

# Modelos no paramétricos y de regresión

## Regresión lineal múltiple

Javier Santibáñez

Facultad de Ciencias, UNAM

`jsantibanez@sigma.iimas.unam.mx`

Semestre 2019-1

# Contenido

- 1 Planteamiento
- 2 Supuestos del modelo RLM
- 3 Estimación puntual
- 4 Inferencias para  $\sigma^2$
- 5 Inferencias para  $\beta$
- 6 Análisis de varianza
- 7 Ajuste del modelo

# Planteamiento

- El modelo de regresión lineal múltiple (RLM) es una extensión del modelo RLS ya que se modela una variable aleatoria  $Y$  condicional a un conjunto de variables auxiliares  $X_1, \dots, X_n$ , mismas que se asumen no aleatorias.
- El modelo RLM queda especificado por las siguientes ecuaciones

$$E(Y | X_1 = x_1, \dots, X_n = x_n) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p,$$

$$V(Y | X_1 = x_1, \dots, X_n = x_n) = \sigma^2,$$

donde  $\beta_0, \beta_1, \dots, \beta_p \in \mathbb{R}$  y  $\sigma \in \mathbb{R}^+$  son los parámetros del modelo y  $x_i$  representa un valor particular de la variable  $X_i$ .

- El objetivo es estimar  $\beta_0, \beta_1, \dots, \beta_p$  y  $\sigma^2$  a partir de un conjunto de observaciones de la población

$$(x_{11}, \dots, x_{1p}, Y_1), \dots, (x_{n1}, \dots, x_{np}, Y_n).$$

# Planteamiento

- Es útil representar el modelo anterior en forma matricial, ya que esto simplifica la notación y facilita la manipulación de las expresiones al realizar las inferencias.
- Se denota por  $\boldsymbol{\beta}$  el vector de coeficientes del modelo,  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$ .
- Para simplificar la notación, se denotará por  $\mathbf{x}_i$  al vector  $(1, x_{i1}, \dots, x_{ip})$ , donde  $x_{ij}$  denota la medición de la  $j$ -ésima variable del  $i$ -ésimo elemento de la muestra,  $i = 1, \dots, n$  e  $j = 1, \dots, p$ .
- De esta forma

$$E(Y_i | \mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta}, \quad i = 1, \dots, n,$$

donde condicionar con respecto a  $\mathbf{x}_i$  es para simplificar que se condiciona con respecto a  $X_1 = x_{i1}, \dots, X_p = x_{ip}$ .

- Al igual que en el modelo simple, es útil representar a  $Y_i$  como

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \dots, n.$$

donde  $\epsilon_i$  es un error aleatorio que representa la variación que tiene  $Y_i$  alrededor de su valor esperado.

- De la representación anterior se sigue que

$$E(\epsilon_i | \mathbf{x}_i) = 0 \quad \text{y} \quad V(\epsilon_i | \mathbf{x}_i) = \sigma^2. \quad i = 1, \dots, n.$$

# Supuestos del modelo RLM

Dado el conjunto de observaciones observaciones

$$(\mathbf{x}_1, Y_1), (\mathbf{x}_2, Y_2), \dots, (\mathbf{x}_n, Y_n),$$

los primeros tres supuestos del modelo RLM son:

## 1. Linealidad

$$E(Y_i | \mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta}, \quad i = 1, \dots, n.$$

## 2. Homocedasticidad (varianza constante)

$$V(Y_i | \mathbf{x}_i) = \sigma^2, \quad i = 1, \dots, n.$$

## 3. No correlación

$$\text{Cov}(Y_i, Y_j | \mathbf{x}_i, \mathbf{x}_j) = 0, \quad i, j = 1, \dots, n \text{ e } i \neq j.$$

# Supuestos del modelo RLM

Los supuestos 2 y 3 se pueden expresar en términos de los errores como sigue

2'. Homocedasticidad (varianza constante).

$$V(\epsilon_i | \mathbf{x}_i) = \sigma^2, \quad i = 1, \dots, n.$$

3'. No correlación.

$$\text{Cov}(\epsilon_i, \epsilon_j | \mathbf{x}_i, \mathbf{x}_j) = 0, \quad i, j = 1, \dots, n \text{ e } i \neq j.$$

El supuesto de no correlación implica que la forma en que dos observaciones cualquiera  $Y_i$  y  $Y_j$  varían alrededor de su valor esperado, no tiene ninguna asociación lineal.

- Para facilitar la resolución del problema de estimar  $\beta$  y  $\sigma^2$  se definen

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \quad \text{y} \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

- La matriz  $\mathbf{X}$  recibe el nombre de matriz de diseño.
- Con lo cual el modelo queda expresado como  $\mathbf{Y} | \mathbf{X} \sim F(\mathbf{X}\beta, \sigma^2\mathbf{I}_n)$ , esto es,  $E(\mathbf{Y} | \mathbf{X}) = \mathbf{X}\beta$  y  $V(\mathbf{Y} | \mathbf{X}) = \sigma^2\mathbf{I}_n$ .
- Si se define  $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$ , entonces el modelo queda representado como  $\mathbf{Y} = \mathbf{X}\beta + \epsilon$  con  $\epsilon | \mathbf{X} \sim F(\mathbf{0}_n, \sigma^2\mathbf{I})$ .

- Si  $\tilde{\beta}$  es un estimador de  $\beta$ , se definen los correspondientes vectores de valores ajustados y de residuos son, respectivamente

$$\tilde{\mathbf{Y}} = \mathbf{X}\tilde{\beta} \quad \text{y} \quad \tilde{\epsilon} = \mathbf{Y} - \tilde{\mathbf{Y}},$$

respectivamente.

- Si  $\tilde{\mathbf{B}}$  es una realización de  $\tilde{\beta}$ , calculada con una realización  $\mathbf{y}$  de  $\mathbf{Y}$ , entonces las correspondientes realizaciones de  $\tilde{\mathbf{Y}}$  y  $\tilde{\epsilon}$  son

$$\tilde{\mathbf{y}} = \mathbf{X}\tilde{\mathbf{B}} \quad \text{y} \quad \tilde{\epsilon} = \mathbf{y} - \tilde{\mathbf{y}},$$

respectivamente.

# Estimación por Mínimos Cuadrados (MCO)

Entonces, la función suma de cuadrados de los errores se define como

$$Q(\beta) = \mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$

El estimador de MCO de  $\beta$  es el vector  $\hat{\beta}$  tal que  $Q(\hat{\beta})$  es mínima.

Primero se debe notar que

$$Q(\beta) = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta$$

Entonces:

$$\nabla Q(\beta) = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\beta = 0 \quad \Rightarrow \quad \mathbf{X}'\mathbf{X}\beta = \mathbf{X}'\mathbf{y}.$$

Las ecuaciones del sistema (expresado en forma matricial)

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$$

reciben el nombre de ecuaciones normales. Si la matriz de diseño  $\mathbf{X}$  es de rango completo (por columnas), entonces  $\mathbf{X}'\mathbf{X}$  es de rango completo y por lo tanto, invertible. Entonces, la solución al sistema de ecuaciones normales es

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}.$$

## Estimación por MCO (cont.)

Para verificar que  $Q(\boldsymbol{\beta})$  tiene un mínimo en  $\hat{\boldsymbol{\beta}}$  utilizamos el criterio de las segundas derivadas parciales. Es sencillo mostrar que

$$H_Q(\boldsymbol{\beta}) = 2\mathbf{X}'\mathbf{X}$$

que no depende de  $\boldsymbol{\beta}$ . Si  $\mathbf{v}$  es un vector de dimensión  $p + 1$  diferente de  $\mathbf{0}$ , y definimos  $\mathbf{u} = \mathbf{X}\mathbf{v}$ , entonces

$$\mathbf{v}'\mathbf{X}'\mathbf{X}\mathbf{v} = (\mathbf{X}\mathbf{v})'\mathbf{X}\mathbf{v} = \mathbf{u}'\mathbf{u}$$

Como asumimos que  $\mathbf{X}$  es de rango completo (por columnas),  $\mathbf{X}\mathbf{v} \neq \mathbf{0}$ . Luego,  $\mathbf{X}'\mathbf{X}$  es positiva definida.

Por lo tanto,  $\hat{\boldsymbol{\beta}}$  es el estimador de MCO del modelo RLM.

# Teorema de Gauss-Markov

En el modelo RLM  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ , bajo las hipótesis:

- $\boldsymbol{\epsilon} \sim (\mathbf{0}, \sigma^2\mathbf{I})$ .
- $\mathbf{X}$  de rango completo.

el estimador de MCO de  $\boldsymbol{\beta}$  es el MELI (BLUE). Es decir,  $\hat{\boldsymbol{\beta}}$  es insesgado para  $\boldsymbol{\beta}$  y si  $\tilde{\boldsymbol{\beta}}$  es otro estimador de insesgado de  $\boldsymbol{\beta}$  y  $\mathbf{v}$  es un vector de dimensión  $p + 1$  distinto de  $\mathbf{0}$ , entonces  $\mathbf{v}'V(\tilde{\boldsymbol{\beta}})\mathbf{v} \geq \mathbf{v}'V(\hat{\boldsymbol{\beta}})\mathbf{v}$ .

# TGM: demostración

Como  $\epsilon \sim (\mathbf{0}, \sigma^2 \mathbf{I})$  se sigue que  $\mathbf{Y} \sim (\mathbf{X}\beta, \sigma^2 \mathbf{I})$ . Luego,

$$\begin{aligned} E(\hat{\beta}) &= E\left((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}\right) \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'E(\mathbf{Y}) \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X}\beta \\ &= \beta. \end{aligned}$$

Entonces,  $\hat{\beta}$  es insesgado para  $\beta$ .

$$\begin{aligned} V(\hat{\beta}) &= V\left((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}\right) \\ &= \left((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\right) V(\mathbf{Y}) \left((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\right)' \\ &= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X}' (\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

# TGM: demostración

Sea  $\tilde{\beta}$  otro estimador lineal insesgado para  $\beta$ . Es decir, existe una matrix  $\mathbf{A}_{(p+1) \times n}$  tal que  $\tilde{\beta} = \mathbf{A}\mathbf{Y}$ . Como se pide que  $\tilde{\beta}$  sea insesgado para  $\beta$  se debe cumplir

$$\begin{aligned} E(\tilde{\beta}) &= E(\mathbf{A}\mathbf{Y}) \\ &= \mathbf{A}E(\mathbf{Y}) \\ &= \mathbf{A}\mathbf{X}\beta. \end{aligned}$$

Lo anterior implica que  $\mathbf{A}\mathbf{X} = \mathbf{I}_{p+1}$ . Ahora sea  $\mathbf{C}_{(p+1) \times n}$  tal que  $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' + \mathbf{C}$ . Es fácil verificar que  $\mathbf{C}\mathbf{X} = \mathbf{0}$ . Ahora calculamos la varianza de  $\tilde{\beta}$ .

$$\begin{aligned} V(\tilde{\beta}) &= V(\mathbf{A}\mathbf{Y}) \\ &= \mathbf{A}V(\mathbf{Y})\mathbf{A}' \\ &= \sigma^2 \left( (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' + \mathbf{C} \right) \left( (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' + \mathbf{C} \right)'. \end{aligned}$$

Sea  $\mathbf{v}$  un vector de dimensión  $p+1$  y  $\mathbf{u} = \mathbf{C}'\mathbf{v}$ , entonces  $\mathbf{v}'\mathbf{C}\mathbf{C}'\mathbf{v} = (\mathbf{C}'\mathbf{v})' \mathbf{C}'\mathbf{v} = \mathbf{u}'\mathbf{u} \geq 0$ . Entonces,

$$\mathbf{v}'V(\tilde{\beta})\mathbf{v} = \mathbf{v}'V(\hat{\beta})\mathbf{v} + \sigma^2\mathbf{v}'\mathbf{C}\mathbf{C}'\mathbf{v} \geq \mathbf{v}'V(\hat{\beta})\mathbf{v}.$$

Por lo tanto, el estimador de MCO de  $\beta$  es el MELI.

# La matriz *sombbrero* $\mathbf{H}$

- Dado que  $\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{X}^{-1}\mathbf{X}'\mathbf{Y}$ , el vector de valores ajustados se puede escribir como

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

- La matriz  $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  se conoce como matriz *sombbrero* (*hat*) y se denota por  $\mathbf{H}$ . El nombre se debe a que  $\mathbf{H}$  le pone el sombrero a  $\mathbf{Y}$ .
- El vector de residuos ahora se puede escribir como

$$\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

- Se puede mostrar que  $\mathbf{H}$  es simétrica e idempotente y en consecuencia,  $\mathbf{I} - \mathbf{H}$  también lo es. Entonces, la SC residual se puede escribir como

$$SC_{error} = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$$

# Estimación por Máxima Verosimilitud

Para hacer estimación por intervalos y pruebas de hipótesis debemos agregar el supuesto de normalidad multivariada en los errores

$$\boldsymbol{\epsilon} \sim N_n(\mathbf{0}_n, \sigma^2 \mathbf{I}_n) \quad \Rightarrow \quad \mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$$

Recordemos que en el caso general si  $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  la densidad de  $Y$  está dada por

$$f(\mathbf{y}) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right\}$$

Entonces, la verosimilitud de  $\mathbf{Y}$  en el modelo RLM es

$$L(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\}$$
$$\ell(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}) = \log L(\cdot) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

# Estimación por Máxima Verosimilitud

Debemos maximizar  $\ell(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X})$  con respecto a  $\boldsymbol{\beta}$  y  $\sigma^2$ . Primero derivamos para encontrar los puntos críticos

$$\frac{\partial}{\partial \boldsymbol{\beta}} \ell = -\frac{1}{\sigma^2} (\mathbf{X}'\mathbf{X}\boldsymbol{\beta} - \mathbf{X}'\mathbf{y})$$

$$\frac{\partial}{\partial \sigma^2} \ell = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Entonces

$$\frac{\partial}{\partial \boldsymbol{\beta}} \ell = 0 \quad \Rightarrow \quad \mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$$

$$\frac{\partial}{\partial \sigma^2} \ell = 0 \quad \Rightarrow \quad \sigma^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

La solución para  $\sigma^2$  depende de  $\boldsymbol{\beta}$  y la solución para  $\boldsymbol{\beta}$  es la misma que por MCO. Entonces, si  $\mathbf{X}$  es de rango completo, los EMV son

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \quad \text{y} \quad \hat{\sigma}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}).$$

- Como el EMV de  $\beta$  coincide con el EMCO, se sigue que

$$E(\hat{\beta}) = \beta \quad \text{y} \quad V(\hat{\beta}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

- Además, por el supuesto de normalidad, se sigue que

$$\hat{\beta} \sim N_{p+1}(\beta, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}).$$

- Para determinar las propiedades de  $\hat{\sigma}^2$  es necesario presentar algunos resultados sobre de formas cuadráticas, ya que

$$\hat{\sigma}^2 = \frac{1}{n} \mathbf{Y}' (\mathbf{I} - \mathbf{H}) \mathbf{Y}.$$

# Distribuciones $\chi^2$ y $\chi^2$ no central

## Distribución $\chi^2$ no central

Sean  $X_1, \dots, X_n$  son v.a. normales independientes tales que  $E(X_i) = \mu_i$  y  $V(X_i) = 1$ ,  $i = 1, \dots, n$ . Entonces, la variable aleatoria  $Y$  dada por

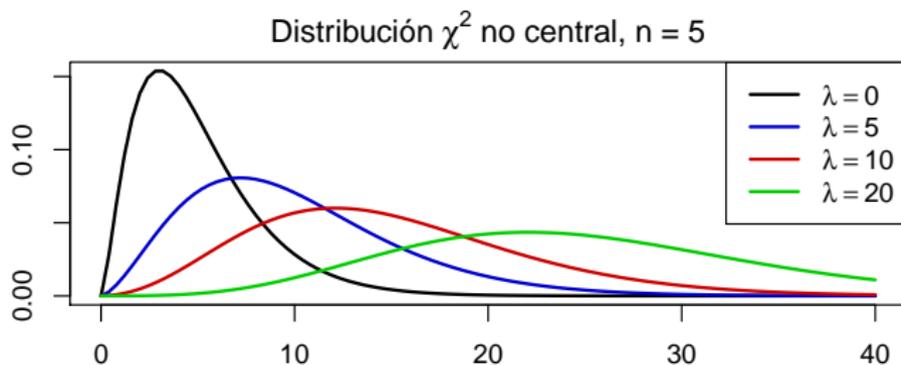
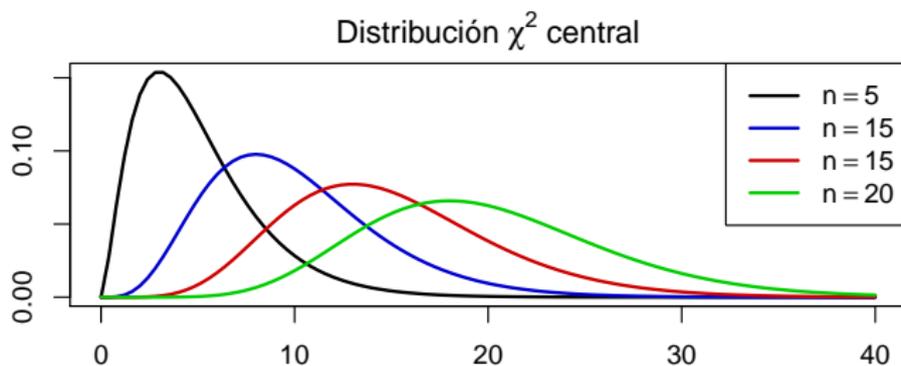
$$Y = \sum_{i=1}^n X_i^2$$

tiene distribución  $\chi^2$  no central con  $n$  grados de libertad y con parámetro de *no centralidad*  $\lambda = \sum_{i=1}^n \mu_i^2$ , lo que se denota como  $Y \sim \chi'_{(n,\lambda)}$ .

## Observación

Cuando  $\lambda = 0$ , la distribución  $\chi'_{(n,\lambda)}$  se reduce a la distribución  $\chi^2_{(n)}$ .

# Distribuciones $\chi^2$ y $\chi^2$ no central



## Teorema

Sean  $\mathbf{Y} \sim N_q(\boldsymbol{\mu}, \mathbf{I})$ ,  $\mathbf{A}_{q \times q}$ ,  $\mathbf{B}_{q \times q}$  son matrices simétricas e idempotentes de rangos  $r_1, r_2 \leq k$ , respectivamente, y  $\mathbf{C}_{s \times q}$  de rango completo, entonces:

- 1  $\mathbf{Y}'\mathbf{A}\mathbf{Y} \sim \chi_{(r_1, \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu})}^2$ .
- 2  $\mathbf{Y}'\mathbf{A}\mathbf{Y} \perp \mathbf{Y}'\mathbf{B}\mathbf{Y}$  si y sólo si  $\mathbf{A}\mathbf{B} = \mathbf{0}_{q \times q}$ .
- 3  $\mathbf{C}\mathbf{Y} \perp \mathbf{Y}'\mathbf{A}\mathbf{Y}$  si y sólo si  $\mathbf{C}\mathbf{A} = \mathbf{0}_{s \times q}$ .

## Resultado

En el modelo RLM, como  $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$  y la matriz  $\mathbf{I} - \mathbf{H}$  es simétrica e idempotente de rango  $n - p - 1$ , se sigue que

$$\frac{n\hat{\sigma}^2}{\sigma^2} = \frac{SC_{error}}{\sigma^2} \sim \chi^2_{(n-p-1)}.$$

Además, por ser  $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  de rango completo y

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

se sigue que  $\hat{\boldsymbol{\beta}} \perp SC_{error}$ , por lo que  $\hat{\boldsymbol{\beta}} \perp \hat{\sigma}^2_{MV}$ .

# Propiedades de $\hat{\sigma}^2$

- Del resultado anterior se sigue que

$$E(\hat{\sigma}^2) = \frac{n-p-1}{n} \sigma^2.$$

Luego,  $\hat{\sigma}^2$  no es insesgado para  $\sigma^2$ .

- Se puede definir el siguiente estimador insesgado para  $\sigma^2$  a partir de  $\hat{\sigma}^2$

$$\hat{\sigma}_{MCO}^2 = \frac{1}{n-p-1} SC_{error}.$$

La notación MCO es solamente para distinguir este estimador de  $\hat{\sigma}^2$ . En realidad el método de MCO no proporciona un estimador de  $\sigma^2$ .

- Para comparar  $\hat{\sigma}^2$  y  $\hat{\sigma}_{MCO}^2$  se utilizar el ECM. Al igual que en el modelo RLS,  $\hat{\sigma}^2$  resulta ser mejor estimador que  $\hat{\sigma}_{MCO}^2$ , a pesar del sesgo.

# Inferencias para $\sigma^2$ : intervalos de confianza

- De los resultados anteriores se tiene una cantidad pivotal para  $\sigma^2$

$$\frac{SC_{error}}{\sigma^2} \sim \chi^2_{(n-p-1)}.$$

- Para  $\alpha \in (0, 1)$ , un intervalo de confianza  $100(1 - \alpha)\%$  para  $\sigma^2$  está dado por

$$\left( \frac{SC_{error}}{\chi^2_{(n-p-1)}(1 - \alpha_2)}, \frac{SC_{error}}{\chi^2_{(n-p-1)}(\alpha_1)} \right)$$

con  $\alpha_1, \alpha_2 > 0$  tales que  $\alpha_1 + \alpha_2 = \alpha$ .

- Como ya se ha mencionado, los valores de  $\alpha_1$  y  $\alpha_2$  que proporcionan el intervalo de menor longitud deben obtenerse numéricamente, aunque siempre es posible utilizar  $\alpha_1 = \alpha_2 = \alpha/2$ .

# Inferencias para $\sigma^2$ : pruebas de hipótesis

- Para contrastar hipótesis sobre  $\sigma^2$  se utiliza el estadístico

$$S = \frac{SC_{error}}{\sigma_0^2}$$

con  $\sigma_0^2 > 0$  una constante conocida (que se fija en  $H_0$ ).

- Para contrastar las hipótesis  $H_0 : \sigma^2 = \sigma_0^2$  vs.  $H_1 : \sigma^2 \neq \sigma_0^2$ , la regla de decisión es rechazar  $H_0$  si  $S < \chi_{n-p-1}^2(\alpha/2)$  o si  $S > \chi_{n-p-1}^2(1 - \alpha/2)$ .
- En la siguiente tabla se resumen las dos pruebas de una cola.

| Hipótesis  | Regla de rechazo                 |
|--|----------------------------------|
| $H_0 : \sigma^2 \leq \sigma_0^2$ vs. $H_1 : \sigma^2 > \sigma_0^2$ | $S > \chi_{n-p-1}^2(1 - \alpha)$ |
| $H_0 : \sigma^2 \geq \sigma_0^2$ vs. $H_1 : \sigma^2 < \sigma_0^2$ | $S < \chi_{n-p-1}^2(\alpha)$     |

- En los tres casos el tamaño de prueba es  $\alpha$ .

# Intervalos de confianza para $\mathbf{a}'\boldsymbol{\beta}$

- También es posible construir intervalos de confianza para c.l. de las componentes de  $\boldsymbol{\beta}$ , de la forma  $\mathbf{a}'\boldsymbol{\beta}$ , con  $\mathbf{a} \in \mathbb{R}^{p+1}$  conocido.
- Por las propiedades de las distribución NMV

$$\mathbf{a}'\hat{\boldsymbol{\beta}} \sim N_1 \left( \mathbf{a}'\boldsymbol{\beta}, \sigma^2 \mathbf{a}' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{a} \right) \quad \text{y} \quad \frac{\mathbf{a}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sigma \sqrt{\mathbf{a}' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{a}}} \sim N(0, 1).$$

- Si se divide la expresión anterior por la raíz de  $\hat{\sigma}_{MCO}^2/(\sigma^2)$  para eliminar a  $\sigma^2$ , resulta que

$$\frac{\mathbf{a}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\hat{\sigma}_{MCO} \sqrt{\mathbf{a}' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{a}}} \sim t_{(n-p-1)}.$$

- Esta última expresión es una cantidad pivotal para  $\mathbf{a}'\boldsymbol{\beta}$ .

# Intervalos de confianza para $\mathbf{a}'\boldsymbol{\beta}$

- Un intervalo de confianza  $100(1 - \alpha)\%$  para  $\mathbf{a}'\boldsymbol{\beta}$  esta dado por

$$\mathbf{a}'\hat{\boldsymbol{\beta}} \pm t_{(n-p-1)}^{(1-\alpha/2)} \hat{\sigma}_{MCO} \sqrt{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}$$

- Cuando es de la forma  $\mathbf{a} = \mathbf{I}_n^{(i+1)}$ , entonces  $\mathbf{a}'\boldsymbol{\beta} = \beta_i$ , para  $i = 0, \dots, p$ .  
Luego, un intervalo de confianza  $100(1 - \alpha)\%$  para  $\beta_i$  esta dado por

$$\hat{\beta}_i \pm t_{(n-p-1)}^{(1-\alpha/2)} \hat{\sigma}_{MCO} \sqrt{(\mathbf{X}'\mathbf{X})_{ii}^{-1}}$$

donde  $\hat{\beta}_i$  es la  $i$ -ésima entrada de  $\hat{\boldsymbol{\beta}}$  y  $(\mathbf{X}'\mathbf{X})_{ii}^{-1}$  es el  $i$ -ésimo elemento de la diagonal de  $(\mathbf{X}'\mathbf{X})^{-1}$ .

# Contraste de hipótesis sobre $\mathbf{a}'\boldsymbol{\beta}$

- Para contrastar hipótesis sobre  $\mathbf{a}'\boldsymbol{\beta}$  se utiliza el estadístico

$$T = \frac{\mathbf{a}'\hat{\boldsymbol{\beta}} - b}{\hat{\sigma}_{MCO} \sqrt{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}}$$

y  $b$  es una constante conocida (que se fija en  $H_0$ ).

- Las tres pruebas de interés se resumen en la siguiente tabla

| Hipótesis  | Regla de rechazo                   |
|--|------------------------------------|
| $H_0 : \mathbf{a}'\boldsymbol{\beta} = b$ vs. $H_1 : \mathbf{a}'\boldsymbol{\beta} \neq b$ | $ T  > t_{(n-p-1)}^{(1-\alpha/2)}$ |
| $H_0 : \mathbf{a}'\boldsymbol{\beta} \geq b$ vs. $H_1 : \mathbf{a}'\boldsymbol{\beta} < b$ | $T < t_{(n-p-1)}^{(\alpha)}$       |
| $H_0 : \mathbf{a}'\boldsymbol{\beta} \leq b$ vs. $H_1 : \mathbf{a}'\boldsymbol{\beta} > b$ | $T > t_{(n-p-1)}^{(1-\alpha)}$     |

# Contraste de hipótesis sobre $\beta_i$

Las pruebas de hipótesis para  $\beta_i$  se resumen en el siguiente cuadro

| Hipótesis                                      | Regla de rechazo                   |
|--|------------------------------------|
| $H_0 : \beta_i = b$ vs. $H_1 : \beta_i \neq b$ | $ T  > t_{(n-p-1)}^{(1-\alpha/2)}$ |
| $H_0 : \beta_i \geq b$ vs. $H_1 : \beta_i < b$ | $T < t_{(n-p-1)}^{(\alpha)}$       |
| $H_0 : \beta_i \leq b$ vs. $H_1 : \beta_i > b$ | $T > t_{(n-p-1)}^{(1-\alpha)}$     |

donde

$$T = \frac{\hat{\beta}_i - b}{\hat{\sigma}_{MCO} \sqrt{(\mathbf{X}'\mathbf{X})_{ii}^{-1}}}$$

y  $b$  es una constante conocida, para  $i = 0, 1, \dots, p$ .

## Intervalos de confianza para $E(Y | \mathbf{x}_0)$ y $Y | \mathbf{x}_0$

Otro caso particular son los intervalos de confianza para la respuesta media y los intervalos de predicción.

Supongamos que  $\mathbf{x}^* = (x_1^*, \dots, x_p^*)'$  es un vector de constantes fijas y conocidas, un estimador del valor esperado de  $Y$  dado  $\mathbf{X} = \mathbf{x}^*$  es  $\mathbf{x}_0' \hat{\boldsymbol{\beta}}$ , con  $\mathbf{x}_0 = (1, \mathbf{x}^*)'$ .

Un intervalo de confianza  $100(1 - \alpha)\%$  para  $E(Y|\mathbf{x}^*)$  está dado por

$$\mathbf{x}_0' \hat{\boldsymbol{\beta}} \pm t_{n-p-a}(\alpha/2) \hat{\sigma}_{MCO} \sqrt{\mathbf{x}_0' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0}.$$

También es posible hacer pruebas de hipótesis sobre  $E(Y|\mathbf{x}^*)$ , basta adecuar las reglas de decisión anteriores. Un intervalo de predicción  $100(1 - \alpha)\%$  para una nueva observación de  $Y$  dado  $\mathbf{X} = \mathbf{x}^*$  está dado por

$$\mathbf{x}_0' \hat{\boldsymbol{\beta}} \pm t_{n-p-a}(\alpha/2) \hat{\sigma}_{MCO} \sqrt{1 + \mathbf{x}_0' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0}.$$

La justificación de este resultado es similar al caso del modelo RLS.

# Intervalos simultáneos para $\mathbf{a}'\boldsymbol{\beta}$ : Hotelling-Scheffé

Si  $\mathbf{a}'$  es un vector de dimensión  $p + 1$ , sabemos que

$$T_{\mathbf{a}} = \frac{\mathbf{a}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\hat{\sigma}_{MCO} \sqrt{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}} \sim t_{n-p-1}.$$

o equivalentemente

$$T_{\mathbf{a}}^2 = \frac{\mathbf{a}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\mathbf{a}}{\hat{\sigma}_{MCO}^2 \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}} \sim F_{1,n-p-1}.$$

Es sencillo mostrar que

$$\max_{\mathbf{a} \in \mathbb{R}^{p+1}} (p+1)^{-1} T_{\mathbf{a}}^2 = \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{X}'\mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{(p+1)\hat{\sigma}_{MCO}^2} \sim F_{p+1,n-p-1}.$$

\* El resultado anterior se usó para construir regiones de confianza para  $\boldsymbol{\beta}$ .

# Intervalos simultáneos para $\mathbf{a}'\boldsymbol{\beta}$ : Hotelling-Scheffé

Por lo tanto, para todo  $\mathbf{a} \in \mathbb{R}^{p+1}$  se cumple

$$P(T_{\mathbf{a}}^2 \leq K_{\alpha}) \geq 1 - \alpha$$

con  $K_{\alpha} = (p + 1)F_{p+1, n-p-1}(\alpha)$ . O equivalentemente

$$P\left(-\sqrt{K_{\alpha}} \leq \frac{\mathbf{a}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\hat{\sigma}_{MCO}\sqrt{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}} \leq \sqrt{K_{\alpha}}\right) \geq 1 - \alpha$$

De donde se obtiene un intervalo de confianza  $100(1 - \alpha)\%$  para  $\mathbf{a}'\boldsymbol{\beta}$  simultáneo para todo  $\mathbf{a} \in \mathbb{R}^{p+1}$

$$\mathbf{a}'\boldsymbol{\beta} \in \mathbf{a}'\hat{\boldsymbol{\beta}} \pm \hat{\sigma}_{MCO}\sqrt{K_{\alpha}\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}.$$

# Regiones de confianza

Como  $\beta$  es un vector, se pueden construir *regiones de confianza* para el vector completo o intervalos de confianza para c.l. de su componentes.

## Definición (región de confianza)

Sea una muestra aleatoria  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$  de una distribución  $F(\mathbf{y}|\boldsymbol{\theta})$ . Una región de confianza  $100(1 - \alpha)\%$  para  $\boldsymbol{\theta}$  es un conjunto aleatorio  $C = C(\mathbf{Y}_1, \dots, \mathbf{Y}_n)$  tal que

$$P(C \ni \boldsymbol{\theta}) \geq 1 - \alpha.$$

Dada una región de confianza para  $\boldsymbol{\theta}$ , se puede construir una prueba de hipótesis para contrastar  $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$  vs.  $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ , para algún vector de constantes conocidas  $\boldsymbol{\theta}_0$ , con regla de decisión rechazar  $H_0$  si  $\boldsymbol{\theta}_0 \notin C$ , con un tamaño de prueba  $\alpha$ .

# Regiones de confianza: ejemplo RLS

- En el modelo RLS, sabemos que

$$\hat{\boldsymbol{\beta}} \sim N_2(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}).$$

- Si se asume que  $\sigma^2$  es conocida,

# Regiones de confianza

Sabemos que  $\hat{\beta} \sim N_{p+1}(\beta, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1})$ , entonces es fácil mostrar que

$$(\hat{\beta} - \beta)' \mathbf{X}'\mathbf{X} (\hat{\beta} - \beta) / \sigma^2 \sim \chi_{(p+1)}^2.$$

Además, como  $(n - p - 1)\hat{\sigma}_{MCO}^2 / \sigma^2 \sim \chi_{n-p-1}^2$  y  $\hat{\beta} \perp \hat{\sigma}_{MCO}^2$ , se sigue que

$$\frac{(\hat{\beta} - \beta)' \mathbf{X}'\mathbf{X} (\hat{\beta} - \beta)}{\hat{\sigma}_{MCO}^2 (p + 1)} \sim F_{(p+1, n-p-1)}.$$

Por lo tanto, una región de confianza  $100(1 - \alpha)\%$  para  $\beta$  está dada por

$$C = \left\{ \mathbf{b} \in \mathbb{R}^{p+1} : \frac{(\hat{\beta} - \mathbf{b})' \mathbf{X}'\mathbf{X} (\hat{\beta} - \mathbf{b})}{\hat{\sigma}_{MCO}^2 (p + 1)} \leq F_{(p+1, n-p-1)}^{(1-\alpha)} \right\}.$$

# Regiones de confianza para $\beta$

Una prueba de hipótesis para contrastar

$$H_0 : \beta = \beta_0 \quad \text{vs.} \quad H_1 : \beta \neq \beta_0$$

consiste en rechazar  $H_0$  con un tamaño de prueba  $\alpha$  si

$$\beta_0 \notin C \quad \Leftrightarrow \quad \frac{(\hat{\beta} - \beta_0)' \mathbf{X}' \mathbf{X} (\hat{\beta} - \beta_0)}{\hat{\sigma}_{MCO}^2 (p+1)} > F_{p+1, n-p-1}(\alpha)$$

donde  $F_{(p+1, n-p-1)}^{(1-\alpha)}$  denota el cuantil  $1 - \alpha$  de la distribución  $F_{(p+1, n-p-1)}$ .

# La hipótesis lineal general

El caso más general que podemos considerar es hacer inferencias sobre  $\mathbf{A}\boldsymbol{\beta}$ , donde  $\mathbf{A}$  es una matriz de dimensión  $q \times (p + 1)$  de rango completo, que representa hacer inferencias sobre varias combinaciones lineales de las componentes de  $\boldsymbol{\beta}$  de manera simultánea. Podemos proceder como anteriormente si consideramos que

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

Por lo que podemos justificar fácilmente que

$$\frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{A}' (\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{A}')^{-1} \mathbf{A}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{q\hat{\sigma}_{MCO}^2} \sim F_{q, n-p-1}$$

Podemos utilizar el resultado anterior para construir una región de confianza y probar hipótesis sobre  $\mathbf{A}\boldsymbol{\beta}$ . Por ejemplo, la hipótesis que establece el análisis de varianza

$$H_0 : \beta_1 = \dots = \beta_p = 0 \quad \text{vs.} \quad H_1 : \beta_i \neq 0, \text{ para algún } i.$$

# La hipótesis lineal general

A continuación vamos a derivar la prueba del cociente de verosimilitudes generalizadas para contrastar las hipótesis

$$H_0 : \mathbf{A}\boldsymbol{\beta} = \mathbf{a} \quad \text{vs.} \quad H_1 : \mathbf{A}\boldsymbol{\beta} \neq \mathbf{a}.$$

con  $\mathbf{A}$  como antes y  $\mathbf{a}$  un vector dimensión  $q$  conocido y constante.

El máximo de la verosimilitud bajo  $H_0$  se obtiene de la manera usual y añadiendo la restricción  $\mathbf{A}\hat{\boldsymbol{\beta}} = \mathbf{a}$ , que se trata de un problema de optimización con restricciones de igualdad por lo que podemos utilizar la técnica de multiplicadores de Lagrange para resolverlo.

Se puede mostrar que la solución al problema anterior se obtiene en

$$\begin{aligned}\tilde{\boldsymbol{\beta}} &= \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}'(\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}')^{-1}(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{a}) \\ \tilde{\sigma}^2 &= \frac{1}{n}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})\end{aligned}$$

# La hipótesis lineal general

El máximo de la verosimilitud sin restricciones se obtiene al evaluar en los EMV

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad \text{y} \quad \hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta})$$

Con los resultados anteriores es sencillo verificar que

$$\begin{aligned} L_0 &= (2\pi\tilde{\sigma}^2)^{n/2} \exp\left\{-\frac{1}{2\tilde{\sigma}^2}(\mathbf{y} - \mathbf{X}\tilde{\beta})'(\mathbf{y} - \mathbf{X}\tilde{\beta})\right\} \\ &= (2\pi\tilde{\sigma}^2)^{-n/2} \exp\left\{-\frac{n}{2}\right\} \\ L_1 &= (2\pi\hat{\sigma}^2)^{-n/2} \exp\left\{-\frac{1}{2\hat{\sigma}^2}(\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta})\right\} \\ &= (2\pi\hat{\sigma}^2)^{-n/2} \exp\left\{-\frac{n}{2}\right\} \end{aligned}$$

Luego, el cociente de verosimilitudes generalizadas es

$$\Lambda = \frac{L_0}{L_1} = \left(\frac{\hat{\sigma}^2}{\tilde{\sigma}^2}\right)^{n/2} = \left(\frac{(\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta})}{(\mathbf{y} - \mathbf{X}\tilde{\beta})'(\mathbf{y} - \mathbf{X}\tilde{\beta})}\right)^{n/2}$$

# La hipótesis lineal general

La prueba de razón de verosimilitudes generalizadas tiene región de rechazo

$$\Lambda = \left( \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})} \right)^{n/2} < k \quad \Leftrightarrow \quad \frac{(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})}{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})} > c$$

Es sencillo, pero un poco fastidioso mostrar que

$$\frac{n-p-1}{q} \left( \frac{(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})}{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})} - 1 \right) = \frac{(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{a})' (\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}')^{-1} (\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{a})}{q\hat{\sigma}_{MCO}^2}$$

que bajo  $H_0$  tiene una distribución  $F_{q,n-p-1}$ .

Con esto justificamos todos los resultados anteriores obtenidos a partir de cantidades pivotaes.

# La hipótesis lineal general

La clave del resultado anterior es escribir  $\tilde{\beta} = \hat{\beta} - \mathbf{C}(\mathbf{A}\hat{\beta} - \mathbf{a})$ , con

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

Así al desarrollar

$$\begin{aligned}(\mathbf{y} - \mathbf{X}\tilde{\beta})'(\mathbf{y} - \mathbf{X}\tilde{\beta}) &= (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}) \\ &\quad - 2(\mathbf{A}\hat{\beta} - \mathbf{a})'\mathbf{C}'\mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\beta}) \\ &\quad + (\mathbf{A}\hat{\beta} - \mathbf{a})'(\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}')^{-1}(\mathbf{A}\hat{\beta} - \mathbf{a})\end{aligned}$$

Por último se debe verificar que

$$(\mathbf{A}\hat{\beta} - \mathbf{a})'\mathbf{C}'\mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\beta}) = 0.$$

## Inferencias para $\beta$ : resumen

Todas las inferencias de interés sobre  $\beta$  se pueden hacer a partir del resultado

$$\frac{(\hat{\beta} - \beta)' \mathbf{A}' (\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{A}')^{-1} \mathbf{A}(\hat{\beta} - \beta)}{q\hat{\sigma}_{MCO}^2}$$

con  $\mathbf{A}$  de dimensión  $q \times (p + 1)$  de rango completo,  $q \leq p + 1$ .

- Inferencias individuales y simultáneas sobre los componentes de  $\beta$ .
- Estimación y predicción de nuevas observaciones.
- Inferencias individuales y simultáneas sobre combinaciones lineales de los componentes de  $\beta$ .

Para las inferencias simultáneas se pueden utilizar regiones de confianza o bien usar intervalos de confianza simultáneos.

# Análisis de varianza

El objetivo del análisis de varianza (ANOVA) es determinar si el modelo

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \epsilon$$

explica mejor la variabilidad en  $Y$  que el modelo más simple

$$Y = \beta_0 + \epsilon.$$

Tradicionalmente el ANOVA se explica a partir de la descomposición de  $SC_{TC}$  como suma de  $SC_{reg}$  y  $SC_{error}$  y sus interpretaciones. Aunque es válido hacerlo de esta forma, también es posible justificar el ANOVA a partir de una prueba de razón de verosimilitudes generalizadas. El juego de hipótesis es

$$H_0 : \beta_1 = \dots = \beta_p = 0 \quad \text{vs.} \quad H_1 : \beta_i \neq 0, \text{ para algún } i.$$

# Análisis de varianza

Primero consideremos la varianza total en  $Y$

$$SC_{TC} = \sum_{i=1}^n (y_i - \bar{y}_n)^2.$$

Esta suma representa la varianza residual o no explicada por el modelo más simple, que se establece en la hipótesis nula. Después tenemos la varianza capturada por el modelo

$$SC_{reg} = \sum_{i=1}^n (\hat{y}_i - \bar{y}_n)^2.$$

Por último tenemos la varianza residual o no explicada por el modelo más completo, que se establece en la hipótesis alternativa.

$$SC_{error} = \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

# Análisis de varianza

Las sumas anteriores se pueden escribir como formas cuadráticas como sigue

$$SC_{TC} = \mathbf{y}' \left( \mathbf{I} - \frac{1}{n} \mathbf{J} \right) \mathbf{y}, \quad SC_{reg} = \mathbf{y}' \left( \mathbf{H} - \frac{1}{n} \mathbf{J} \right) \mathbf{y}, \quad SC_{error} = \mathbf{y}' (\mathbf{I} - \mathbf{H}) \mathbf{y}.$$

Con las expresiones anteriores es fácil probar que

$$SC_{TC} = SC_{reg} + SC_{error}.$$

Regresando a la prueba de hipótesis, si la varianza residual del modelo más completo es *mucho* menor que la varianza residual del modelo más simple, hay evidencia en contra de  $H_0$ , por lo que la prueba de hipótesis tendría una región de rechazo de la forma

$$\frac{SC_{error}}{SC_{tot}} < k \quad \Leftrightarrow \quad \frac{SC_{reg}}{SC_{error}} > c$$

donde  $k$  o  $c$  deben ser determinadas según el tamaño de prueba.

De los resultados sobre formas cuadráticas sabemos que

- $SC_{reg}/\sigma^2 \sim \chi_{(p+1,\lambda)}^{*2}$ , con  $\lambda = \beta' \mathbf{X}' (\mathbf{H} - \frac{1}{n} \mathbf{J}) \mathbf{X} \beta$ .
- $SC_{error}/\sigma^2 \sim \chi_{(n-p-1)}^2$ , y
- $SC_{reg} \perp SC_{error}$ .

En consecuencia

$$F := \frac{SC_{reg}/p}{SC_{error}/(n-p-1)} \sim F_{(p,n-p-1,\lambda)}^*$$

donde  $F_{p,n-p-1}^*$  representa una distribución  $F$  no central con  $p$  g.l. en el numerador,  $n-p-1$  g.l. en el denominador y parámetro de no centralidad  $\lambda$ .

Bajo la hipótesis nula,  $\beta' = (\beta_0, 0, \dots, 0)$  por lo que  $\mathbf{X}\beta = \beta_0\mathbf{1}_n$ . Luego

$$\beta' \mathbf{X}' \left( \mathbf{H} - \frac{1}{n} \mathbf{J} \right) \mathbf{X} \beta = 0.$$

Entonces, bajo  $H_0$

$$F \sim F_{(p, n-p-1)}$$

y podemos definir una prueba para contrastar  $H_0$  con el estadístico anterior. Tal prueba tiene como regla de decisión es *rechazar*  $H_0$  con un tamaño de prueba  $\alpha$  si

$$F > F_{(p, n-p-1)}^{(1-\alpha)}.$$

donde  $F_{(p, n-p-1)}^{(1-\alpha)}$  es el cuantil  $1 - \alpha$  de una distribución  $F_{(p, n-p-1)}$

## \*El ANOVA y la hipótesis lineal general

La prueba anterior también puede obtenerse a partir de cociente de verosimilitudes generalizadas. De acuerdo con resultados anteriores, para contrastar la hipótesis lineal general  $H_0 : \mathbf{A}\boldsymbol{\beta} = \mathbf{a}$  el estadístico de prueba es

$$F = \frac{(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{a})' (\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}')^{-1} (\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{a})}{q\hat{\sigma}_{MCO}^2}.$$

Si  $A = (\mathbf{0} \mid \mathbf{I}_p)$  y  $\mathbf{a} = \mathbf{0}$ , se tiene la prueba de significancia de la regresión, que se contrasta con el ANOVA.

Para justificar la validez del ANOVA, a partir de los resultados sobre la hipótesis lineal general, hace falta verificar que

$$\hat{\boldsymbol{\beta}}' \mathbf{A}' (\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}')^{-1} \mathbf{A}\hat{\boldsymbol{\beta}} = \mathbf{y}' \left( \mathbf{H} - \frac{1}{n}\mathbf{J} \right) \mathbf{y},$$

## \*El ANOVA y la hipótesis lineal general

Para verificar el resultado anterior resulta muy útil escribir a  $\mathbf{X}$  en forma particionada como  $(\mathbf{1} \mid \mathbf{X}_1)$ , donde  $\mathbf{X}_1$  la matriz de datos. Se puede comprobar que

$$\mathbf{X}'\mathbf{X} = \left( \begin{array}{c|c} n & n\bar{\mathbf{x}}' \\ \hline n\bar{\mathbf{x}} & \mathbf{X}'_1\mathbf{X}_1 \end{array} \right)$$
$$(\mathbf{X}'\mathbf{X})^{-1} = \left( \begin{array}{c|c} n^{-1} + \bar{\mathbf{x}}'\mathbf{S}^{-1}\bar{\mathbf{x}} & -\bar{\mathbf{x}}'\mathbf{S}^{-1} \\ \hline -\mathbf{S}^{-1}\bar{\mathbf{x}} & \mathbf{S}^{-1} \end{array} \right)$$

Donde  $\mathbf{S} = \mathbf{X}'_1(\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{X}_1$  es la matriz de suma de cuadrados y productos cruzados de  $\mathbf{X}_1$  y  $\bar{\mathbf{x}} = \frac{1}{n}\mathbf{X}'_1\mathbf{1}$  es el vector de promedios de las columnas de  $\mathbf{X}_1$ . También es importante notar que premultiplicar por  $A = (\mathbf{0} \mid \mathbf{I}_p)$  equivale a eliminar la primera fila y posmultiplicar por  $\mathbf{A}$  equivale a eliminar la primera columna.

## \*El ANOVA y la hipótesis lineal general

También es posible justificar el ANOVA directamente maximizando la verosimilitud con la restricción  $\beta_1 = \dots = \beta_p = 0$ , en cuyo caso la verosimilitud se reduce a

$$L = (2\pi\sigma^2)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} (y_i - \beta_0)^2 \right\}$$

Que se maximiza cuando  $\tilde{\beta}_0 = \bar{y}$  y  $\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = SC_{TC}/n$ . Luego,

$$L_0 \propto (SC_{TC})^{-n/2}$$

A partir de este resultado se sigue que la región de rechazo de la prueba de cociente de verosimilitudes generalizadas es de la forma

$$\frac{SC_{error}}{SC_{TC}} < k \quad \Leftrightarrow \quad \frac{SC_{reg}}{SC_{error}} > c$$

que coincide con el resultado obtenido solamente a partir de las sumas de cuadrados.

# La tabla ANOVA

- La prueba que se hace en el ANOVA se conoce como prueba  $F$  y es importante en el análisis de regresión, pues indica si realmente existe algún efecto de las variables auxiliares en la variable respuesta  $Y$ .
- Los resultados de la prueba  $F$  se resumen en la tabla ANOVA.

| F.V.  | G.L.        | S.C.  | C.M.                     | F                     |
|-------|-------------|---|--------------------------|-----------------------|
| Reg   | $p$         | $\mathbf{y}'(\mathbf{H} - \frac{1}{n}\mathbf{J})\mathbf{y}$ | $SC_{reg}/p$             | $CM_{reg}/CM_{error}$ |
| Error | $n - p - 1$ | $\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}$            | $SC_{error}/(n - p - 1)$ |                       |
| TC    | $n - 1$     | $\mathbf{y}'(\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{y}$ | $SC_{TC}/(n - 1)$        |                       |

# Casos patológicos del ANOVA

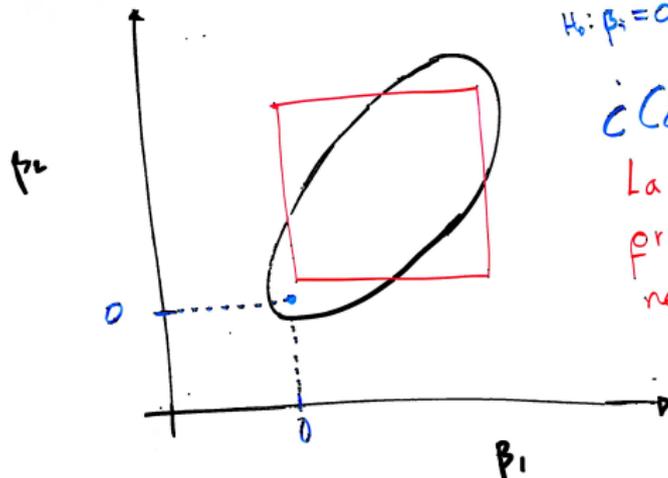
Prueba F no rechaza  $H_0: \beta_1 = \beta_2 = 0$

✗ Pruebas t individuales, ambas rechazan

$H_0: \beta_1 = 0$ .

¿Cómo proceder?

La significancia conjunta de las pruebas t individuales no es necesariamente  $100(1-\alpha)\%$ .



• Región de confianza  $100(1-\alpha)\%$

• Intervalos individuales  $100(1-\alpha)\%$

# Casos patológicos del ANOVA

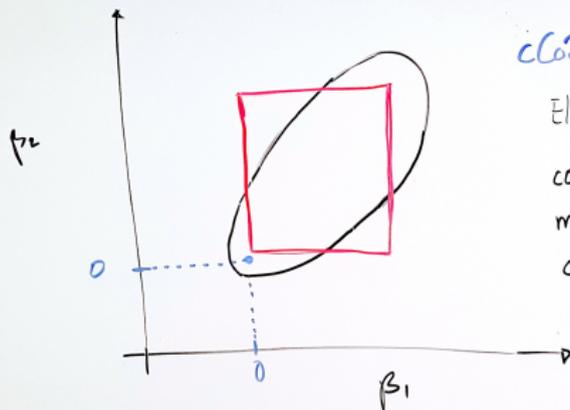
Prueba F no rechaza  $H_0: \beta_1 = \beta_2 = 0$ .

Pruebas t simultáneas rechazan  $H_0: \beta_i = 0$ .

¿Cómo proceder?

El planteamiento es correcto. La contradicción se debe a que el método de Bonferroni acota la confianza (significancia)

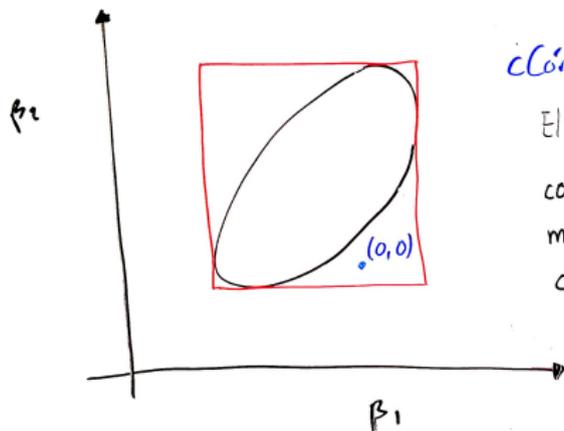
simultánea con el método de inclusión y exclusión, por lo que la confianza (significancia) exacta no es conocida.



• Región de confianza  $100(1-\alpha)\%$

• Intervalos de confianza simultáneos  
 $100(1-\alpha)\%$  (Bonferroni)

# Casos patológicos del ANOVA



- Región de confianza  $100(1-\alpha)\%$
- Intervalos de confianza simultáneos  $100(1-\alpha)\%$  (Hotelling-Scheffé)

Prueba F rechaza  $H_0: \beta_1 = \beta_2 = 0$ .

Pruebas  $t$  simultáneas no rechazan  $H_0: \beta_i = 0$

¿Cómo proceder?

El planteamiento es correcto. La contradicción se debe a que el método de HS acota la confianza (significancia)

simultánea con un argumento de optimización, por lo que la confianza (significancia) exacta no es conocida.

## Coefficiente de determinación vs $R^2_{adj}$

- También en el modelo RLM el coeficiente de determinación sirve como una medida del ajuste del modelo. Se calcula como en el caso simple

$$R^2 = \frac{SC_{reg}}{SC_{TC}} = 1 - \frac{SC_{error}}{SC_{TC}}$$

- El coeficiente  $R^2$  siempre crece conforme se incluyen más variables en el modelo, esto impide que pueda ser utilizado para comparar modelos.
- Si se utiliza el  $R^2$  para seleccionar el mejor modelo, siempre será mejor el modelo con más variables, lo cual no necesariamente es cierto.
- Para resolver este problema se define el  $R^2$  ajustado, que penaliza el número de variables que tiene el modelo. Se calcula como

$$R^2_{adj} = 1 - \frac{SC_{error}/(n - p - 1)}{SC_{TC}/(n - 1)} = 1 - \frac{CM_{error}}{CM_{TC}}$$