

# Unidad 3 Regresión Lineal Múltiple

La idea de la regresión lineal múltiple es modelar el valor esperado de la variable respuesta como combinación lineal de más de una variable explicativa. Si  $Y$  denota la variable respuesta y  $X_1, \dots, X_p$  las variables explicativas, entonces el primer supuesto del modelo de regresión lineal múltiple es

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

El supuesto de varianza constante se mantiene

$$V(Y|X_1 = x_1, \dots, X_p = x_p) = \sigma^2$$

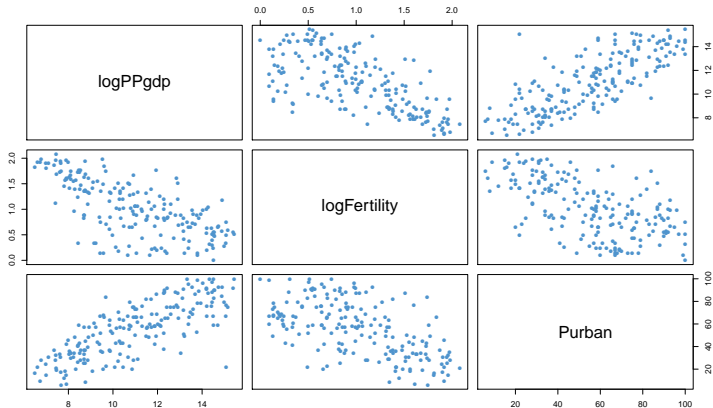
Las solución al problema de inferencia de este modelo (estimación puntual, por intervalos y pruebas de hipótesis) es relativamente *fácil*.

La principal dificultad del análisis de RLM es la interpretación del modelo. Incrementar el número de variables repercute en lo siguiente:

- Entender la relación entre las  $X$  y  $Y$ , porque ya no es posible graficar.
- La interpretación de los parámetros.
- Hacer inferencias simultáneas.
- Interacción entre las variables.
- Selección del *mejor* modelo.
- Identificar desviaciones de los supuestos.

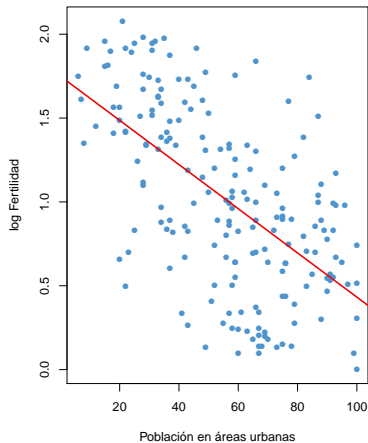
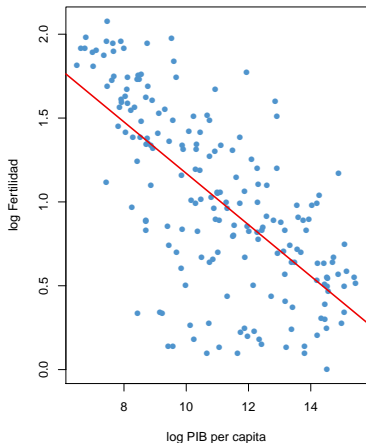
# Ejemplo

Se tiene un conjunto de datos con observaciones de 193 países de la ONU. De cada país se registraron tres variables: tasa de fecundidad (*fecundidad*), producto interno bruto per capita (*pibp*) y porcentaje de población en áreas urbanas (*purbana*). El objetivo es modelar la fecundidad a partir de las otras dos variables.



## Ejemplo (cont.)

Como la relación entre  $\log(\text{fecundidad})$  y  $\log(\text{pibpc})$  y entre  $\log(\text{fecundidad})$  y  $\text{purbana}$  parece ser lineal, podemos ajustar un modelo RLS en cada caso.



## Ejemplo (cont.)

- El coeficiente  $R^2$  del modelo que relaciona  $\log(\text{fecundidad})$  y  $\log(\text{pibpc})$  es 0.4591. Esto significa que  $\log(\text{pibpc})$  explica 46% de la variabilidad de  $\log(\text{fecundidad})$ .
- En el modelo de  $\log(\text{fecundidad})$  contra  $\text{purbana}$  el valor del  $R^2$  es 0.3482.
- ¿Cuál será el valor del  $R^2$  en un modelo que combine las dos variables explicativas? ¿Será la suma de los  $R^2$  de los modelos individuales?
- En el gráfico de dispersión de  $\log(\text{pibpc})$  y  $\text{purbana}$  se aprecia alguna relación entre estas dos variables. Posiblemente, parte de la variabilidad de  $\log(\text{fecundidad})$  que explican  $\log(\text{pibpc})$  y  $\text{purbana}$  sea *común*.
- El valor del  $R^2$  en el modelo con las dos variables es 0.4689.

# Planteamiento del Problema

El objetivo es modelar  $Y$  a partir de  $p$  variables explicativas  $X_1, \dots, X_p$ . Al igual que en el modelo RLS se asume que el valor esperado de  $Y$  se puede expresar como función lineal de las  $X_i$ :

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

Se asume que  $Y$  varía alrededor de su media de manera independiente de las  $X_i$ :

$$V(Y|X_1 = x_1, \dots, X_p = x_p) = \sigma^2$$

Lo anterior se puede representar como

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

donde  $\epsilon \sim (0, \sigma^2)$ .

## Planteamiento (cont.)

El modelo tiene  $p + 2$  parámetros desconocidos a estimar. Para ello se asume que se tienen  $n$  observaciones de la variable respuesta con sus respectivas variables explicativas.

$$Y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_p x_{1p} + \epsilon_1$$

$$Y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_p x_{2p} + \epsilon_2$$

$\vdots$

$$Y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_p x_{np} + \epsilon_n$$

Además se asume que los errores no están correlacionados, es decir,  $\text{Cov}(\epsilon_i, \epsilon_j) = 0$ , con  $i, j = 1, \dots, n$ , y  $i \neq j$ .



## Planteamiento (cont.)

Para facilitar la resolución del problema se utilizar la siguiente representación matricial:

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

Con lo cual el modelo queda expresado como

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

con  $\boldsymbol{\epsilon} \sim (\mathbf{0}_n, \sigma^2 \mathbf{I}_n)$ . La matriz  $\mathbf{X}$  recibe el nombre de matriz de diseño.

# Estimación por MCO

Al igual que en el caso *simple*, dado un valor específico de  $\beta$ , se define el vector de valores ajustados como  $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$  y el vector de errores como  $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$ . 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  $Q(\beta)$  tiene un mínimo en  $\hat{\beta}$  utilizamos el criterio de las segundas derivadas parciales. Es sencillo mostrar que

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

que no depende de  $\beta$ . 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,  $\hat{\beta}$  es el estimador de MCO del modelo RLM.

# Teorema de Gauss-Markov

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

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

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

# 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  $\tilde{\beta}$ .

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

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.



# Valores ajustados, residuos y suma de cuadrados residual

El vector de valores ajustados se define 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 *sombrero* (*hat*) y se denota por  $\mathbf{H}$ . El nombre se debe a que  $\mathbf{H}$  le pone el sombrero a  $\mathbf{y}$ . Se puede mostrar que  $\mathbf{H}$  es simétrica e idempotente. El vector de residuos entonces es

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{H}\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{y}$$

Como  $\mathbf{H}$  es simétrica e idempotente,  $\mathbf{I} - \mathbf{H}$  también lo es. Entonces, la suma de cuadrados residual se calcula como

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

Otra expresión útil y que se demuestra fácilmente es la siguiente

$$SC_{error} = \mathbf{y}'\mathbf{y} - \hat{\mathbf{y}}'\hat{\mathbf{y}}.$$

# Estimación por Máxima Verosimilitud

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

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

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

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

Entonces

$$\begin{aligned}\frac{\partial}{\partial \beta} \ell = 0 &\Rightarrow \mathbf{X}'\mathbf{X}\beta = \mathbf{X}'\mathbf{y} \\ \frac{\partial}{\partial \sigma^2} \ell = 0 &\Rightarrow \sigma^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}\beta)' (\mathbf{y} - \mathbf{X}\beta)\end{aligned}$$

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

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

# Propiedades de los EMV

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 adicionales sobre distribuciones de formas cuadráticas, aunque debemos notar que

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n} (\mathbf{y} - \mathbf{X}\hat{\beta})' (\mathbf{y} - \mathbf{X}\hat{\beta}) \\ &= \frac{1}{n} SC_{error} = \frac{1}{n} \\ &= \frac{1}{n} \mathbf{e}'\mathbf{e} \\ &= \frac{1}{n} \mathbf{y}' (\mathbf{I} - \mathbf{H}) \mathbf{y}.\end{aligned}$$

# Distribución $\chi^2$ no central

## Definición (distribución $\chi^2$ no central)

Se dice que una variable aleatoria (absolutamente) continua  $Y$  tiene una distribución  $\chi^2$  no central con parámetros  $\nu$  y  $\lambda$  si su densidad está dada por

$$f(y|\nu, \lambda) = \sum_{j=0}^{\infty} Po(j|\lambda/2) \chi^2(y|\nu + 2j)$$

donde  $Po(\cdot|\lambda/2)$  denota la densidad Poisson con parámetro  $\lambda/2$  y  $\chi^2(\cdot|\nu + 2j)$  denota la densidad  $\chi^2$  (central) con  $\nu + j$  grados de libertad.

Lo anterior se denota como  $Y \sim \chi^{*2}(\nu, \lambda)$ ,  $\nu$  son grados de libertad y  $\lambda$  se conoce como parámetro de no centralidad.

Cuando  $\lambda = 0$ , la distribución  $\chi^{*2}$  se reduce a la distribución  $\chi^2$  (central) con los mismos grados de libertad.

## Resultados

Sean  $\mathbf{Y} \sim N_k(\boldsymbol{\mu}, \mathbf{I})$ ,  $\mathbf{A}_{k \times k}$ ,  $\mathbf{B}_{k \times k}$  idempotentes de rango  $r_1, r_2 \leq k$ , respectivamente, y  $\mathbf{C}_{l \times k}$  de rango completo, entonces:

- $\mathbf{Y}'\mathbf{A}\mathbf{Y} \sim \chi^{*2}(r, \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu})$ .
- $\mathbf{Y}'\mathbf{A}\mathbf{Y} \perp \mathbf{Y}'\mathbf{B}\mathbf{Y}$  si y sólo si  $\mathbf{A}\mathbf{B} = \mathbf{0}$ .
- $\mathbf{Y}'\mathbf{A}\mathbf{Y} \perp \mathbf{C}\mathbf{Y}$  si y sólo si  $\mathbf{C}\mathbf{A} = \mathbf{0}$ .

# Propiedades de los EMV

Como  $\hat{\sigma}^2 = \frac{1}{n} \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$ ,  $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$  y  $\mathbf{I} - \mathbf{H}$  es idempotente de rango  $n - p - 1$ , se sigue que

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^{*2}(n - p - 1, \lambda)$$

donde

$$\begin{aligned}\lambda &= (\mathbf{X}\boldsymbol{\beta})'(\mathbf{I} - \mathbf{H})\mathbf{X}\boldsymbol{\beta} \\ &= \boldsymbol{\beta}'\mathbf{X}'(\mathbf{I} - \mathbf{H})\mathbf{X}\boldsymbol{\beta} \\ &= \boldsymbol{\beta}'(\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{H}\mathbf{X})\boldsymbol{\beta} \\ &= 0.\end{aligned}$$

Por lo tanto,

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n - p - 1)$$

De lo anterior se sigue que

$$E(\hat{\sigma}^2) = \frac{n-p-1}{n}\sigma^2$$
$$V(\hat{\sigma}^2) = 2\frac{n-p-1}{n^2}\sigma^4.$$

El EMV de  $\sigma^2$  es sesgado pero es fácil mostrar que el EMCO es insesgado. Como  $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ ,  $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  es de rango completo y

$$\begin{aligned}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I} - \mathbf{H}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= 0.\end{aligned}$$

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



# Inferencias para $\sigma^2$

Debido a que

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

es sencillo hacer inferencias sobre  $\sigma^2$ .

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

$$\left( \frac{SC_{error}}{\chi_{n-p-1}^2(\alpha/2)}, \frac{SC_{error}}{\chi_{n-p-1}^2(1 - \alpha/2)} \right)$$

donde  $\chi_{n-p-1}^2(\alpha/2)$  denota el cuantil superior  $\alpha/2$  de una distribución  $\chi_{n-p-1}^2$

El intervalo anterior es sólo uno de los múltiples que se pueden construir. Cualquier par de cuantiles  $\chi_{n-p-1}^2(1 - \alpha_1)$  y  $\chi_{n-p-1}^2(\alpha_2)$  tales que  $\alpha_1 + \alpha_2 = \alpha$  y  $\alpha_1, \alpha_2 \geq 0$  se pueden utilizar para un intervalo de confianza  $100(1 - \alpha)\%$ . La principal diferencia está en la longitud de los intervalos calculados.

# Inferencias para $\sigma^2$

Para contrastar las hipótesis

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{vs.} \quad H_1 : \sigma^2 \neq 0$$

con  $\sigma_0^2 > 0$  fijo y conocido, se utiliza el estadístico

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

y la regla de decisión es *rechazar*  $H_0$  si

$$S < \chi_{n-p-1}^2 (1 - \alpha/2) \quad \text{o} \quad S > \chi_{n-p-1}^2 (\alpha/2).$$

En la siguiente tabla se resumen los resultados para las pruebas de una cola

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

# Inferencias para $\beta$

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

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

Dada una región de confianza para  $\theta$ , se puede construir una prueba de hipótesis para contrastar

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta \neq \theta_0$$

para algún vector de constantes conocidas  $\theta_0$  con regla de decisión *rechazar*  $H_0$  si

$$\theta_0 \notin C$$

con una significancia  $\alpha$ .

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

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

# Inferencias para $\beta$

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

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

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 una significancia  $\alpha$  si

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

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

## Inferencias para $\beta$

Otra posibilidad es hacer inferencias sobre c.l. de las componentes de  $\beta$  del tipo  $\mathbf{a}'\beta$ , con  $\mathbf{a}$  un vector de dimensión  $p + 1$  de constantes conocidas.

Por las propiedades de la distribución NMV

$$\mathbf{a}'\hat{\beta} \sim N_1 \left( \mathbf{a}'\beta, \sigma^2 \mathbf{a}' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{a} \right)$$

Entonces,

$$\frac{\mathbf{a}' (\hat{\beta} - \beta)}{\sigma \sqrt{\mathbf{a}' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{a}}} \sim N(0, 1)$$

Si *reemplazamos*  $\sigma^2$  por  $\hat{\sigma}_{MCO}^2$ , entonces

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

El estadístico  $T_{\mathbf{a}}$  se puede utilizar para construir intervalos de confianza y pruebas de hipótesis para  $\mathbf{a}'\beta$ .

# Inferencias para $\beta$

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

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

donde  $t_{n-p-a}(\alpha/2)$  denota el cuantil superior  $\alpha/2$  de una distribución  $t_{n-p-a}$ . Las pruebas de hipótesis se resumen en el siguiente cuadro

Hipótesis	Regla de rechazo
$H_0 : \mathbf{a}'\beta = b$ vs. $H_1 : \mathbf{a}'\beta \neq b$	$ T_{\mathbf{a}}  > t_{n-p-a}(\alpha/2)$
$H_0 : \mathbf{a}'\beta \geq b$ vs. $H_1 : \mathbf{a}'\beta < b$	$T_{\mathbf{a}} < -t_{n-p-a}(\alpha)$
$H_0 : \mathbf{a}'\beta \leq b$ vs. $H_1 : \mathbf{a}'\beta > b$	$T_{\mathbf{a}} > t_{n-p-a}(\alpha)$

donde

$$T_{\mathbf{a}} = \frac{\mathbf{a}'\hat{\beta} - b}{\hat{\sigma}_{MCO}\sqrt{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}}$$

y  $b$  es una constante conocida

# Inferencias para $\beta$

Como casos particulares, cuando es de la forma  $\mathbf{a} = \mathbf{1}_{(i+1)}$ , entonces  $\mathbf{a}'\beta = \beta_i$ , para  $i = 0, \dots, p$ .

Un intervalo de confianza  $100(1 - \alpha)\%$  para  $\beta_i$  esta dado por

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

donde  $\hat{\beta}_i$  es la  $i$ -ésima entrada de  $\hat{\beta}$  y  $(\mathbf{X}'\mathbf{X})_{ii}^{-1}$  es el  $i$ -ésimo elemento de la diagonal de  $(\mathbf{X}'\mathbf{X})^{-1}$ . 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-a}(\alpha/2)$
$H_0 : \beta_i \geq b$ vs. $H_1 : \beta_i < b$	$t < -t_{n-p-a}(\alpha)$
$H_0 : \beta_i \leq b$ vs. $H_1 : \beta_i > b$	$t > t_{n-p-a}(\alpha)$

donde

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

y  $b$  es una constante conocida.



# Intervalos de confianza y predicción

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

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

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  $\mathbf{a} \in \mathbb{R}^{p+1}$

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

# La hipótesis lineal general

El caso más general que podemos considerar es hacer inferencias sobre  $\mathbf{A}\beta$ , 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  $\beta$  de manera simultánea. Podemos proceder como anteriormente si consideramos que

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

Por lo que podemos justificar fácilmente que

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

Podemos utilizar el resultado anterior para construir una región de confianza y probar hipótesis sobre  $\mathbf{A}\beta$ . 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

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

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

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



# Resumen: inferencias para $\beta$

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

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

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

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

Para hacer inferencias simultaneas podemos tomar en cuenta la covarianza y hacer regiones de confianza o bien dejarla a un lado y hacer 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

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

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

$$\frac{SC_{error}}{SC_{tot}} < k$$

donde  $k$  debe ser determinada por el nivel de significancia.

Por la descomposición de  $SC_{TC}$  tenemos la siguiente equivalencia

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

donde  $k$  debe ser determinada por el nivel de significancia. Ahora, 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
- $SC_{reg} \perp SC_{error}$ .

Por lo que

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

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  $H_0$ ,  $\beta = (\beta_0, 0, \dots, 0)$  por lo que  $\mathbf{X}\beta = \beta_0\mathbf{1}_n$ . Luego

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

Entonces, bajo  $H_0$

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

y podemos definir una prueba para contrastar  $H_0$  con el estadístico anterior. La regla de decisión es *rechazar*  $H_0$  con un nivel de significancia  $\alpha$  si

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

La prueba anterior también puede obtenerse a partir de cociente de verosimilitudes generalizadas. Recordemos que para 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 elegimos  $\mathbf{A} = (\mathbf{0} \mid \mathbf{I}_p)$  y  $\mathbf{a} = \mathbf{0}$ , recuperamos la hipótesis que se prueba con el ANOVA, sólo hace falta verificar que la prueba el mismo. En concreto, hace falta verificar que

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

lo cual es *muy sencillo*... (no es cierto).

Como consecuencia, el ANOVA es un caso particular de 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. Con esta definición, 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 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  $\mathbf{A} = (\mathbf{0} \mid \mathbf{I}_p)$  equivale a eliminar la primera fila y posmultiplicar por  $\mathbf{A}$  equivale a eliminar la primera columna.



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,

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

Con este resultado y el obtenido anteriormente 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 son los mismos resultados que obtuvimos estudiando solamente las sumas de cuadrados.

# Tabla ANOVA

La prueba que se hace en el ANOVA se conoce como prueba  $F$  y es muy importante en el análisis de regresión, pues nos indica si realmente existe algún efecto de las variables independientes y la variable respuesta. Cuando no rechazamos  $H_0$ , entonces el modelo de regresión planteado no es válido. 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	$n - p - 1$	$\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}$	$SC_{error}/(n - p - 1)$	
T.C.	$n - 1$	$\mathbf{y}'(\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{y}$	$SC_{TC}/(n - 1)$	

## 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  $R^2$  tiene el problema de que siempre crece conforme vamos incluyendo mas variables lo cual crea un conflicto al momento de querer seleccionar modelos pues si utilizamos el  $R^2$  para determinar el mejor modelo, siempre sería 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

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

# Casos patológicos del ANOVA

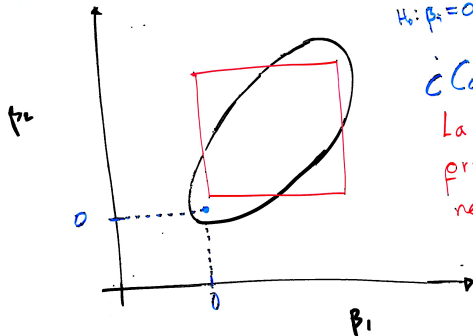
Prueba F no rechaza  $H_0: \beta_1 = \beta_2 = 0$

✗ Pruebas t individuales, ambas rechazan

$H_0: \beta_2 = 0$ .

¿Cómo proceder?

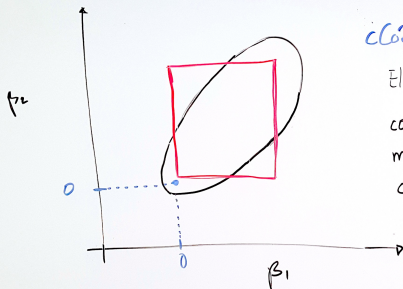
La significancia conjunta de las pruebas t individuales no es necesariamente  $100(1-\alpha)\%$ .



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

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

# Casos patológicos del ANOVA



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

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

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

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

¿Cómo proceder?

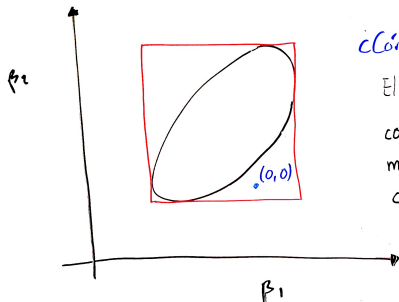
El planteamiento es correcto. La contradicción se debe a que el método de Bonferroni acota la confianza (significancia)

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

# Casos patológicos del ANOVA

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.

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

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