Unidad 4
Regresión Lineal Múltiple

Introducción

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 $X_1, \ldots, X_p$ las variables explicativas, entonces el primer supuesto del modelo de regresión lineal múltiple es

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

El supuesto de varianza constante se mantiene

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

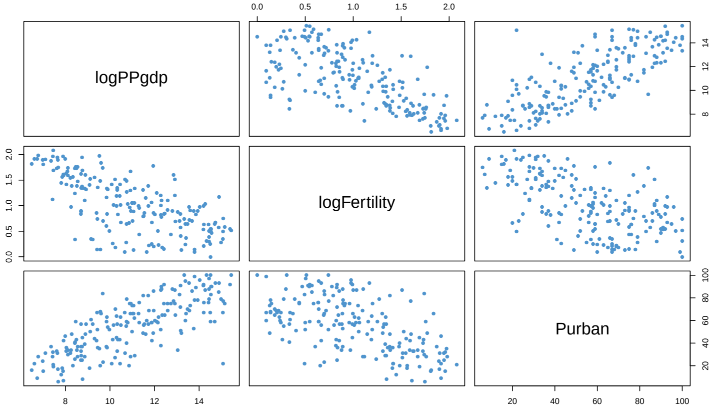
Las solución al problema de inferencia de este modelo (estimación puntual, por intervalos y pruebas de hipótesis) es relativamente *fácil*.

Introducción (cont.)

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

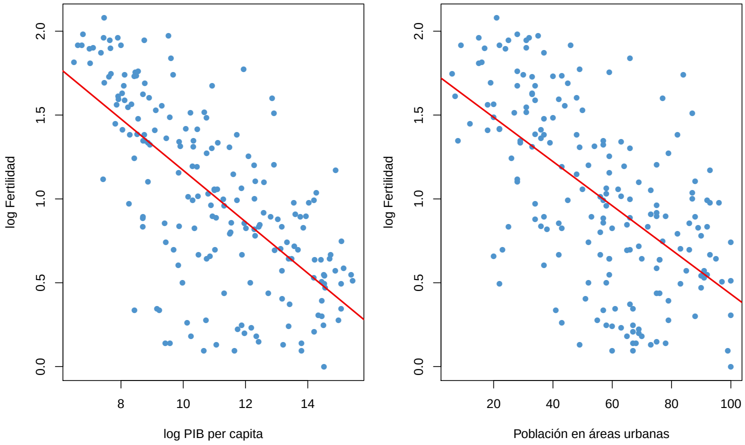
- Entender la relación entre las X y Y, porque ya no es posible graficar.
- La interpretación de los parámetros.
- Hacer inferencias simultáneas.
- Interacción entre las variables.
- Selección del *mejor* modelo.
- Identificar desviaciones de los supuestos.

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(`fecundidad`) y log(`pibpc`) y entre log(`fecundidad`) y `purbana` parece ser lineal, podemos ajustar un modelo RLS en cada caso.



Ejemplo (cont.)

- El coeficiente R^2 del modelo que relaciona $\log(\texttt{fecundidad})$ y $\log(\texttt{pibpc})$ es 0.4591. Esto significa que $\log(\texttt{pibpc})$ explica 46 % de la variabilidad de $\log(\texttt{fecundidad})$.
- En el modelo de $\log(\texttt{fecundidad})$ contra `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 $\log(\texttt{pibpc})$ y `purbana` se aprecia alguna relación entre estas dos variables. Posiblemente, parte de la variabilidad de $\log(\texttt{fecundidad})$ que explican $\log(\texttt{pibpc})$ y `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 $X_1, \ldots, X_p$. Al igua que en el modelo RLS se asume que el valor esperado de Y se puede expresar como función lineal de las X_i:

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

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

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

Lo anterior se puede representar como

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

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.

$$
\begin{aligned}
Y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \ldots + \beta_p x_{1p} + \epsilon_1 \\
Y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \ldots + \beta_p x_{2p} + \epsilon_2 \\
&\vdots \\
Y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_p x_{np} + \epsilon_n
\end{aligned}
$$

Además se asume que los errores no están correlacionados, es decir, $Cov\left(\epsilon_i, \epsilon_j\right) = 0$, con $i, j = 1, \ldots, n$, 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 $\boldsymbol{\epsilon} \sim \left(\mathbf{0}_n, \sigma^2 \mathbf{I}_n\right)$. 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 $\boldsymbol{\beta}$, 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

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

El estimdor de MCO de $\boldsymbol{\beta}$ es el vector $\hat{\beta}$ tal que $Q\left(\hat{\beta}\right)$ es mínima.
Primero se debe notar que

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

Entonces:

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

Estimación por MCO (cont.)

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

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

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 $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$, bajo las hipótesis:

- $\epsilon \sim \left(\mathbf{0}, \sigma^2\mathbf{I}\right)$.
- $\mathbf{X}$ 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 $\mathbf{v}'V\left(\tilde{\boldsymbol{\beta}}\right)\mathbf{v} \geq \mathbf{v}'V\left(\hat{\boldsymbol{\beta}}\right)\mathbf{v}$.

TGM: demostración

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

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

Entonces, $v\hat{b}et$ es insesgado para $\boldsymbol{\beta}$.

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

TGM: demostración

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

$$\begin{aligned} E\left(\tilde{\boldsymbol{\beta}}\right) &= E\left(\mathbf{AY}\right) \\ &= \mathbf{A}E\left(\mathbf{Y}\right) \\ &= \mathbf{AX}\beta. \end{aligned}$$

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

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

TGM: demostración

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

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

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

Valores ajustados, residuos y suma de cuadrados residual

El vector de valores ajustados se define como

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

La matriz $\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$ 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

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

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

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

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

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

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

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

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

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

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

Estimación por Máxima Verosimilitud

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

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

Entonces

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

La solución para σ^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

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

Propiedades de los EMV

Como el EMV de $\boldsymbol{\beta}$ coincide con el EMCO, se sigue que

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

Además, por el supuesto de normalidad, se sigue que

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

Para determinar las propiedades de $\hat{\sigma}^2$ es necesario presentar algunos resultados adicionales sobre distribuciones de formas cuadráticas, aunque debemos notar que

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

Distribución χ^2 no central

Definición (distribución χ^2 no central)

Se dice que una variable aleatoria (absolutamente) continua Y tiene una distribución χ^2 no central con parámetros ν y λ 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 χ^2 (central) con $\nu+j$ grados de libertad.
Lo anterior se denota como $Y \sim \chi^{*2}\,(\nu,\lambda)$, ν son grados de libertad y λ se conoce como parámetro de no centralidad.

Cuando $\lambda = 0$, la distribución χ^{*2} se reduce a la distribución χ^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 $\mathbf{AB} = \mathbf{0}$.
- $\mathbf{Y}'\mathbf{A}\mathbf{Y} \perp \mathbf{C}\mathbf{Y}$ si y sólo si $\mathbf{CA} = \mathbf{0}$.

Propiedades de los EMV

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

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

donde

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

Por lo tanto,

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

Propiedades de los EMV

De lo anterior se sigue que

$$
\begin{aligned}
E\left(\hat{\sigma}^2\right) &= \frac{n-p-1}{n}\sigma^2 \\
V\left(\hat{\sigma}^2\right) &= 2\frac{n-p-1}{n^2}\sigma^4.
\end{aligned}
$$

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

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

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

Debido a que

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

es sencillo hacer inferencias sobre σ^2.
Por ejemplo, un intervalo de confianza $100(1-\alpha)\,\%$ para está dado por

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

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

El intervalo anterior es sólo uno de los múltiples que se pueden construir. Cualquier par de cuantiles $\chi^2_{n-p-1}\,(1-\alpha_1)$ y $\chi^2_{n-p-1}\,(\alpha_2)$ 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 σ^2

Para contrastar las hipótesis

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

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

$$S < \chi^2_{n-p-1}(1-\alpha/2) \qquad \text{o} \qquad S > \chi^2_{n-p-1}(\alpha/2)\,.$$

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. $H_1 : \sigma^2 > \sigma_0^2$	$S > \chi^2_{n-p-1}(\alpha)$
$H_0 : \sigma^2 \geq \sigma_0^2$ vs. $H_1 : \sigma^2 < \sigma_0^2$	$S > \chi^2_{n-p-1}(1-\alpha)$

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

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

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

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \qquad \text{vs.} \qquad H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$$

para algún vector de constantes conocidas $\boldsymbol{\theta}_0$ con regla de decisión *rechazar* H_0 si

$$\boldsymbol{\theta}_0 \notin C$$

con una significancia α.

Inferencias para β

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

$$\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)' \mathbf{X}'\mathbf{X}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)/\sigma^2 \sim \chi^2_{p+1}.$$

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

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

Inferencias para β

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

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

Una prueba de hipótesis para contrastar

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \qquad \text{vs.} \qquad H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$$

consiste en rechazar H_0 con una significancia α si

$$\boldsymbol{\beta}_0 \notin C \quad \Leftrightarrow \quad \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)' \mathbf{X}'\mathbf{X} \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) > \hat{\sigma}^2_{MCO}(p+1)F_{p+1,n-p-1}\left(\alpha\right)$$

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

Inferencias para β

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

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

Entonces,

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

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

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

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

Inferencias para β

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

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

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

donde

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

y b es una constante conocida

Inferencias para β

Como casos particulares, cuando es de la forma $\mathbf{a} = \mathbf{I}_{(i+1)}$, entonces$\mathbf{a}'\boldsymbol{\beta} = \beta_i$, para $i = 0, \ldots, p$.
Un intervalo de confianza $100(1-\alpha)\,\%$ para β_i esta dado por

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

donde $\hat{\boldsymbol{\beta}}_i$ es la i-ésima entrada de $\hat{\boldsymbol{\beta}}$ y $(\mathbf{X}'\mathbf{X})^{-1}_{ii}$ es el i-ésimo elemento de la diagonal de $(\mathbf{X}'\mathbf{X})^{-1}$. Las pruebas de hipótesis para β_i se resumen en el siguiente cuadro

Hipótesis	Regla de rechazo
$H_0 : \beta_i = b$ vs. $H_1 : \beta_i \neq b$	$\lvert t\rvert > t_{n-p-a}(\alpha/2)$
$H_0 : \beta_i \geq b$ 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

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

y b es una constante conocida.

Intevalos de confianza y predicción

Otro caso particular son los intervalos de confianza para la respuesta media y los intervalos de predicción.
Supongamos que $\mathbf{x}^* = \left(x_1^*, \ldots, x_p^*\right)'$ 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 $E\left(Y|\mathbf{x}^*\right)$ está dado por

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

También es posible hacer pruebas de hipótesis sobre $E\left(Y|\mathbf{x}^*\right)$, 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

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

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

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

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

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

o equivalentemente

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

Es sencillo mostrar que

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

* 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

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

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

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

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

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