]	[,2]	[,3]
[1,]	0	1	0
[2,]	0	0	1

Ejemplo

Calculemos $\widehat{\boldsymbol{\theta}}^*$

```
theta.star=A%*%tita.est
theta.star
```

Ejemplo

\# Calculemos $\widehat{\boldsymbol{\theta}}^*$

```
theta.star=A%*%tita.est
theta.star
```

```
             [,1]
[1,]  0.01637148
[2,]  0.73201132
```

Ejemplo

Calculemos $\widehat{\boldsymbol{\theta}}^*$

```
theta.star=A%*%tita.est
theta.star
```

```
              [,1]
[1,]  0.01637148
[2,]  0.73201132
```

```
CovMat¡- (ese∧2)*A%*%solve(XtX)%*%t(A)
CovMat
```

Ejemplo

Calculemos $\widehat{\boldsymbol{\theta}}^*$

```
theta.star=A%*%tita.est
theta.star
```

```
                [,1]
[1,]    0.01637148
[2,]    0.73201132
```

```
CovMat¡- (ese∧2)*A%*%solve(XtX)%*%t(A)
CovMat
```

```
                 [,1]           [,2]
[1,]    0.0001235781   -0.0001351733
[2,]   -0.0001351733    0.0051407373
```

Ejemplo

Dado que

$$V_1 = \frac{(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)^{\mathrm{T}} \left(\mathbf{A} \left(\mathbf{X}^{\mathrm{T}} \mathbf{X} \right)^{-1} \mathbf{A}^{\mathrm{T}} \right)^{-1} (\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)}{\sigma^2} \sim \chi^2_2$$

Ejemplo

Dado que

$$V_1 = \frac{(\widehat{\theta}^* - \theta^*)^{\mathrm{T}} \left(\mathbf{A} \left(\mathbf{X}^{\mathrm{T}} \mathbf{X} \right)^{-1} \mathbf{A}^{\mathrm{T}} \right)^{-1} (\widehat{\theta}^* - \theta^*)}{\sigma^2} \sim \chi^2_2$$

es independiente de

$$V_2 = \frac{(n-3)s^2}{\sigma^2} \sim \chi^2_{n-3}$$

Ejemplo

Dado que

$$V_1 = \frac{(\widehat{\theta}^* - \theta^*)^{\mathrm{T}} \left(\mathbf{A} \left(\mathbf{X}^{\mathrm{T}} \mathbf{X} \right)^{-1} \mathbf{A}^{\mathrm{T}} \right)^{-1} (\widehat{\theta}^* - \theta^*)}{\sigma^2} \sim \chi^2_2$$

es independiente de

$$V_2 = \frac{(n-3)s^2}{\sigma^2} \sim \chi^2_{n-3}$$

resulta que el cociente $(V_1/2)/(V_2/(n-3))$ tiene distribución $F_{2,n-3}$

$$\frac{(\widehat{\theta}^* - \theta^*)^{\mathrm{T}} \left(\mathbf{A} \left(\mathbf{X}^{\mathrm{T}} \mathbf{X} \right)^{-1} \mathbf{A}^{\mathrm{T}} \right)^{-1} (\widehat{\theta}^* - \theta^*)}{2\, s^2} \sim F_{2,n-3}$$

Ejemplo

Dado que

$$V_1 = \frac{(\widehat{\theta}^* - \theta^*)^{\mathrm{T}} \left(\mathbf{A} \left(\mathbf{X}^{\mathrm{T}} \mathbf{X} \right)^{-1} \mathbf{A}^{\mathrm{T}} \right)^{-1} (\widehat{\theta}^* - \theta^*)}{\sigma^2} \sim \chi^2_2$$

es independiente de

$$V_2 = \frac{(n-3)s^2}{\sigma^2} \sim \chi^2_{n-3}$$

resulta que el cociente $(V_1/2)/(V_2/(n-3))$ tiene distribución $F_{2,n-3}$

$$\frac{(\widehat{\theta}^* - \theta^*)^{\mathrm{T}} \left(\mathbf{A} \left(\mathbf{X}^{\mathrm{T}} \mathbf{X} \right)^{-1} \mathbf{A}^{\mathrm{T}} \right)^{-1} (\widehat{\theta}^* - \theta^*)}{2\, s^2} \sim F_{2,n-3}$$

Lo computamos como

Ejemplo

```
t(theta.star)%*%(solve(CovMat))%*%(theta.star)/2
```

```
          [,1]
[1,] 57.40246
```

Ejemplo

```
t(theta.star)%*%(solve(CovMat))%*%(theta.star)/2
```

```
          [,1]
[1,] 57.40246
```

que coincide con el valor reportado en el summary en el F-statistic con 2 y 94 grados de libertad.

Funciones Lineales de los Parámetros

En muchas aplicaciones estamos más interesados en estimar funciones lineales de $\boldsymbol{\theta}$ que en estimar $\boldsymbol{\theta}$ en sí mismo:

$$\mathbf{a}^{\mathrm{T}}\boldsymbol{\theta} \quad \text{o} \quad \mathbf{A}\boldsymbol{\theta}$$

Funciones Lineales de los Parámetros

En muchas aplicaciones estamos más interesados en estimar funciones lineales de $\boldsymbol{\theta}$ que en estimar $\boldsymbol{\theta}$ en sí mismo:

$$\mathbf{a}^{\mathrm{T}}\boldsymbol{\theta} \quad \text{o} \quad \mathbf{A}\boldsymbol{\theta}$$

Un caso de especial interés:

En un futuro valor $\mathbf{x}_0$, ¿qué esperamos observar? $\mathbb{E}(y_0) =?$

Funciones Lineales de los Parámetros

En muchas aplicaciones estamos más interesados en estimar funciones lineales de $\boldsymbol{\theta}$ que en estimar $\boldsymbol{\theta}$ en sí mismo:

$$\mathbf{a}^{\mathrm{T}}\boldsymbol{\theta} \quad \text{o} \quad \mathbf{A}\boldsymbol{\theta}$$

Un caso de especial interés:

En un futuro valor $\mathbf{x}_0$, ¿qué esperamos observar? $\mathbb{E}(y_0) =$?

$$\mathbb{E}(y_0) = \mathbf{x}_0^{\mathrm{T}}\boldsymbol{\theta} \Rightarrow \widehat{\mathbb{E}(y_0)} = \mathbf{x}_0^{\mathrm{T}}\widehat{\boldsymbol{\theta}}$$

Observemos que si $\widehat{\boldsymbol{\theta}}$ es insesgado entonces el estimador propuesto sería insesgado de $\mathbb{E}(y_0)$.

Funciones Lineales de los Parámetros

En muchas aplicaciones estamos más interesados en estimar funciones lineales de $\boldsymbol{\theta}$ que en estimar $\boldsymbol{\theta}$ en sí mismo:

$$\mathbf{a}^{\mathrm{T}}\boldsymbol{\theta} \quad \text{o} \quad \mathbf{A}\boldsymbol{\theta}$$

Un caso de especial interés:

En un futuro valor $\mathbf{x}_0$, ¿qué esperamos observar? $\mathbb{E}(y_0) =?$

$$\mathbb{E}(y_0) = \mathbf{x}_0^{\mathrm{T}}\boldsymbol{\theta} \Rightarrow \widehat{\mathbb{E}(y_0)} = \mathbf{x}_0^{\mathrm{T}}\widehat{\boldsymbol{\theta}}$$

Observemos que si $\widehat{\boldsymbol{\theta}}$ es insesgado entonces el estimador propuesto sería insesgado de $\mathbb{E}(y_0)$.

Podría haber muchos estimadores de una función lineal $\mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$ o $\mathbf{A}\boldsymbol{\theta}$, nos enfocaremos en los estimadores lineales: funciones lineales de $y_1, \ldots, y_n$.

Funciones Estimables

Definición 1. Una función paramétrica φ se dice que es una función lineal de los paramétros $\{\theta_1, \ldots, \theta_p\}$ si existen $\{a_1, \ldots, a_p\}$ constantes conocidas tal que

$$\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta} = \sum_{j=1}^{p} a_j \theta_j$$

donde $\mathbf{a} = (a_1, \ldots, a_p)^{\mathrm{T}}$.

Funciones Estimables

Definición 1. Una función paramétrica φ se dice que es una función lineal de los paramétros $\{\theta_1, \ldots, \theta_p\}$ si existen $\{a_1, \ldots, a_p\}$ constantes conocidas tal que

$$\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta} = \sum_{j=1}^{p} a_j \theta_j$$

donde $\mathbf{a} = (a_1, \ldots, a_p)^{\mathrm{T}}$.

Definición 2. Decimos que una función paramétrica $\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$ es estimable si tiene un estimador lineal (en $\mathbf{y}$) insesgado, es decir si existe $\mathbf{c} \in \mathbb{R}^n$ tal que

$$\mathbb{E}(\mathbf{c}^{\mathrm{T}}\mathbf{y}) = \varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta} \quad \forall \boldsymbol{\theta} \in \mathbb{R}^p$$

Funciones Estimables

Definición 1. Una función paramétrica φ se dice que es una función lineal de los paramétros $\{\theta_1, \ldots, \theta_p\}$ si existen $\{a_1, \ldots, a_p\}$ constantes conocidas tal que

$$\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta} = \sum_{j=1}^{p} a_j \theta_j$$

donde $\mathbf{a} = (a_1, \ldots, a_p)^{\mathrm{T}}$.

Definición 2. Decimos que una función paramétrica $\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$ es estimable si tiene un estimador lineal (en $\mathbf{y}$) insesgado, es decir si existe $\mathbf{c} \in \mathbb{R}^n$ tal que

$$\mathbb{E}(\mathbf{c}^{\mathrm{T}}\mathbf{y}) = \varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta} \quad \forall \boldsymbol{\theta} \in \mathbb{R}^p$$

¿Hay funciones que no son estimables?

Funciones Estimables: Caracterización

El siguiente resultado caracteriza las funciones paramétricas estimables suponiendo el modelo

$$\Omega : \mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\theta} \qquad \boldsymbol{\Sigma}_{\mathbf{y}} = \sigma^2 \mathbf{I}$$

Teorema 5. La función paramétrica $\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$ es estimable si y sólo si $\mathbf{a}$ es una combinación lineal de las filas de $\mathbf{X}$, o sea si existe $\mathbf{c} \in \mathbb{R}^n$ tal que

$$\mathbf{a} = \mathbf{X}^{\mathrm{T}}\,\mathbf{c}$$

Funciones Estimables: Caracterización

El siguiente resultado caracteriza las funciones paramétricas estimables suponiendo el modelo

$$\Omega : \mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\theta} \qquad \boldsymbol{\Sigma}_{\mathbf{y}} = \sigma^2 \mathbf{I}$$

Teorema 5. La función paramétrica $\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$ es estimable si y sólo si $\mathbf{a}$ es una combinación lineal de las filas de $\mathbf{X}$, o sea si existe $\mathbf{c} \in \mathbb{R}^n$ tal que

$$\mathbf{a} = \mathbf{X}^{\mathrm{T}}\,\mathbf{c}$$

Notemos que

- Si rango$(\mathbf{X}) = p$, toda función $\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$ es estimable.

Funciones Estimables: Caracterización

El siguiente resultado caracteriza las funciones paramétricas estimables suponiendo el modelo

$$\Omega : \mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\theta} \qquad \boldsymbol{\Sigma}_{\mathbf{y}} = \sigma^2 \mathbf{I}$$

Teorema 5. La función paramétrica $\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$ es estimable si y sólo si $\mathbf{a}$ es una combinación lineal de las filas de $\mathbf{X}$, o sea si existe $\mathbf{c} \in \mathbb{R}^n$ tal que

$$\mathbf{a} = \mathbf{X}^{\mathrm{T}}\,\mathbf{c}$$

Notemos que

- Si rango$(\mathbf{X}) = p$, toda función $\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$ es estimable.

- Si rango$(\mathbf{X}) = q < p$, y $\boldsymbol{\eta} = \mathbb{E}(\mathbf{y})$, las funciones estimables son combinaciones lineales de $\boldsymbol{\eta}$.

Ejemplo

Supongamos que

$$y_{i,j} = \mu + \alpha_j + \epsilon_{ij} \qquad 1 \leq j \leq 2, \quad 1 \leq i \leq n_j$$

entonces

$$p = 3 \qquad \mathbf{y} = \begin{pmatrix} y_{1,1} \\ y_{2,1} \\ \vdots \\ y_{n_1,1} \\ y_{1,2} \\ y_{2,2} \\ \vdots \\ y_{n_2,2} \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{pmatrix} \qquad \boldsymbol{\theta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix}$$

Ejemplo

Puedo estimar las combinaciones lineales de las componentes de

$$\boldsymbol{\eta} = \begin{pmatrix} \mu + \alpha_1 \\ \vdots \\ \mu + \alpha_1 \\ \mu + \alpha_2 \\ \vdots \\ \mu + \alpha_2 \end{pmatrix}$$

es decir, puedo estimar $\mu + \alpha_1$ y $\alpha_2 - \alpha_1$, pero no puedo estimar ni μ ni α_1.

Mejor Estimador Lineal Insesgado (BLUE)

Lema 3. Supongamos que vale el modelo Ω. Sean

- $\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$ una función estimable
- $\mathcal{V}$ el espacio generado por las columnas de $\mathbf{X}$ ($q = \text{rango}(\mathbf{X}) \leq p$).

Luego,

- existe un único $\mathbf{c}_0 \in \mathcal{V}$ tal que el estimador lineal

$$\widehat{\varphi} = \mathbf{c}_0^{\mathrm{T}}\mathbf{y}$$

es un estimador insesgado de φ.

Mejor Estimador Lineal Insesgado (BLUE)

Lema 3. Supongamos que vale el modelo Ω. Sean

- $\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$ una función estimable
- $\mathcal{V}$ el espacio generado por las columnas de $\mathbf{X}$ ($q = \text{rango}(\mathbf{X}) \leq p$).

Luego,

- existe un único $\mathbf{c}_0 \in \mathcal{V}$ tal que el estimador lineal

$$\widehat{\varphi} = \mathbf{c}_0^{\mathrm{T}}\mathbf{y}$$

es un estimador insesgado de φ.

- Más aún, si $\mathbf{c}^{\mathrm{T}}\mathbf{y}$ es un estimador insesgado de φ, $\mathbf{c}_0$ es la proyección ortogonal de $\mathbf{c}$ sobre $\mathcal{V}$ y

$$\textsc{Var}(\mathbf{c}_0^{\mathrm{T}}\mathbf{y}) \leq \textsc{Var}(\mathbf{c}^{\mathrm{T}}\mathbf{y})$$

y vale el igual sólo si $\mathbf{c} = \mathbf{c}_0$.

Teorema de Gauss–Markov

Teorema de Gauss–Markov

Supongamos que vale el modelo $\Omega : \mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\theta} \qquad \boldsymbol{\Sigma}_{\mathbf{y}} = \sigma^2\mathbf{I}$.

Toda función estimable $\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$ tiene un único estimador $\widehat{\varphi}$ lineal insesgado de mínima varianza (BLUE).

Este estimador $\widehat{\varphi}$ se puede obtener reemplazando a $\boldsymbol{\theta}$ en $\mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$ por $\widehat{\boldsymbol{\theta}}$, siendo $\widehat{\boldsymbol{\theta}}$ un estimador de mínimos cuadrados, es decir, tal que

$$\mathbf{X}^{\mathrm{T}}\mathbf{X}\widehat{\boldsymbol{\theta}} = \mathbf{X}^{\mathrm{T}}\mathbf{y}\,.$$

Definición 3. Para cada función estimable $\varphi = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$, llamaremos EMC de φ al único estimador lineal de mínima varianza, o sea, a

$$\widehat{\varphi} = \mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}$$

Notación

El modelo se escribe como

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

Notación

El modelo se escribe como

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

$$\text{rango}(\mathbf{X}) = p$$

Notación

El modelo se escribe como

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

$$\text{rango}(\mathbf{X}) = p$$

$$\widehat{\boldsymbol{\theta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

Notación

El modelo se escribe como

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

$$\text{rango}(\mathbf{X}) = p$$

$$\widehat{\boldsymbol{\theta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

Dada una matriz $\mathbf{A} \in \mathbb{R}^{p\times p}$ llamaremos

$$\lambda_1(\mathbf{A}) \leq \lambda_2(\mathbf{A}) \leq \cdots \leq \lambda_p(\mathbf{A})$$

a los autovalores de $\mathbf{A}$

Consistencia

a) **Si** $\lambda_1(\mathbf{X}^{\mathrm{T}}\mathbf{X}) \to \infty$ **entonces** $\widehat{\theta} \xrightarrow{p} \theta$.

b) **Supongamos que se cumple H3, entonces la condición** $\lambda_1(\mathbf{X}^{\mathrm{T}}\mathbf{X}) \to \infty$ **también es necesaria.**

Ejemplos

a) Regresión por el origen: $y_i = \theta_1 x_i + \epsilon_i \quad 1 \leq i \leq n$

$$\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad \mathbf{X}^{\mathrm{T}}\mathbf{X} = \sum_{i=1}^{n} x_i^2$$

La condición $\lambda_1(\mathbf{X}^{\mathrm{T}}\mathbf{X}) \to \infty$ dice que $\sum_{i=1}^{n} x_i^2 \to \infty$

Ejemplos

b) Regresión simple: $y_i = \theta_1 + \theta_2\, x_i + \epsilon_i \quad 1 \leq i \leq n$

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

Ejemplos

b) Regresión simple: $y_i = \theta_1 + \theta_2\, x_i + \epsilon_i \quad 1 \leq i \leq n$

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

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

La condición $\lambda_1(\mathbf{X}^{\mathrm{T}}\mathbf{X}) \to \infty$ es equivalente a $\sum_{i=1}^{n}(x_i - \overline{x})^2 \to 0$.

Distribución asintótica

Bajo H3,

$$\widehat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1})$$

Luego si $\mathbf{S} \in \mathbb{R}^{p\times p}$ es tal que

$$\mathbf{S}\mathbf{S}^{\mathrm{T}} = \mathbf{X}^{\mathrm{T}}\mathbf{X}$$

Tenemos que

$$\mathbf{S}^{\mathrm{T}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim N(\mathbf{0}, \sigma^2\mathbf{I}_p)$$

Distribución asintótica

$$\mathbf{z}_i = \mathbf{S}^{-1}\mathbf{x}_i \qquad \boldsymbol{\gamma} = \mathbf{S}^{\mathrm{T}}\boldsymbol{\theta}$$

Distribución asintótica

$$\mathbf{z}_i = \mathbf{S}^{-1}\mathbf{x}_i \qquad \boldsymbol{\gamma} = \mathbf{S}^{\mathrm{T}}\boldsymbol{\theta}$$

El modelo se reescribe

$$y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta} + \epsilon_i = (\mathbf{S}\,\mathbf{z}_i)^{\mathrm{T}}(\mathbf{S}^{\mathrm{T}})^{-1}\boldsymbol{\gamma} + \epsilon_i = \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\gamma} + \epsilon_i$$

El EMC de $\boldsymbol{\gamma}$ es

$$\widehat{\boldsymbol{\gamma}} = \mathbf{S}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}$$

Basta ver que

$$\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \xrightarrow{D} N(\mathbf{0}, \sigma^2\mathbf{I}_p)$$

Distribución asintótica

Propiedad. Bajo H1 y H2,

$$\frac{\widehat{\gamma} - \gamma}{p} = O_{\mathbb{P}}(1)$$

Distribución asintótica

Propiedad. Bajo H1 y H2,

$$\frac{\widehat{\gamma} - \gamma}{p} = O_{\mathbb{P}}(1)$$

Lema. Sean $\{V_j\}_{j \geq 1}$ i.i.d $\mathbb{E}(V_j) = 0$, $\text{VAR}(V_j) = \sigma^2$. Definamos

$$W_n = \sum_{j=1}^{n} a_{j,n} V_j \text{ donde } \sum_{j=1}^{n} a_{j,n}^2 = 1$$

Si

$$\max_{1 \leq j \leq n} a_{j,n}^2 \to 0$$

entonces $W_n \xrightarrow{D} N(0, \sigma^2)$.

Distribución asintótica

Supongamos que

$$\max_{1\leq j\leq n} \mathbf{x}_j^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{x}_j \to 0$$

entonces

$$\mathbf{S}^{\mathrm{T}}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) = \widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma} \xrightarrow{D} N(\mathbf{0},\sigma^2\mathbf{I}_p)$$

Observación

Observemos que

$$\mathbf{x}_j^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{x}_j$$

es el $j-$ésimo elemento diagonal de la matriz de proyección

$$\mathbf{P} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}$$