Capítulo 3

MODELOS DE REGRESIÓN

PRUEBA DE HIPÓTESIS E INTERVALOS DE CONFIANZA

En esta parte se desarrollan las pruebas de hipótesis e intervalos de confianzas referentes al modelo lineal general, en primer lugar se discute la prueba de la hipótesis $\boldsymbol{H\beta} = \boldsymbol{h}$, la cual es la hipótesis más general y a partir de ella se desarrollan algunas pruebas de hipótesis particulares. También proveemos un intervalo de confianza para σ^2. Asumiremos en todo el capítulo que $\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma^2\mathbf{I})$, donde $\boldsymbol{X}$ es $n \times p$, $p = k + 1$, de rango $p < n$.

3.1 Prueba de la Hipótesis $\boldsymbol{H\beta} = \boldsymbol{h}$

En esta sección se deriva y discute una prueba de la hipótesis $\boldsymbol{H\beta} = \boldsymbol{h}$ para el modelo lineal general. En las secciones siguientes se discuten varios casos especiales de esta prueba. Las pruebas que son generalmente de interés son aquellas que envuelven funciones lineales de los parámetros desconocidos β_i. A continuación algunos ejemplos

EJEMPLO 3.1

Considere un modelo lineal simple

$$Y_i = \beta_0 + \beta X_i + \varepsilon_i \qquad i = 1, 2, ..., n$$

los siguientes son ejemplos de pruebas que pueden ser de interés:

1. $\beta_0 = 0$ (o $\beta_0 = b_0$, donde b_0 es una constante dada); esta es una prueba de que el intercepto es cero (o igual a una constante b_0).

2. $\beta = 1$ (o $\beta = b$, donde b es una constante dada); esta es una prueba de que la pendiente es uno (o igual a una constante b).

EJEMPLO 3.2

Considere un modelo lineal simple

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i \qquad i = 1, 2, ..., n$$

los siguientes son ejemplos de pruebas que pueden ser de interés:

1. $\beta_1 = \beta_2$; una prueba de que β_1 sea igual a β_2, ignora los valores de los β_i restantes.
2. $\beta_1 = \beta_2$ y $\beta_3 = \beta_4$, una prueba de que β_1 sea igual a β_2 y β_3 sea igual a β_4 ignorando el valor de β_0
3. $\beta_1 = \beta_2 = 6$; esta es una prueba diferente a la 1, aquí la prueba es que β_1 y β_2 son ambos iguales a 6; en la parte 1 la prueba es que β_1 y β_2 son ambos iguales sin establecer cual es el valor común.
4. $\beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$; una prueba de que los β_j son simultaneamente iguales a cero.
5. $\beta_1 = \beta_2 = \beta_3 = \beta_4$; una prueba de que todos los β_j (excepto β_0) son iguales, pero no se establece cual es el valor común.
6. $\begin{matrix} \beta_1 - 2\beta_2 = 4\beta_3 \\ \beta_1 + 2\beta_2 = 6 \end{matrix}$ una prueba de 2 relaciones lineales especificas entre los β_j.

Una cosa a notar es que cada hipótesis es un caso especial de

$$\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}$$

donde $\boldsymbol{H}$ es una matriz $q \times p$ y $\boldsymbol{h}$ es un vector $q \times 1$. A continuación se muestra $\boldsymbol{H}$ y $\boldsymbol{h}$ para cada hipótesis del ejemplo 3.2. El vector $\boldsymbol{\beta}' = [\beta_0, \beta_1, \beta_2, \beta_3, \beta_4]$

1. $\boldsymbol{H} = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \end{bmatrix}$; $\boldsymbol{h} = 0$
2. $\boldsymbol{H} = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$; $\boldsymbol{h} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$
3. $\boldsymbol{H} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$; $\boldsymbol{h} = \begin{bmatrix} 6 \\ 6 \end{bmatrix}$
4. $\boldsymbol{H} = \mathbf{I}_5$; $\boldsymbol{h} = 0$
5. $\boldsymbol{H} = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \end{bmatrix}$; $\boldsymbol{h} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

6. $\boldsymbol{H} = \begin{bmatrix} 0 & 1 & -2 & -4 & 0 \\ 0 & 1 & 2 & 0 & 0 \end{bmatrix}; \boldsymbol{h} = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$

Nota 3.1 *Las matrices $\boldsymbol{H}$ y $\boldsymbol{h}$ no son necesariamente únicas. Por ejemplo considere la matriz del item 5. La misma puede escribirse como*

$$\boldsymbol{H} = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & 1 & -2 & 0 \\ 0 & 1 & 1 & 1 & -3 \end{bmatrix} \text{ ó } \boldsymbol{H} = \begin{bmatrix} 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

Las únicas restricciones que se consideran sobre $\boldsymbol{H}$ y $\boldsymbol{h}$ son la siguientes:

1. $\boldsymbol{H\beta} = \boldsymbol{h}$ es un conjunto consistente de ecuaciones.

2. La matriz $\boldsymbol{H}_{q\times p}$ tiene rango q. (esto es conveniente para que $\boldsymbol{HH}'$ sea de rango completo).

A continuación se deriva una prueba de $\boldsymbol{H\beta} = \boldsymbol{h}$ para el modelo lineal general $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ donde $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. En primer lugar se construye el estadístico de prueba, luego se identifica la distribución de dicho estadístico y por último se calcula la potencia de dicha prueba. Todo se resume en el teorema 3.1.

3.1.1 Cálculo del Estadístico de Prueba

Para calcular el estadístico de prueba se usa la razón de verosimilitud generalizada. Esta metodología compara el valor máximo de $L(\boldsymbol{\beta}, \sigma^2)$ restringido por $H_0 : \boldsymbol{H\beta} = \boldsymbol{h}$ con el valor máximo de $L(\boldsymbol{\beta}, \sigma^2)$ bajo la $H_1 : \boldsymbol{H\beta} \neq \boldsymbol{h}$, la cual en esencia no esta restringida. Por lo tanto, la razón de verosimilitud generalizada está dada por:

$$v(y) = \frac{\max\limits_{(\boldsymbol{\beta},\sigma^2)\in\omega}[L(\boldsymbol{\beta}, \sigma^2 : \boldsymbol{y})]}{\max\limits_{(\boldsymbol{\beta},\sigma^2)\in\Omega}[L(\boldsymbol{\beta}, \sigma^2 : \boldsymbol{y})]} = \frac{L(\hat{\omega})}{L(\hat{\Omega})} \tag{3.1}$$

donde los espacios muestrales Ω y ω están dados por:

$$\begin{aligned} \Omega &= \{(\boldsymbol{\beta}, \sigma^2) : \boldsymbol{\beta} \in E_p; \sigma^2 > 0\} \\ \omega &= \{(\boldsymbol{\beta}, \sigma^2) : \boldsymbol{\beta} \in E_p; \boldsymbol{H\beta} = \boldsymbol{h}; \sigma^2 > 0\} \end{aligned} \tag{3.2}$$

Es decir, Ω es el espacio de los parámetros bajo la hipótesis alternativa (sin resctricciones), y ω es el espacio de los parámetros bajo la hipótesis nula (restringido a que $\boldsymbol{H\beta} = \boldsymbol{h}$). Si $\boldsymbol{H\beta}$ es igual o cercano a $\boldsymbol{h}$, entonces $L(\hat{\omega})$ debe ser cercano a $L(\hat{\Omega})$. Si $L(\hat{\omega})$ no es cercano a $L(\hat{\Omega})$ concluimos que la muestra, $yv = (y_1, ..., y_n)'$ aparentemente no viene de una $N_n(\boldsymbol{X\beta}, \sigma^2\mathbf{I})$ con $\boldsymbol{H\beta} = \boldsymbol{h}$.

Recordemos que la función de verosimilitud $L(\boldsymbol{\beta}, \sigma^2 : \boldsymbol{y})$, esta dada por

$$L(\boldsymbol{\beta}, \sigma^2 : \boldsymbol{y}) = (2\pi)^{-n/2}(\sigma^2)^{-n/2}e^{-(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta})/2\sigma^2} \tag{3.3}$$

El denominador, $L(\hat{\Omega})$, se obtiene fácilmente, ya que este es la función de verosimilitud evaluada en el valor de máxima verosimilitud de los parámetros $\boldsymbol{\beta}$ y σ^2, cuando estos pueden variar en Ω. Dichos estimadores se obtuvieron en el capítulo anterior, representados por las ecuaciones 2.1.3 y 2.26 pero que en este caso se van a denotar por $\hat{\boldsymbol{\beta}}_{\boldsymbol{\Omega}}$ y $\hat{\sigma}^2_{\Omega}$, respectivamente. Por lo tanto sustituyendo estas dos ecuaciones en la ecuación 2.18 se obtiene que

$$L(\hat{\Omega}) = (2\pi)^{-n/2}(\hat{\sigma}^2_{\Omega})^{-n/2}e^{-n/2} \tag{3.4}$$

Para hallar el numerador, $L(\hat{\omega})$, se debe encontrar el máximo de la función de verosimilitud $L(\boldsymbol{\beta}, \sigma^2 : \boldsymbol{y}; X)$ con respecto a $\boldsymbol{\beta}$ y σ^2 cuando estos parámetros pueden variar en ω. Esto puede hacerse por dos métodos:

1. Al resolver $\boldsymbol{H\beta} = \boldsymbol{h}$ para $\boldsymbol{\beta}$, sustituyendo la restricción en la función de verosimilitud y maximizando la función resultante.

2. Al usar Multiplicadores de Lagrange y maximizar la función de verosimilitud sujeta a la restricción $\boldsymbol{H\beta} = \boldsymbol{h}$.

1. **Método de Sustitución.** Este método consiste en sustituir la restricción, $\boldsymbol{H\beta} = \boldsymbol{h}$, en el modelo $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, obteniéndose de esta manera un modelo nuevo conocido como el modelo reducido por la hipótesis. La cuestión ahora a resolver es como introducir la restricción en el modelo. Si la matriz $\boldsymbol{H}$ fuese una matriz cuadrada de rango completo, entonces tuviese inversa, con lo cual $\boldsymbol{H}^{-1}\boldsymbol{H} = \mathbf{I}$ y asi

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} = \boldsymbol{X}\mathbf{I}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \boldsymbol{XH}^{-1}\boldsymbol{H\beta} + \boldsymbol{\varepsilon} = \boldsymbol{XH}^{-1}\boldsymbol{h} + \boldsymbol{\varepsilon}$$

con lo cual se tendría el modelo nuevo con la restricción incluida. El problema es que la matriz $\boldsymbol{H}$ no es invertible. Para solucionar esto, se completa a $\boldsymbol{H}$ con otra matriz $\boldsymbol{G}$, de manera que la matriz completa $\begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{G} \end{bmatrix}$ tenga inversa. Recordemos que la matriz $\boldsymbol{H}$ es $q \times p$, de rango q, por lo tanto la matriz $\boldsymbol{G}$ debe ser $(p-q) \times p$ de rango $p-q$, $0 < q < p$, que además cumpla con la condición de que tal $\boldsymbol{HG}' = 0$, es decir que $\boldsymbol{G}$ es una completación ortogonal de $\boldsymbol{H}$. Siendo así, se tiene que la matriz $\begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{G} \end{bmatrix}$ es $p \times p$ de rango p. Por lo tanto se cumple que

$$\begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{G} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{G} \end{bmatrix} = \mathbf{I}$$

Recordemos que si una matriz $\boldsymbol{A}$ es invertible se cumple que $\boldsymbol{A}^{-1} = \boldsymbol{A}^{-}$, además por propiedades de inversa se tiene que

$$\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{bmatrix}^{-} = \begin{bmatrix} \boldsymbol{A}^{-} & \boldsymbol{B}^{-} \end{bmatrix}$$

Por lo tanto se tiene que

$$\mathbf{I} = \begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{G} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{G} \end{bmatrix} = \begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{G} \end{bmatrix}^{-} \begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{G} \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}^{-} & \boldsymbol{G}^{-} \end{bmatrix} \begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{G} \end{bmatrix} = \boldsymbol{H}^{-}\boldsymbol{H} + \boldsymbol{G}^{-}\boldsymbol{G}$$

Reescribiendo el modelo

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{X\beta} + \varepsilon = \boldsymbol{X} \begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{G} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{G} \end{bmatrix} \beta + \varepsilon = \boldsymbol{X}(\boldsymbol{H}^{-}\boldsymbol{H} + \boldsymbol{G}^{-}\boldsymbol{G})\beta + \varepsilon \\ &= \boldsymbol{XH}^{-}\boldsymbol{H}\beta + \boldsymbol{XG}^{-}\boldsymbol{G}\beta + \varepsilon \end{aligned}$$

sustituyendo $\boldsymbol{H\beta} = \boldsymbol{h}$, el modelo será

$$\boldsymbol{Y} = \boldsymbol{XH}^{-}\boldsymbol{h} + \boldsymbol{XG}^{-}\boldsymbol{G}\beta + \varepsilon$$

lo cual es equivalente a

$$\boldsymbol{Y} - \boldsymbol{XH}^{-}\boldsymbol{h} = \boldsymbol{XG}^{-}\boldsymbol{G}\beta + \varepsilon$$

Haciendo $\boldsymbol{Z} = \boldsymbol{Y} - \boldsymbol{XH}^{-}\boldsymbol{h}$, $\boldsymbol{B} = \boldsymbol{XG}^{-}$ y $\gamma = \boldsymbol{G\beta}$, el modelo se escribe como $\boldsymbol{Z} = \boldsymbol{B}\gamma + \varepsilon$, el cual es el modelo $\boldsymbol{Y} = \boldsymbol{X\beta} + \varepsilon$ con la restricción impuesta por H_0.

Este es un modelo lineal general ya que $\boldsymbol{Z}$ es un vector aleatorio $n \times 1$ observable, $\boldsymbol{B}$ es una matriz $n \times (p-q)$ no observables de rango $p-q$ (falta probarlo) y $\boldsymbol{\gamma}$ es un vector $(p-q) \times 1$ de parámetros desconocidos y puede ser cualquier vector en E_{p-q} y $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$.
El modelo reducido significa:"El modelo completo reducido por las restricciones de la hipótesis H_0".

Solo fata probar que $\boldsymbol{B} = \boldsymbol{XG}^{-}$ es una matriz $n \times (p-q)$ de rango $p-q$. Como $\boldsymbol{G}$ es una matriz de rango $(p-q) \times p$ de rango $p-q$, entonces $\boldsymbol{G}^{-} = \boldsymbol{G}'(\boldsymbol{GG}')^{-1}$, lo cual implica que $\boldsymbol{G}^{-}$ es $p \times (p-q)$, y por consiguiente $\boldsymbol{XG}^{-}$ es una matriz $n \times (p-q)$. Para el rango usamos el hecho de que el $r(G^{-}) = p-q$, $\boldsymbol{X}^{-}\boldsymbol{X} = \mathbf{I}$ y que $r(AB) \leq r(B)$, entonces

$$p - q = r(\boldsymbol{G}^{-}) = r(\boldsymbol{X}^{-}\boldsymbol{XG}^{-}) \leq r(\boldsymbol{XG}^{-}) \leq r(\boldsymbol{G}^{-}) = p - q$$

por lo tanto $r(\boldsymbol{XG}^{-}) = p - q$

Usando el nuevo modelo $\boldsymbol{Z} = \boldsymbol{B\gamma} + \boldsymbol{\varepsilon}$, los estimadores máximo verosímiles de $\boldsymbol{\gamma}$ y σ^2 están dados también en las ecuaciones 2.1.3 y 2.26 con $\boldsymbol{B}$ reemplazando a $\boldsymbol{X}$ y $\boldsymbol{Z}$ reemplazando a $\boldsymbol{Y}$. Ellos son denotados por $\hat{\gamma}_\omega$ y $\hat{\sigma}^2_\omega$, y están dados por

$$\hat{\gamma}_\omega = \boldsymbol{B}^{-}\boldsymbol{Z} \tag{3.5}$$

$$\hat{\sigma}^2_\omega = \frac{1}{n}\boldsymbol{Z}'(\mathbf{I} - \boldsymbol{BB}^{-})\boldsymbol{Z} \tag{3.6}$$

Entonces,

$$L(\hat{\omega}) = (2\pi)^{-n/2}(\hat{\sigma}^2_\omega)^{-n/2}e^{-n/2} \tag{3.7}$$

Por lo tanto, el estadístico de prueba de la razón de verosimilitud generalizada es

$$V(y) = V = \left(\frac{\hat{\sigma}^2_\Omega}{\hat{\sigma}^2_\omega}\right)^{n/2} \tag{3.8}$$

Nota 3.2 *Si $q = p$, entonces $\boldsymbol{B} = 0$ y el modelo reducido es $\boldsymbol{Z} = \boldsymbol{\varepsilon}$, donde $\boldsymbol{Z} = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{H}^{-1}\boldsymbol{h}$ y el espacio del parámetro ω es $\omega = \{(\boldsymbol{\beta}, \sigma^2) : \boldsymbol{\beta} = \boldsymbol{H}^1\boldsymbol{h} = b; \sigma^2 > 0\}$. Además $L(\hat{\omega}) = (2\pi)^{-n/2}(\hat{\sigma}^2_\omega)^{-n/2}e^{-n/2}$ donde $\hat{\sigma}^2_\omega = \frac{1}{n}\boldsymbol{Z}'\boldsymbol{Z} = \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{X}b)'(\boldsymbol{Y} - \boldsymbol{X}b)$.*

En lugar de V se usa un estadístico que es una función monótona de V, esto es, se usa W, donde $W = (V^{-2/n} - 1)\frac{(n-p)}{q}$, lo cual da:

$$W = \left(\frac{n-p}{q}\right)\left(\frac{\hat{\sigma}^2_\omega - \hat{\sigma}^2_\Omega}{\hat{\sigma}^2_\Omega}\right) = \left(\frac{n-p}{q}\right)\left[\frac{\boldsymbol{Z}'(\mathbf{I} - \boldsymbol{B}\boldsymbol{B}^-)\boldsymbol{Z} - \boldsymbol{Y}'(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Y}}{\boldsymbol{Y}'(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Y}}\right] \tag{3.9}$$

Terminología: Se llama W al estadístico de prueba de la razón de verosimilitud generalizada.(Aunque estrictamente debería ser es una función del estadístico de prueba de la razón de verosimilitud generalizada)

2. **El método de Lagrange**

   Este método consiste en construir una nueva función, llamada función de lagrange, la cual está formada por la función que se quiere optimizar y las restricciones que esta contiene. Una vez construida dicha función, el objetivo es hallar los valores que alcanzan el óptimo de la misma. En éste caso la función de lagrange está dada por:

$$\begin{aligned} L &= L(\boldsymbol{\beta}, \sigma^2 : \boldsymbol{y}; \boldsymbol{X}) - \boldsymbol{\lambda}'(\boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{h}) \\ &= (2\pi\sigma^2)^{-n/2}\exp\left[\frac{-1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right] - \boldsymbol{\lambda}'(\boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{h}) \\ &= (2\pi\sigma^2)^{-n/2}\exp\left[\frac{-1}{2\sigma^2}(\boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta})\right] - \boldsymbol{\lambda}'(\boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{h}) \end{aligned}$$

   Derivando esta expresión con respecto a $\boldsymbol{\beta}, \sigma^2$ y $\boldsymbol{\lambda}$ respectivamente se obtiene

$$\begin{aligned} \frac{\partial L}{\partial \boldsymbol{\beta}} &= L(\boldsymbol{\beta}, \sigma^2 : \boldsymbol{y}; \boldsymbol{X})\left[-(2\sigma^2)^{-1}(-\boldsymbol{y}'\boldsymbol{X} - \boldsymbol{X}'\boldsymbol{y} + \boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X})\right] - \boldsymbol{H}'\boldsymbol{\lambda} \\ &= L(\boldsymbol{\beta}, \sigma^2 : \boldsymbol{y}; \boldsymbol{X})\left[-(2\sigma^2)^{-1}(-2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta})\right] - \boldsymbol{H}'\boldsymbol{\lambda} \\ &= L(\boldsymbol{\beta}, \sigma^2 : \boldsymbol{y}; \boldsymbol{X})\left[-(2\sigma^2)^{-1}(2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} - 2\boldsymbol{X}'\boldsymbol{y})\right] - \boldsymbol{H}'\boldsymbol{\lambda} & (3.10) \\ \frac{\partial L}{\partial \sigma^2} &= -\frac{n}{2}(2\pi\sigma^2)^{-n/2-1}\exp\left[\frac{-1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right] \\ &\quad + L(\boldsymbol{\beta}, \sigma^2 : \boldsymbol{y}; \boldsymbol{X})(2\sigma^4)^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\ &= -n\pi(2\pi\sigma^2)^{-1}L(\boldsymbol{\beta}, \sigma^2) + (2\sigma^4)^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})L(\boldsymbol{\beta}, \sigma^2) & (3.11) \\ \frac{\partial L}{\partial \boldsymbol{\lambda}} &= -(\boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{h}) & (3.12) \end{aligned}$$

   Igualando esta derivadas a cero y denotando como $\tilde{\boldsymbol{\beta}}_\omega, \tilde{\sigma}^2_\omega$ y $\tilde{\boldsymbol{\lambda}}$ como las soluciones a las ecuaciones resultantes, se obtienen las siguientes expresiones (sea $\boldsymbol{\lambda}^* = \frac{\tilde{\sigma}^2_\omega\tilde{\boldsymbol{\lambda}}}{L(\tilde{\boldsymbol{\beta}}_\omega, \tilde{\sigma}^2_\omega : \boldsymbol{y}, \boldsymbol{x})}$, y note que $L(\boldsymbol{\beta}, \sigma^2 : \boldsymbol{y}, \boldsymbol{x}) > 0$):

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} & \boldsymbol{H}' \\ \boldsymbol{H} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\beta}}_\omega \\ \boldsymbol{\lambda}^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{y} \\ \boldsymbol{h} \end{bmatrix} \tag{3.13}$$

$$\tilde{\sigma}^2_\omega = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}_\omega)'(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}_\omega) \tag{3.14}$$

Resolviendo estas ecuaciones se tiene que

$$\boldsymbol{\lambda}^* = [\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\tilde{\beta}_\Omega - \boldsymbol{h}) \tag{3.15}$$

$$\tilde{\boldsymbol{\beta}}_\omega = \hat{\beta}_{\boldsymbol{\Omega}} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\tilde{\beta}_\Omega - \boldsymbol{h}) \tag{3.16}$$

$$\tilde{\sigma}^2_\omega = \tilde{\sigma}^2_\Omega + \left(\frac{1}{n}\right)(\boldsymbol{H}\tilde{\beta}_\Omega - \boldsymbol{h})[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\tilde{\beta}_\Omega - \boldsymbol{h}) \tag{3.17}$$

donde $\hat{\beta}_{\boldsymbol{\Omega}} = \boldsymbol{X}^-\boldsymbol{Y}$ es el Estimador máximo verosímil de $\boldsymbol{\beta}$ en el modelo completo, $\tilde{\sigma}^2_\Omega$ es el estimador máximo verosímil de σ^2 en el modelo completo. Si se sustituye $\tilde{\boldsymbol{\beta}}_\omega$ y $\tilde{\sigma}^2_\omega$ en la ecuación 3.9 y realizando ciertas operaciones algebraicas se obtiene la siguiente expresión para W.

$$W = \left(\frac{n-p}{q}\right)\left\{\frac{(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})}{\boldsymbol{Y}'(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Y}}\right\} \tag{3.18}$$

3.1.1.1 Equivalencia de los 2 estadísticos A continuación se demuestra que la fórmula para el estadístico de prueba obtenido con el método de selección, ecuación 3.9, es igual al obtenido con el método de Lagrange, ecuación 3.18. Para ello sólo basta con probar que

$$\boldsymbol{Z}'(\mathbf{I} - \boldsymbol{B}\boldsymbol{B}^-)\boldsymbol{Z} - \boldsymbol{Y}'(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Y} \tag{3.19}$$

en la ecuación 3.9 sea igual a

$$(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h}) \tag{3.20}$$

en la ecuación 3.18.

Para ello es necesario realizar algunas transformaciones en las dos ecuaciones. Comencemos con la ecuación 3.9, en la cual sustituyendo $\boldsymbol{Z} = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{H}^-\boldsymbol{h}, \boldsymbol{B} = \boldsymbol{X}\boldsymbol{G}^-$ en dicha ecuación y simplificando se obtiene que

$$\boldsymbol{Z}'(\mathbf{I} - \boldsymbol{B}\boldsymbol{B}^-)\boldsymbol{Z} - \boldsymbol{Y}'(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Y} = \boldsymbol{Z}'[\boldsymbol{X}\boldsymbol{X}^- - (\boldsymbol{X}\boldsymbol{G}^-)(\boldsymbol{X}\boldsymbol{G}^-)^-]\boldsymbol{Z} \tag{3.21}$$

Para obtener este resultado se procede de la siguiente manera. En principio note que como $\boldsymbol{Z} = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{H}^-\boldsymbol{h}$, entonces

$$\begin{aligned}
(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Z} &= (\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{H}^-\boldsymbol{h}) = (\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Y} - (\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)(\boldsymbol{X}\boldsymbol{H}^-\boldsymbol{h}) \\
&= (\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{H}^-\boldsymbol{h} - \boldsymbol{X}\boldsymbol{X}^-\boldsymbol{X}\boldsymbol{H}^-\boldsymbol{h} \\
&= (\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{H}^-\boldsymbol{h} - \boldsymbol{X}\boldsymbol{H}^-\boldsymbol{h} \\
&= (\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Y}
\end{aligned}$$

y como $(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)$ es idempotente y simétrica entonces

$$\begin{aligned}
\boldsymbol{Z}'(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Z} &= \boldsymbol{Z}'(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Z} \\
&= [(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Z}]'[(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Z}] \\
&= [(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Y}]'[(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Y}] \\
&= \boldsymbol{Y}'(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Y} \\
&= \boldsymbol{Y}'(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Y}
\end{aligned}$$

Por lo tanto,

$$\begin{aligned} \boldsymbol{Z}'(\mathbf{I}-\boldsymbol{B}\boldsymbol{B}^-)\boldsymbol{Z}-\boldsymbol{Y}'(\mathbf{I}-\boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Y} &= \boldsymbol{Z}'(\mathbf{I}-\boldsymbol{B}\boldsymbol{B}^-)\boldsymbol{Z}-\boldsymbol{Z}'(\mathbf{I}-\boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Z} \\ &= \boldsymbol{Z}'(\mathbf{I}-\boldsymbol{B}\boldsymbol{B}^- - \mathbf{I}+\boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Z} \\ &= \boldsymbol{Z}'(\boldsymbol{X}\boldsymbol{X}^- - \boldsymbol{B}\boldsymbol{B}^-)\boldsymbol{Z} \end{aligned}$$

Ahora sustituyendo $\boldsymbol{B}=\boldsymbol{X}\boldsymbol{G}^-$ en la ecuación anterior se obtiene la ecuación 3.21.
Con respecto a la ecuación 3.20, al sustituir $\hat{\boldsymbol{\beta}}=\boldsymbol{X}^-\boldsymbol{Y}$ y realizando algunas simplificaciones se obtiene que

$$(\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h})'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h})=\boldsymbol{Z}'(\boldsymbol{H}\boldsymbol{X}^-)^-(\boldsymbol{H}\boldsymbol{X}^-)\boldsymbol{Z} \tag{3.22}$$

Para obtener dicho resultado es necesario notar que

1. $(\boldsymbol{H}\boldsymbol{X}^-)(\boldsymbol{X}\boldsymbol{H}^-)=\mathbf{I}$;
2. $\boldsymbol{H}\boldsymbol{X}^-\boldsymbol{Y}-\boldsymbol{h}=\boldsymbol{H}\boldsymbol{X}^-(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{H}^-\boldsymbol{h})=\boldsymbol{H}\boldsymbol{X}^-\boldsymbol{Z}$;
3. $[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}=(\boldsymbol{H}\boldsymbol{X}^-)'^-(\boldsymbol{H}\boldsymbol{X}^-)^-$.
4. $(\boldsymbol{H}\boldsymbol{X}^-)^-(\boldsymbol{H}\boldsymbol{X}^-)$ es simetrica e idempotente.

Dichos resultados se muestran a continuación: Para el inciso 1., como la matriz $\boldsymbol{H}$ es $q\times p$ de rango q y $\boldsymbol{X}$ es $n\times p$, entonces $\boldsymbol{H}\boldsymbol{X}^-$ es una matriz $q\times n$ de rango q, lo cual implica que es de rango por filas y se cumple que $(\boldsymbol{H}\boldsymbol{X}^-)(\boldsymbol{X}\boldsymbol{H}^-)=\mathbf{I}$. El rango de $\boldsymbol{H}\boldsymbol{X}^-$ se obtiene de la siguiente manera

$$q=r(\boldsymbol{H})=r(\boldsymbol{H}\boldsymbol{X}^-\boldsymbol{X})\leq r(\boldsymbol{H}\boldsymbol{X}^-)\leq r(\boldsymbol{H})=q$$

para el inciso 2., se procede de la siguiente manera

$$\begin{aligned} \boldsymbol{H}\boldsymbol{X}^-\boldsymbol{Y}-\boldsymbol{h} &= \boldsymbol{H}\boldsymbol{X}^-\boldsymbol{Y}-\boldsymbol{H}\boldsymbol{X}^-\boldsymbol{X}\boldsymbol{H}^-\boldsymbol{h} \\ &= \boldsymbol{H}\boldsymbol{X}^-(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{H}^-\boldsymbol{h}) \\ &= \boldsymbol{H}\boldsymbol{X}^-\boldsymbol{Z} \end{aligned}$$

para el inciso 3., se tiene que

$$\begin{aligned} [\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1} &= [\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{H}']^- = [\boldsymbol{H}\boldsymbol{X}^-\boldsymbol{X}'^-\boldsymbol{H}']^- = [(\boldsymbol{H}\boldsymbol{X}^-)(\boldsymbol{X}'^-\boldsymbol{H}')]^- \\ &= [(\boldsymbol{H}\boldsymbol{X}^-)(\boldsymbol{H}\boldsymbol{X}^-)']^- = (\boldsymbol{H}\boldsymbol{X}^-)'^-(\boldsymbol{H}\boldsymbol{X}^-)^- \end{aligned}$$

y para el inciso 4., a continuación se evalua la simetría y la idempotencia.

- Simetría: Note que

$$\begin{aligned} [(\boldsymbol{H}\boldsymbol{X}^-)^-(\boldsymbol{H}\boldsymbol{X}^-)]' &= [(\boldsymbol{H}\boldsymbol{X}^-)'[(\boldsymbol{H}\boldsymbol{X}^-)(\boldsymbol{H}\boldsymbol{X}^-)']^{-1}(\boldsymbol{H}\boldsymbol{X}^-)]' \\ &= (\boldsymbol{H}\boldsymbol{X}^-)'[(\boldsymbol{H}\boldsymbol{X}^-)(\boldsymbol{H}\boldsymbol{X}^-)']^{-1}(\boldsymbol{H}\boldsymbol{X}^-)=(\boldsymbol{H}\boldsymbol{X}^-)^-(\boldsymbol{H}\boldsymbol{X}^-) \end{aligned}$$

- Idempotencia: Observe que:

$$(\boldsymbol{H}\boldsymbol{X}^-)^-\underbrace{(\boldsymbol{H}\boldsymbol{X}^-)(\boldsymbol{H}\boldsymbol{X}^-)^-}_{\mathbf{I}}(\boldsymbol{H}\boldsymbol{X}^-)=(\boldsymbol{H}\boldsymbol{X}^-)^-(\boldsymbol{H}\boldsymbol{X}^-)$$

Entonces

$$\begin{aligned}
(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h}) &= (\boldsymbol{H}\boldsymbol{X}^{-}\boldsymbol{Z})'(\boldsymbol{H}\boldsymbol{X}^{-})'^{-}(\boldsymbol{H}\boldsymbol{X}^{-})^{-}(\boldsymbol{H}\boldsymbol{X}^{-}\boldsymbol{Z}) \\
&= \boldsymbol{Z}'(\boldsymbol{H}\boldsymbol{X}^{-})'(\boldsymbol{H}\boldsymbol{X}^{-})'^{-}(\boldsymbol{H}\boldsymbol{X}^{-})^{-}\boldsymbol{H}\boldsymbol{X}^{-}\boldsymbol{Z} \\
&= \boldsymbol{Z}'[(\boldsymbol{H}\boldsymbol{X}^{-})'(\boldsymbol{H}\boldsymbol{X}^{-})'^{-}][(\boldsymbol{H}\boldsymbol{X}^{-})^{-}\boldsymbol{H}\boldsymbol{X}^{-}]\boldsymbol{Z} \\
&= \boldsymbol{Z}'[(\boldsymbol{H}\boldsymbol{X}^{-})^{-}\boldsymbol{H}\boldsymbol{X}^{-}]'[(\boldsymbol{H}\boldsymbol{X}^{-})^{-}\boldsymbol{H}\boldsymbol{X}^{-}]\boldsymbol{Z} \\
&= \boldsymbol{Z}'[(\boldsymbol{H}\boldsymbol{X}^{-})^{-}\boldsymbol{H}\boldsymbol{X}^{-}][(\boldsymbol{H}\boldsymbol{X}^{-})^{-}\boldsymbol{H}\boldsymbol{X}^{-}]\boldsymbol{Z} \\
&= \boldsymbol{Z}'(\boldsymbol{H}\boldsymbol{X}^{-})^{-}\boldsymbol{H}\boldsymbol{X}^{-}\boldsymbol{Z}
\end{aligned}$$

Por lo tanto, para probar que el estadístico en las ecuaciones 3.9 y 3.18 son los mismos, se debe mostrar que las ecuaciones 3.22 y 3.21 son las mismas, y esto significa que se debe probar que

$$\boldsymbol{X}\boldsymbol{X}^{-} - (\boldsymbol{X}\boldsymbol{G}^{-})(\boldsymbol{X}\boldsymbol{G}^{-})^{-} = (\boldsymbol{H}\boldsymbol{X}^{-})^{-}(\boldsymbol{H}\boldsymbol{X}^{-}) \tag{3.23}$$

lo cual es equivalente a probar que

$$\boldsymbol{X}\boldsymbol{X}^{-} = (\boldsymbol{X}\boldsymbol{G}^{-})(\boldsymbol{X}\boldsymbol{G}^{-})^{-} + (\boldsymbol{H}\boldsymbol{X}^{-})^{-}(\boldsymbol{H}\boldsymbol{X}^{-}) \tag{3.24}$$

Haciendo

$$\boldsymbol{N} = (\boldsymbol{X}\boldsymbol{G}^{-})(\boldsymbol{X}\boldsymbol{G}^{-})^{-} + (\boldsymbol{H}\boldsymbol{X}^{-})^{-}(\boldsymbol{H}\boldsymbol{X}^{-}) \tag{3.25}$$

entonces lo que hay que probar es que $\boldsymbol{N} = \boldsymbol{X}\boldsymbol{X}^{-}$. Es fácil ver que la matriz $\boldsymbol{N}$ es una matriz $n \times n$, simetrica, idempotente y de rango p. A continuación se demuestran cada una de las características de $\boldsymbol{N}$.

- Simetría: Observe que

$$\begin{aligned}
N' &= [(\boldsymbol{X}\boldsymbol{G}^{-})(\boldsymbol{X}\boldsymbol{G}^{-})^{-} + (\boldsymbol{H}\boldsymbol{X}^{-})^{-}(\boldsymbol{H}\boldsymbol{X}^{-})]' \\
&= [(\boldsymbol{X}\boldsymbol{G}^{-})(\boldsymbol{X}\boldsymbol{G}^{-})^{-}]' + [(\boldsymbol{H}\boldsymbol{X}^{-})^{-}(\boldsymbol{H}\boldsymbol{X}^{-})]'
\end{aligned}$$

ya se demostró que $[(\boldsymbol{H}\boldsymbol{X}^{-})^{-}(\boldsymbol{H}\boldsymbol{X}^{-})]' = (\boldsymbol{H}\boldsymbol{X}^{-})^{-}(\boldsymbol{H}\boldsymbol{X}^{-})$ solo falta probar que $[(\boldsymbol{X}\boldsymbol{G}^{-})(\boldsymbol{X}\boldsymbol{G}^{-})^{-}]' = (\boldsymbol{X}\boldsymbol{G}^{-})(\boldsymbol{X}\boldsymbol{G}^{-})^{-}$, para ello es necesario tener en cuenta que $\boldsymbol{X}\boldsymbol{G}^{-}$ es una matriz $n \times p - q$ de rango $p - q$ lo cual se muestra a continuación.

$$p - q = r(\boldsymbol{G}^{-}) = r(\boldsymbol{X}^{-}\boldsymbol{X}\boldsymbol{G}^{-}) \leq r(\boldsymbol{X}\boldsymbol{G}^{-}) \leq r(\boldsymbol{G}^{-}) = p - q$$

Por lo tanto, $\boldsymbol{X}\boldsymbol{G}^{-}$ es de rango por columnnas y asi tiene las siguientes propiedades $(\boldsymbol{X}\boldsymbol{G}^{-})^{-}\boldsymbol{X}\boldsymbol{G}^{-} = \mathbf{I}$ y $(\boldsymbol{X}\boldsymbol{G}^{-})^{-} = [(\boldsymbol{X}\boldsymbol{G}^{-})'(\boldsymbol{X}\boldsymbol{G}^{-})]^{-1}(\boldsymbol{X}\boldsymbol{G}^{-})'$. Entonces para la simetría se tiene que

$$\begin{aligned}
[(\boldsymbol{X}\boldsymbol{G}^{-})(\boldsymbol{X}\boldsymbol{G}^{-})^{-}]' &= [(\boldsymbol{X}\boldsymbol{G}^{-})[(\boldsymbol{X}\boldsymbol{G}^{-})'(\boldsymbol{X}\boldsymbol{G}^{-})]^{-1}(\boldsymbol{X}\boldsymbol{G}^{-})']' \\
&= (\boldsymbol{X}\boldsymbol{G}^{-})[(\boldsymbol{X}\boldsymbol{G}^{-})'(\boldsymbol{X}\boldsymbol{G}^{-})]^{-1}(\boldsymbol{X}\boldsymbol{G}^{-})' = (\boldsymbol{X}\boldsymbol{G}^{-})(\boldsymbol{X}\boldsymbol{G}^{-})^{-}
\end{aligned}$$

- Idempotencia: Se debe probar que $\boldsymbol{NN} = \boldsymbol{N}$, veamos

$$
\begin{aligned}
\boldsymbol{NN} &= [(\boldsymbol{XG}^-)(\boldsymbol{XG}^-)^- + (\boldsymbol{HX}^-)^-(\boldsymbol{HX}^-)][(\boldsymbol{XG}^-)(\boldsymbol{XG}^-)^- + (\boldsymbol{HX}^-)^-(\boldsymbol{HX}^-)] \\
&= (\boldsymbol{XG}^-)\underbrace{(\boldsymbol{XG}^-)^-(\boldsymbol{XG}^-)}_{\mathbf{I}}(\boldsymbol{XG}^-)^- + (\boldsymbol{XG}^-)(\boldsymbol{XG}^-)^-(\boldsymbol{HX}^-)^-(\boldsymbol{HX}^-) \\
&\quad + (\boldsymbol{HX}^-)^-(\boldsymbol{HX}^-)(\boldsymbol{XG}^-)(\boldsymbol{XG}^-)^- + (\boldsymbol{HX}^-)^-\underbrace{(\boldsymbol{HX}^-)(\boldsymbol{HX}^-)^-}_{\mathbf{I}}(\boldsymbol{H}\,X^-) \\
&= (\boldsymbol{XG}^-)(\boldsymbol{XG}^-)^- \\
&\quad + (\boldsymbol{XG}^-)[(\boldsymbol{XG}^-)'(\boldsymbol{XG}^-)]^{-1}(\boldsymbol{XG}^-)'(\boldsymbol{HX}^-)'[(\boldsymbol{HX}^-)(\boldsymbol{HX}^-)']^{-1}(\boldsymbol{HX}^-) \\
&\quad + (\boldsymbol{HX}^-)^-\boldsymbol{H}\underbrace{\boldsymbol{X}^-\boldsymbol{X}}_{\mathbf{I}}\boldsymbol{G}^-(\boldsymbol{XG}^-)^- + (\boldsymbol{HX}^-)^-(\boldsymbol{H}\,X^-) \\
&= (\boldsymbol{XG}^-)(\boldsymbol{XG}^-)^- \\
&\quad + (\boldsymbol{XG}^-)[(\boldsymbol{XG}^-)'(\boldsymbol{XG}^-)]^{-1}\boldsymbol{G}'^-\boldsymbol{X}'\boldsymbol{X}'^-\boldsymbol{H}'[(\boldsymbol{HX}^-)(\boldsymbol{HX}^-)']^{-1}(\boldsymbol{HX}^-) \\
&\quad + (\boldsymbol{HX}^-)^-\underbrace{\boldsymbol{HG}^-}_{\mathbf{0}}(\boldsymbol{XG}^-)^- + (\boldsymbol{HX}^-)^-(\boldsymbol{H}\,X^-) \\
&= (\boldsymbol{XG}^-)(\boldsymbol{XG}^-)^- \\
&\quad + (\boldsymbol{XG}^-)[(\boldsymbol{XG}^-)'(\boldsymbol{XG}^-)]^{-1}\boldsymbol{G}'^-\underbrace{(\boldsymbol{X}^-\boldsymbol{X})'}_{\mathbf{I}}\boldsymbol{H}'[(\boldsymbol{HX}^-)(\boldsymbol{HX}^-)']^{-1}(\boldsymbol{HX}^-) \\
&\quad + (\boldsymbol{HX}^-)^-(\boldsymbol{H}\,X^-) \\
&= (\boldsymbol{XG}^-)(\boldsymbol{XG}^-)^- \\
&\quad + (\boldsymbol{XG}^-)[(\boldsymbol{XG}^-)'(\boldsymbol{XG}^-)]^{-1}\underbrace{(\boldsymbol{HG}^-)'}_{\mathbf{0}}[(\boldsymbol{HX}^-)(\boldsymbol{HX}^-)']^{-1}(\boldsymbol{HX}^-) \\
&\quad + (\boldsymbol{HX}^-)^-(\boldsymbol{H}\,X^-) \\
&= (\boldsymbol{XG}^-)(\boldsymbol{XG}^-)^- + (\boldsymbol{HX}^-)^-(\boldsymbol{H}\,X^-) = \boldsymbol{N}
\end{aligned}
$$

- Rango de $\boldsymbol{N}$. Como $\boldsymbol{N}$ es una matriz simétrica e idempotente entonces

$$
\begin{aligned}
r(\boldsymbol{N}) &= r[(\boldsymbol{XG}^-)(\boldsymbol{XG}^-)^- + (\boldsymbol{HX}^-)^-(\boldsymbol{H}\,X^-)] \\
&= tr[(\boldsymbol{XG}^-)(\boldsymbol{XG}^-)^- + (\boldsymbol{HX}^-)^-(\boldsymbol{H}\,X^-)] \\
&= tr[(\boldsymbol{XG}^-)(\boldsymbol{XG}^-)^-] + tr[(\boldsymbol{HX}^-)^-(\boldsymbol{H}\,X^-)]
\end{aligned}
$$

  Además, como $(\boldsymbol{XG}^-)(\boldsymbol{XG}^-)^-$ y $(\boldsymbol{HX}^-)^-(\boldsymbol{H}\;X^-)$ son matrices simétricas e idempotentes de rango $p-q$ y q respectivamente, entonces

$$
\begin{aligned}
r(\boldsymbol{N}) &= r[(\boldsymbol{XG}^-)(\boldsymbol{XG}^-)^-] + r[(\boldsymbol{HX}^-)^-(\boldsymbol{H}\,X^-)] \\
&= p - q + q = p
\end{aligned}
$$

Para probar que $\boldsymbol{N} = \boldsymbol{XX}^-$ es necesario definir la siguiente matriz

$$\boldsymbol{A} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'(\boldsymbol{HX}^-)'^- + \boldsymbol{G}^-(\boldsymbol{XG}^-)^- \tag{3.26}$$

la cual tiene las siguientes características

1. $\boldsymbol{XA} = \boldsymbol{N}$;
2. $\boldsymbol{A}$ tiene rango p;
3. $\boldsymbol{XAX} = \boldsymbol{X}$.

A continuación se comprueban las 3 características.

1. $\boldsymbol{XA} = \boldsymbol{N}$. Veamos que

$$\begin{aligned}
\boldsymbol{XA} &= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'(\boldsymbol{HX}^-)'^- + \boldsymbol{XG}^-(\boldsymbol{XG}^-)^- \\
&= [(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']'\boldsymbol{H}'(\boldsymbol{HX}^-)'^- + \boldsymbol{XG}^-(\boldsymbol{XG}^-)^- \\
&= \boldsymbol{X}'^-\boldsymbol{H}'(\boldsymbol{HX}^-)'^- + \boldsymbol{XG}^-(\boldsymbol{XG}^-)^- \\
&= (\boldsymbol{HX}^-)'(\boldsymbol{HX}^-)'^- + \boldsymbol{XG}^-(\boldsymbol{XG}^-)^- \\
&= [(\boldsymbol{HX}^-)^-(\boldsymbol{HX}^-)]' + \boldsymbol{XG}^-(\boldsymbol{XG}^-)^- \\
&= (\boldsymbol{HX}^-)^-(\boldsymbol{HX}^-) + \boldsymbol{XG}^-(\boldsymbol{XG}^-)^- = \boldsymbol{N}
\end{aligned}$$

2. $\boldsymbol{A}$ tiene rango p. En este caso

$$p = r(\boldsymbol{N}) = r(\boldsymbol{XA}) \leq r(\boldsymbol{A}) = r(\boldsymbol{X}^- X\boldsymbol{A}) \leq r(\boldsymbol{XA}) = r(\boldsymbol{N}) = p$$

3. $\boldsymbol{XAX} = \boldsymbol{X}$.

$$\boldsymbol{XAX} = \boldsymbol{XAXAA}^- = \boldsymbol{NNA}^- = \boldsymbol{NA}^- = \boldsymbol{XAA}^- = \boldsymbol{X}$$

Por lo tanto, se puede probar que $\boldsymbol{N} = \boldsymbol{XX}^-$ ya que

$$\begin{aligned}
\boldsymbol{XX}^- &= (\boldsymbol{XX}^-)' = [(\boldsymbol{XAX})\boldsymbol{X}^-]' = [(\boldsymbol{XA})(\boldsymbol{XX}^-)]' = (\boldsymbol{XX}^-)'(\boldsymbol{XA})' \\
&= (\boldsymbol{XX}^-)'\boldsymbol{N}' = (\boldsymbol{XX}^-)\boldsymbol{N} = \boldsymbol{XX}^-\boldsymbol{XA} = \boldsymbol{XA} = \boldsymbol{N}
\end{aligned}$$

Esto prueba que W en la ecuación 3.9 es el mismo W en la ecuación 3.18.

3.1.2 Distribución del Estadístico de Prueba

La distribución del Estadístico de Prueba puede ser la distribución del estadístico obtenido con el método de sustitución o con el método de Lagrange, pues como estos dos estadísticos son equivalentes, la distribución es la misma. A continuación se calcula la distribución del estadístico obtenido con el método de sustitución. Recuerde que el estadístico de prueba por dicho método está dado por la ecuación 3.9. Ya que

$$\boldsymbol{Y}'(\mathbf{I} - \boldsymbol{XX}^-)\boldsymbol{Y} = \boldsymbol{Z}'(\mathbf{I} - \boldsymbol{XX}^-)\boldsymbol{Z}$$

Entonces el estadístico de prueba se puede escribir como

$$W = \frac{\left(\frac{1}{q}\right)\boldsymbol{Z}'\left[\frac{\boldsymbol{XX}^- - \boldsymbol{BB}^-}{\sigma^2}\right]\boldsymbol{Z}}{\left[\frac{1}{n-p}\right]\boldsymbol{Z}'\left(\frac{\mathbf{I}-\boldsymbol{XX}^-}{\sigma^2}\right)\boldsymbol{Z}} = \frac{\left(\frac{1}{q}\right)\boldsymbol{Z}'\boldsymbol{A}_1\boldsymbol{Z}}{\left(\frac{1}{n-p}\right)\boldsymbol{Z}'\boldsymbol{A}_2\boldsymbol{Z}} = \frac{\left(\frac{1}{q}\right)U_1}{\left(\frac{1}{n-p}\right)U_2}$$

Dado que

- $\boldsymbol{Z} \sim N(\boldsymbol{X\beta} - \boldsymbol{XH}^{-}\boldsymbol{h}; \sigma^2\mathbf{I})$
- $\boldsymbol{A}_1(\sigma^2\mathbf{I}) = \boldsymbol{XX}^{-} - \boldsymbol{BB}^{-}$ es idempotente de rango q, y
- $\boldsymbol{A}_2(\sigma^2\mathbf{I}) = \mathbf{I} - \boldsymbol{XX}^{-}$ es idempotente de rango $n-p$

se cumple que

- $U_1 \sim \chi^2(q, \lambda)$, donde

$$\lambda = \frac{1}{2\sigma^2}[E(\boldsymbol{Z})]'(\boldsymbol{XX}^{-} - \boldsymbol{BB}^{-})[E(\boldsymbol{Z})]$$

  como $E(\boldsymbol{Z}) = \boldsymbol{X\beta} - \boldsymbol{XH}^{-}\boldsymbol{h}$ y $\boldsymbol{XX}^{-} - \boldsymbol{BB}^{-} = (\boldsymbol{HX}^{-})^{-}(\boldsymbol{HX}^{-})$ entonces

$$\begin{aligned} \lambda &= \frac{1}{2\sigma^2}[\boldsymbol{X\beta} - \boldsymbol{XH}^{-}\boldsymbol{h}]'[(\boldsymbol{HX}^{-})^{-}(\boldsymbol{HX}^{-})][\boldsymbol{X\beta} - \boldsymbol{XH}^{-}\boldsymbol{h}] \\ &= \end{aligned}$$

- $U_2 \sim \chi^2(n-p, \lambda_2)$ Claramente el parámetro de no centralidad es cero pues

$$\begin{aligned} \lambda_2 &= \frac{1}{2\sigma^2}[E(\boldsymbol{Z})]'(\mathbf{I} - \boldsymbol{XX}^{-})[E(\boldsymbol{Z})] \\ &= \frac{1}{2\sigma^2}[\boldsymbol{X\beta} - \boldsymbol{XH}^{-}\boldsymbol{h}]'(\mathbf{I} - \boldsymbol{XX}^{-})[\boldsymbol{X\beta} - \boldsymbol{XH}^{-}\boldsymbol{h}] \\ &= \frac{1}{2\sigma^2}[\boldsymbol{X}(\boldsymbol{\beta} - \boldsymbol{H}^{-}\boldsymbol{h})]'(\mathbf{I} - \boldsymbol{XX}^{-})\boldsymbol{X}(\boldsymbol{\beta} - \boldsymbol{H}^{-}\boldsymbol{h}) = 0 \end{aligned}$$

  ya que $(\mathbf{I} - \boldsymbol{XX}^{-})\boldsymbol{X} = \boldsymbol{X} - \boldsymbol{X}\overbrace{\boldsymbol{X}^{-}\boldsymbol{X}}^{\mathbf{I}} = \boldsymbol{X} - \boldsymbol{X} = 0$

Adémas, es fácil demostrar que U_1 y U_2 son independientes, (Tarea), entonces

$$W \sim F(q, n-p, \lambda)$$

donde

$$\lambda = \frac{1}{2\sigma^2}(\boldsymbol{H\beta} - \boldsymbol{h})'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H\beta} - \boldsymbol{h})$$

y $\lambda = 0$ si y sólo si $\boldsymbol{H\beta} - \boldsymbol{h} = 0$, esto es si y sólo si H_0 es cierta ya que $\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'$ es una matriz definida positiva.

Teorema 3.1 *En el modelo lineal general $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ donde $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$, W es un estadístico de prueba de la razón de verosimilitud generalizada para probar la hipótesis*

$$\begin{aligned} H_0 &: \boldsymbol{H\beta} = \boldsymbol{h} \\ H_1 &: \boldsymbol{H\beta} \neq \boldsymbol{h} \end{aligned}$$

donde W está dado en las 2 formas siguientes

Forma 1

$$\begin{aligned} W &= \frac{\frac{1}{q}(\boldsymbol{H\hat{\beta}} - \boldsymbol{h})'[Cov(\boldsymbol{H\beta} - \boldsymbol{h})]^{-1}(\boldsymbol{H\hat{\beta}} - \boldsymbol{h})}{\frac{\hat{\sigma}^2}{\sigma^2}} \\ &= \left(\frac{n-p}{q}\right)\left\{\frac{(\boldsymbol{H\hat{\beta}} - \boldsymbol{h})'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H\hat{\beta}} - \boldsymbol{h})}{\boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{XX}^{-})\boldsymbol{Y}}\right\} \\ &= \frac{1}{q}(\boldsymbol{H\hat{\beta}} - \boldsymbol{h})'[\widehat{Cov}(\boldsymbol{H\hat{\beta}} - \boldsymbol{h})]^{-1}(\boldsymbol{H\hat{\beta}} - \boldsymbol{h}) \end{aligned} \tag{3.27}$$

donde $\hat{\boldsymbol{\beta}} = \boldsymbol{X}^{-}\boldsymbol{Y}$ y $\hat{\sigma}^2 = \frac{1}{n-p}\boldsymbol{Y}'(\boldsymbol{I}-\boldsymbol{X}\boldsymbol{X}^{-})\boldsymbol{Y}$ son los EIMV de $\boldsymbol{\beta}$ y σ^2, respectivamente.

Forma 2

$$W = \left(\frac{n-p}{q}\right)\left(\frac{\hat{\sigma}^2_{\omega} - \hat{\sigma}^2_{\Omega}}{\hat{\sigma}^2_{\Omega}}\right) \tag{3.28}$$

donde

$$\hat{\sigma}^2_{\Omega} = \frac{1}{n}\boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^{-})\boldsymbol{Y}' = \frac{1}{n}(\boldsymbol{Y}'\boldsymbol{Y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y})$$

es el estimador máximo verosímil de σ^2 en el modelo completo $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, y donde

$$\hat{\sigma}^2_{\omega} = \frac{1}{n}\boldsymbol{Z}'(\boldsymbol{I} - \boldsymbol{B}\boldsymbol{B}^{-})\boldsymbol{Z}' = \frac{1}{n}(\boldsymbol{Z}'\boldsymbol{Z} - \hat{\gamma}'\boldsymbol{B}'\boldsymbol{Z})$$

es el estimador máximo verosímil de σ^2 en el modelo reducido $\boldsymbol{Z} = \boldsymbol{B\gamma} + \boldsymbol{\varepsilon}$, (donde $\boldsymbol{Z} = \boldsymbol{B\gamma} + \boldsymbol{\varepsilon}$ es el modelo completo reducido por H_0)

Además, $W \sim F(q, n-p, \lambda)$ donde

$$\lambda = \frac{1}{2\sigma^2}(\boldsymbol{H\beta} - \boldsymbol{h})'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H\beta} - \boldsymbol{h})$$

La prueba de la razón de verosimilitud generalizada es como sigue: **Rechazar H_0 si y sólo si w satisface que $w > F_{\alpha,q,n-p}$ donde $F_{\alpha,q,n-p}$ es el punto de la distribución F cuya cola derecha es α con q y $n-p$ grados de libertad; w es el valor calculado de W.**
Las ecuaciones 3.27 y 3.28 son las fórmulas básicas para hallar pruebas estadísticas, y aún más importante, para hallar intervalos de confianza sobre funciones lineales de los β_i.

3.1.3 Potencia de la Prueba

Recuerde que la potencia de una prueba, $\Pi(\lambda)$, se define como la probabilidad de rechazar la hipótesis nula cuando ella es falsa, es decir

$$\Pi(\lambda) \quad = \quad P[\text{Rechazar } H_0 | H_0 \text{ es falsa}]$$

por lo tanto, la potencia de esta prueba está dada por

$$\begin{aligned} \Pi(\lambda) &= P[\text{Rechazar } H_0 | H_0 \text{ es falsa}] = P[W > F_{\alpha,q,n-p} / \boldsymbol{H\beta} \neq \boldsymbol{h}] \\ &= \int_{F_{\alpha,q,n-p}}^{\infty} F(w : q, n-p, \lambda) dw \end{aligned}$$

Por lo tanto, para determinar la potencia de la prueba es necesario conocer $P[W > F_{\alpha,q,n-p} | \boldsymbol{H\beta} \neq \boldsymbol{h}]$. Existen diversas maneras de obtener dicha probabilidad

- Usando los gráficos de las cartas de Pearson-Heartley, las cuales se encuentran en Stapleton (1995).
- Usando las tablas de la distribución F no central disponibles en Graybill (1976).
- Usando cualquier paquete estadístico (en este caso se recomienda R).

En cualquiera de los dos primeros casos es necesario calcular la siguiente cantidad

$$\phi = \sqrt{\frac{2\lambda}{q+1}}$$

Para las tablas en Graybill y Para los gráficos en Stapleton.
Para hacerlo en R se usa la siguiente instrucción

pf(q=qf(p=α, df1=, df2=, lower.tail = FALSE, log.p = FALSE), df1=, df2=, ncp=λ,
lower.tail = TRUE, log.p = FALSE)

3.1.4 Expresión alterna para $\boldsymbol{W}$

Una expresión alterna para W se obtiene al definir $\boldsymbol{\theta} = \boldsymbol{H\beta} - \boldsymbol{h}$. Con esta definición entonces la hipótesis a probar es

$$\begin{aligned} H_0 &: \boldsymbol{\theta} = \mathbf{0} \\ H_1 &: \boldsymbol{\theta} \neq \mathbf{0} \end{aligned}$$

Por el teorema 2.13 el estimador máximo verosímil de $\boldsymbol{\theta}$ es $\hat{\boldsymbol{\theta}} = \boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h}$. Además $\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \sigma^2 \boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}')$.

Por lo tanto, el estadístico de prueba obtenido por le método de Lagrange se puede escribir como

$$W = \left(\frac{n-p}{q}\right)\left\{\frac{\hat{\boldsymbol{\theta}}'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}\hat{\boldsymbol{\theta}}}{\boldsymbol{Y}'(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^{-})\boldsymbol{Y}}\right\} \tag{3.29}$$

En cambio, el estadístico obtenido por el método de sustitución permanece sin cambio.

3.2 Pruebas de hipótesis comúnmente usadas

A continuación se presenta el desarrollo de las pruebas de algunas hipótesis que son comúnmente usadas en el análisis de un conjunto de datos.

3.2.1 Prueba sobre un subconjunto de $\boldsymbol{\beta}$

De manera general, supongamos que se desea probar la hipótesis de que un subconjunto de los x no son usados en la predicción de y. Un simple ejemplo es $H_0 : \beta_j = 0$ para un solo β_j. Si H_0 es rechazada, deberíamos mantener a $\beta_j x_j$ en el modelo. Como otra ilustración, considere el modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \varepsilon$$

para el cual deseamos probar la hipótesis $H_0 : \beta_3 = \beta_4 = \beta_5 = 0$. Si H_0 es rechazada, debemos seleccionar el modelo completo de segundo orden en vez del modelo reducido de primer orden.

Matricialmente la hipótesis es planteada de la siguiente manera:

$$H_0 : \boldsymbol{\beta}_2 = \begin{pmatrix} \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = \mathbf{0}$$

Por lo tanto la hipótesis esta particionando al vector $\boldsymbol{\beta}$ en dos partes como se muestra a continuación

$$H_0 : \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \cdots \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}$$

Dicha partición genera la siguiente partición de la matriz $\boldsymbol{X}$

$$\boldsymbol{X} = \left(\begin{array}{ccc|ccc} 1 & x_{11} & x_{12} & x_{11}^2 & x_{12}^2 & x_{11}x_{12} \\ 1 & x_{21} & x_{22} & x_{21}^2 & x_{22}^2 & x_{21}x_{22} \\ 1 & x_{31} & x_{32} & x_{31}^2 & x_{32}^2 & x_{31}x_{32} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n1}^2 & x_{n2}^2 & x_{n1}x_{n2} \end{array}\right) = \begin{pmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 \end{pmatrix}$$

Por lo tanto dicha hipotesis esta generando la siguiente partición en el modelo

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \boldsymbol{\varepsilon} \\ &= \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \end{aligned} \tag{3.30}$$

Volviendo a la generalidad, supongamos que se desea probar la hipótesis $H_0 : \boldsymbol{\beta}_2 = b_2$, donde $\boldsymbol{\beta}_2$ es un vector que contiene los últimos q elementos del vector $\boldsymbol{\beta}$, $0 < q < p$. Dicha hipótesis genera entonces la siguiente partición del vector $\boldsymbol{\beta}$

$$\boldsymbol{\beta} = \left(\begin{array}{c} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-q-1} \\ \hline \beta_{p-q} \\ \vdots \\ \beta_p \end{array}\right) = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}$$

Dicha partición genera la siguiente partición de la matriz $\boldsymbol{X}$

$$\boldsymbol{X} = \left(\begin{array}{cccc|ccc} 1 & x_{11} & \cdots & x_{1(p-q-1)} & x_{1(p-q)} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2(p-q-1)} & x_{2(p-q)} & \cdots & x_{2p} \\ 1 & x_{31} & \cdots & x_{3(p-q-1)} & x_{3(p-q)} & \cdots & x_{3p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ 1 & x_{21} & \cdots & x_{2(p-q-1)} & x_{2(p-q)} & \cdots & x_{2p} \end{array}\right) = \begin{pmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 \end{pmatrix}$$

Sin pérdida de generalidad, asumamos que los β's incluidos en la hipótesis se pueden ordenar de mas manera que sean los último elementos del vector $\boldsymbol{\beta}$, con su correspondiente arreglo de la matriz $\boldsymbol{X}$. **Nota: Esto debe quedar claro, por lo tanto hay que mostrar algunos ejemplos en clases**. Entonces $\boldsymbol{\beta}$ y $\boldsymbol{X}$ se pueden particionar convenientemente, y por lo tanto el modelo sería el siguiente

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \boldsymbol{\varepsilon} \\ &= \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \end{aligned} \tag{3.31}$$

donde $\boldsymbol{\beta}_2$ contiene los β a ser probados. El intercepto β_0 por lo general se incluye en $\boldsymbol{\beta}_1$.

La hipótesis de interés es $H_0 : \boldsymbol{\beta}_2 = 0$. Si designamos el número de parámetros en $\boldsymbol{\beta}_2$ por h, entonces $\boldsymbol{X}_2$ es $n \times h$, $\boldsymbol{\beta}_1$ es $(k-h+1) \times 1$, y $\boldsymbol{X}_1$ es $n \times (k-h+1)$. Por lo tanto, $\boldsymbol{\beta}_1 = (\beta_0, \beta_1, \cdots, \beta_{k-h})'$ y $\boldsymbol{\beta}_2 = (\beta_{k-h+1}, \cdots, \beta_k)'$. En términos de la ilustración al inicio de esta sección, deberíamos tener $\boldsymbol{\beta}_1 = (\beta_0, \beta_1, \beta_2)'$ y $\boldsymbol{\beta}_2 = (\beta_3, \beta_4, \beta_5)'$. Note que $\boldsymbol{\beta}_1$ en esta sección es diferente al $\boldsymbol{\beta}_1$ de la sección anterior, en el cual $\boldsymbol{\beta}$ fue particionado como $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \boldsymbol{\beta}_1 \end{pmatrix}$ y $\boldsymbol{\beta}_1$ constituido por todos los β excepto por β_0.

Para probar $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ versus $H_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$, usamos una técnica modelo-completo-reducido. El modelo completo esta dado por (3.40). Bajo $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$, el modelo reducido sería

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 * + \boldsymbol{\varepsilon} * \tag{3.32}$$

Usamos la notación $\boldsymbol{\beta}_1^*$ y $\boldsymbol{\varepsilon}^*$, ya que en el modelo reducido, $\boldsymbol{\beta}_1^*$ y $\boldsymbol{\varepsilon}^*$ serán típicamente diferentes a $\boldsymbol{\beta}_1$ y $\boldsymbol{\varepsilon}$ en el modelo completo (a menos que $\boldsymbol{X}_1$ y $\boldsymbol{X}_2$ sean ortogonales, ver teorema 7.9a y su corolario, Rencher). El estimador de $\boldsymbol{\beta}_1^*$ en el modelo reducido (3.41) es $\hat{\boldsymbol{\beta}}_1^* = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{y}$, el cual es, en general, no el mismo como los primeros $k-h+1$ elementos de $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ del modelo completo (3.40)(a menos que $\boldsymbol{X}_1$ y $\boldsymbol{X}_2$ sean ortogonales, ver teorema 7.10, Rencher)

Para comparar el ajuste del modelo completo (3.40) con el ajuste del modelo reducido (3.41), sumamos y restamos $\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}$ y $\hat{\boldsymbol{\beta}}_1^*\boldsymbol{X}_1'\boldsymbol{y}$ a la suma de cuadrados total corregida $\sum_{i=1}^n (y_i - \bar{y})^2 = \boldsymbol{y}'\boldsymbol{y} - n\bar{y}^2$ de manera de obtener la partición

$$\boldsymbol{y}'\boldsymbol{y} - n\bar{y}^2 = (\boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}) + (\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}_1^{*\prime}\boldsymbol{X}_1'\boldsymbol{y}) + (\hat{\boldsymbol{\beta}}_1^{*\prime}\boldsymbol{X}_1'\boldsymbol{y} - n\bar{y}^2) \tag{3.33}$$

o

$$SCT = SCE + SC(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) + SC(reducido) \tag{3.34}$$

donde $SC(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}_1^{*\prime}\boldsymbol{X}_1'\boldsymbol{y}$ es la suma de cuadrados de regresión "extra" debido a $\boldsymbol{\beta}_2$ luego de ajustar por $\boldsymbol{\beta}_1$. Note que $SC(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)$ puede también ser expresada como

$$\begin{aligned} SC(\boldsymbol{\beta}_1|\boldsymbol{\beta}_1) &= \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - n\bar{y}^2 - (\hat{\boldsymbol{\beta}}_1^{*\prime}\boldsymbol{X}_1'\boldsymbol{y} - n\bar{y}^2) \\ &= SCR(completo) - SCR(reducido) \end{aligned}$$

lo cual es la diferencia entre la suma de cuadrados de regresión para el modelo completo y la suma de cuadrados de regresión para el modelo reducido.

3.3 Prueba General de la regresión

La prueba de hipótesis es una herramienta formal para, entre otras cosas, seleccionar entre un modelo reducido y un modelo completo asociado. La hipótesis H_0 expresa el modelo reducido en términos de valores de un subconjunto de los β_j en $\boldsymbol{\beta}$. La hipótesis alternativa, H_1, esta asociada con el modelo completo.

Para ilustrar esta herramienta comenzamos con una prueba común, la prueba de la hipótesis general de regresión de que ninguna de las x predice a y. Esta hipótesis (que conduce al modelo reducido) puede expresarse como $H_0 : \boldsymbol{\beta}_1 = 0$, donde $\boldsymbol{\beta}_1 = (\beta_1, \beta_2, ..., \beta_k)'$. Note que deseamos probar $H_0 : \boldsymbol{\beta}_1 = 0$, no $H_0 : \boldsymbol{\beta} = 0$, donde

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \boldsymbol{\beta}_1 \end{pmatrix}$$

Ya que β_0 es usualmente diferente de cero, no es de interés incluir $\beta_0 = 0$ en la hipótesis. El rechazo de $H_0 : \boldsymbol{\beta} = 0$ podría ser debido solamente a β_0, y podría no dejarnos conocer si las variables x predicen a y.

Procedemos proponiendo un estadístico de prueba que se distribuye como una F central si H_0 es cierta y como una F no central en el caso contrario. Nuestro procedimiento para obtener un estadístico de prueba.

La suma de cuadrados total corregida $SCT = \sum_{i=1}^{n}(y_i - \bar{y})^2$ se puede particionar como

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \tag{3.35}$$

$$\begin{aligned} \sum_{i=1}^{n}(y_i - \bar{y})^2 &= \hat{\boldsymbol{\beta}}_1' \boldsymbol{X}_c' \boldsymbol{y} + \left[\sum_{i=1}^{n}(y_i - \bar{y})^2 - \hat{\boldsymbol{\beta}}_1' \boldsymbol{X}_c' \boldsymbol{y} \right] \\ &= \hat{\boldsymbol{\beta}}_1' \boldsymbol{X}_c' \boldsymbol{X}_c \hat{\boldsymbol{\beta}}_1 + SCE = SCR + SCE \end{aligned} \tag{3.36}$$

donde $SCE = \sum_{i=1}^{n}(y_i - \bar{y})^2 - \hat{\boldsymbol{\beta}}_1' \boldsymbol{X}_c' \boldsymbol{y}$. La suma de cuadrados de regresión $SCR = \hat{\boldsymbol{\beta}}_1' \boldsymbol{X}_c' \boldsymbol{X}_c \hat{\boldsymbol{\beta}}_1$ es claramente debido a $\boldsymbol{\beta}_1$.

Para construir una prueba F, primero expresamos las sumas de cuadrado en (3.36) como formas cuadráticas en $\boldsymbol{y}$ y demostramos que SCR y SCR tienen distribución chi-cuadrado y son independientes. Usando $\sum_{i=1}^{n}(y_i - \bar{y})^2 = \boldsymbol{y}'[\mathbf{I} - (\frac{1}{n})\boldsymbol{J}]\boldsymbol{y}$, $\hat{\boldsymbol{\beta}}_1 = (\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{X}_c'\boldsymbol{y}$, y $SCE = \sum_{i=1}^{n}(y_i - \bar{y})^2 - \hat{\boldsymbol{\beta}}_1' \boldsymbol{X}_c' \boldsymbol{y}$, podemos escribir (3.36) como

$$\begin{aligned}
\boldsymbol{y}'\left(\mathbf{I}-\frac{1}{n}\boldsymbol{J}\right)\boldsymbol{y} &= SCR + SCE \\
&= \boldsymbol{y}'\boldsymbol{X}_c(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{X}_c'\boldsymbol{y} + \boldsymbol{y}'\left(\mathbf{I}-\frac{1}{n}\boldsymbol{J}\right)\boldsymbol{y}\boldsymbol{y}'\boldsymbol{X}_c(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{X}_c'\boldsymbol{y} \\
&= \boldsymbol{y}'\boldsymbol{H}_c\boldsymbol{y} + \boldsymbol{y}'\left(\mathbf{I}-\frac{1}{n}\boldsymbol{J}-\boldsymbol{H}_c\right)\boldsymbol{y} \qquad (3.37)
\end{aligned}$$

donde $H_c = \boldsymbol{X}_c(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{X}_c'$.

En el síguete teorema establecemos algunas propiedades de las tres matrices de las formas cuadráticas en (3.37)

Teorema 3.2 *Las matrices $\boldsymbol{I}-(1/n)\boldsymbol{J}$, $\boldsymbol{H}_c = \boldsymbol{X}_c(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{X}_c'$ y $\boldsymbol{I}-(1/n)\boldsymbol{J}-\boldsymbol{H}_c$ tienen las siguientes propiedades*

1. $\boldsymbol{H}_c\left[\boldsymbol{I}-(1/n)\boldsymbol{J}\right] = \boldsymbol{H}_c$
2. *Hm_c es idempotente de rango k.*
3. *$\boldsymbol{I}-(1/n)\boldsymbol{J}-\boldsymbol{H}_c$ es idempotente de rango $n-k-1$.*
4. $\boldsymbol{H}_c\left[\boldsymbol{I}-(1/n)\boldsymbol{J}-\boldsymbol{H}_c\right] = 0$

Prueba: *Hacerla en clase*

Las distribuciones de SCR/σ^2 y SCE/σ^2 son dadas en el siguiente teorema

Teorema 3.3 *Si $\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$, entonces $SCR/\sigma^2 = \hat{\boldsymbol{\beta}}'\boldsymbol{X}_c'\boldsymbol{X}_c\hat{\boldsymbol{\beta}}_1/\sigma^2$ y $SCE/\sigma^2 = [\sum_{i=1}^{n}(y_i-\bar{y})^2 - \hat{\boldsymbol{\beta}}_1'\boldsymbol{X}_c'\boldsymbol{y}]/\sigma^2$ tienen las siguientes distribuciones*

1. *$SCR/\sigma^2 \sim \chi^2(k, \lambda_1)$, donde $\lambda_1 = \mu' A\mu/2\sigma^2 = \hat{\boldsymbol{\beta}}_1'\boldsymbol{X}_c'\boldsymbol{y}]/2\sigma^2$.*
2. $SCE/\sigma^2 \sim \chi^2(n-k-1)$

Prueba: *Hacerla en clase*

La independencia entre SCR y SCE es dada en el siguiente teorema

Teorema 3.4 *Si $\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$, entonces SCR y SCE son independientes.*
Prueba: *Hacerla en clase*

Ahora podemos establecer una prueba F para $H_0 : \boldsymbol{\beta}_1 = 0$ versus $H_1 : \boldsymbol{\beta}_1 \neq 0$.

Teorema 3.5 *Si $\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$, la distribución de*

$$F = \frac{SCR/(k\sigma^2)}{SCE/[(n-k-1)\sigma^2]} = \frac{SCR/k}{SCE/(n-k-1)} \qquad (3.38)$$

Es como sigue

1. *Si $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{0}$ es falsa, entonces $F \sim F(k, n-k-1, \lambda_1)$ donde,*

$$\lambda_1 = \boldsymbol{\beta}'\boldsymbol{X}_c'\boldsymbol{X}_c\boldsymbol{\beta}_1/2\sigma^2$$

2. *Si $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{0}$ es cierta, entonces $F \sim F(k, n-k-1)$*

Prueba: *Hacerla en clase*

La prueba para $H_0 : \boldsymbol{\beta}_1 = 0$ es llevada a cabo como sigue. Rechazamos H_0 si $F \geq F_{\alpha,k,n-k-1}$ donde $F_{\alpha,k,n-k-1}$ es el punto por encima del cual se encuentra el α por ciento de la distribución F central. Alternativamente un p valor puede ser usado para llevar a cabo la prueba. Un p valor es el área de la cola de la distribución F central perteneciente al valor calculado F, es decir, la probabilidad de exceder el valor calculado F, asumiendo que $H_0 : \boldsymbol{\beta}_1 = 0$ sea cierta. Un p valor menor que α es equivalente a $F \geq F_{\alpha,k,n-k-1}$.

La tabla de análisis de varianza (ANOVA) resume los resultados y los cálculos llevados a cabo en la prueba F. Los cuadrados medios son las sumas de cuadrados divididos por los grados de libertad de las distribuciones chi-cuadrado asociadas.

Table 3.1 Tabla ANOVA para la prueba F de $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{0}$

Fuente de variación	gl	Suma de cuadrados	Cuadrado Medio	Cuadrado medio esperado
Debido a $\boldsymbol{\beta}_1$	k	$SCR = \hat{\boldsymbol{\beta}}_1' \boldsymbol{X}_c' \boldsymbol{y} = \boldsymbol{\beta}' \boldsymbol{X}' \boldsymbol{y} - n\bar{y}^2$	SCR/k	$\sigma^2 + \frac{1}{k}\boldsymbol{\beta}_1' \boldsymbol{X}_c' \boldsymbol{X} - c\boldsymbol{\beta}_1$
Error	$n-k-1$	$SCE = \sum_{i=1}^n (y_i - \bar{y})^2 - \hat{\boldsymbol{\beta}}_1' \boldsymbol{X}_c' \boldsymbol{y}$	$SCE/(n-k-1)$	σ^2
Total	$n-1$	$SCE = \sum_{i=1}^n (y_i - \bar{y})^2$		

Las entradas en la columna para el cuadrado medio esperado en la tabla son simplemente $E(SCR/k)$ y $E[SCE/(n-k-1)]$.

Si $H_0 : \boldsymbol{\beta}_1 = 0$ es cierta, los valores esperados de los cuadrados medios son iguales a σ^2, y esperamos que el valor de F sea cercano a 1. Si $\boldsymbol{\beta}_1 \neq \boldsymbol{0}$, entonces $E(SCR/k) > \sigma^2$ ya que $\boldsymbol{X}_c' \boldsymbol{X}_c$ es definida positiva, y esperamos que F exceda a 1. Por lo tanto rechazamos H_0 para valores grandes de F.

La prueba de $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{0}$ ha sido desarrollada usando el modelo centrado. También podemos expresar SCR y SCE en terminos del modelo no centrado $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$,

$$SCR = \boldsymbol{\beta}' \boldsymbol{X}' \boldsymbol{y}, \qquad SCE = \boldsymbol{y}' \boldsymbol{y} - \boldsymbol{\beta}' \boldsymbol{X}' \boldsymbol{y} \tag{3.39}$$

EJEMPLO 3.3

Usando los datos del capítulo anterior, ilustramos la prueba de $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{0}$ donde, en este caso, $\boldsymbol{\beta}_1 = (\beta_1, \beta_2)'$. Anteriormente encontramos que $\boldsymbol{X}' \boldsymbol{y} = (90, 482, 872)'$ y $\hat{\boldsymbol{\beta}} = (5.3754, 3.0118, -1.2855)'$. Las cantidades $\boldsymbol{y}'\boldsymbol{y}$, $\hat{\boldsymbol{\beta}}' \boldsymbol{X}' \boldsymbol{y}$, y $n\bar{y}^2$ son dadas por

$$\boldsymbol{y}'\boldsymbol{y} = \sum_{i=1}^{1} 2y_i^2 = 2^2 + 3^2 + \cdots + 14^2 = 840$$

$$\hat{\boldsymbol{\beta}}' \boldsymbol{X}' \boldsymbol{y} = \begin{pmatrix} 5.3754 & 3.0118 & -1.2855 \end{pmatrix} \begin{pmatrix} 90 \\ 482 \\ 872 \end{pmatrix} = 814.5410$$

$$n\bar{y}^2 = 675$$

Así, obtenemos

$$SCR = \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} - n\bar{y}^2 = 139.5410$$

$$SCE = \boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} = 25.4590$$

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \boldsymbol{y}'\boldsymbol{y} - n\bar{y}^2 = 165$$

La prueba F esta dada en la tabla 3.2. Ya que $24.665 > F_{0.05,2,9} = 4.26$, rechazamos $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ y concluimos que al menos uno de β_1 o β_2 no es cero. El p valor es 0.000223.

Table 3.2 ANOVA para la prueba de la regresión general

Fuente	gl	SC	CM	F
Debido a $\boldsymbol{\beta}_1$	2	139.5410	69.7705	24.665
Error	9	25.4590	2.8288	
Total	11	165		

3.4 Prueba de un Subconjunto de los β

De manera más general, suponga que deseamos probar la hipótesis de que un subconjunto de los x no es útil para predecir y. Un ejemplo simple es $H_0 : \beta_j = 0$ para un solo β_j. Si H_0 es rechazada, debemos mantener $\beta_j x_j$ en el modelo. Como otra ilustración, considere el modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \varepsilon$$

para el cual deseamos probar la hipótesis $H_0 : \beta_3 = \beta_4 = \beta_5 = 0$. Si H_0 es rechazada, deberíamos seleccionar el modelo de segundo orden completo por encima del modelo reducido de primer orden.

Sin perdida de generalidad, asumamos que los β_j a ser probados son los últimos en el vector $\boldsymbol{\beta}$, con un correspondiente arreglo de las columnas de $\boldsymbol{X}$. Entonces $\boldsymbol{\beta}$ y $\boldsymbol{X}$ pueden ser particionadas convenientemente. De esta manera el modelo puede escribirse como

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \boldsymbol{\varepsilon} \\ &= \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \end{aligned} \tag{3.40}$$

donde $\boldsymbol{\beta}_2$ contiene los β a ser probados. El intercepto β_0 por lo general se incluye en $\boldsymbol{\beta}_1$.

La hipótesis de interés es $H_0 : \boldsymbol{\beta}_2 = 0$. Si designamos el número de parámetros en $\boldsymbol{\beta}_2$ por h, entonces $\boldsymbol{X}_2$ es $n \times h$, $\boldsymbol{\beta}_1$ es $(k - h + 1) \times 1$, y $\boldsymbol{X}_1$ es $n \times (k - h + 1)$. Por lo tanto, $\boldsymbol{\beta}_1 = (\beta_0, \beta_1, \cdots, \beta_{k-h})'$ y $\boldsymbol{\beta}_2 = (\beta_{k-h+1}, \cdots, \beta_k)'$. En términos de la ilustración al inicio de esta sección, deberíamos tener $\boldsymbol{\beta}_1 = (\beta_0, \beta_1, \beta_2)'$ y $\boldsymbol{\beta}_2 = (\beta_3, \beta_4, \beta_5)'$. Note que $\boldsymbol{\beta}_1$ en esta sección es diferente al $\boldsymbol{\beta}_1$ de la sección anterior, en el cual $\boldsymbol{\beta}$ fue particionado como $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \boldsymbol{\beta}_1 \end{pmatrix}$ y $\boldsymbol{\beta}_1$ constituido por todos los β excepto por β_0.

Para probar $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ versus $H_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$, usamos una técnica modelo-completo-reducido. El modelo completo esta dado por (3.40). Bajo $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$, el modelo reducido sería

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 * + \boldsymbol{\varepsilon} * \tag{3.41}$$

Usamos la notación $\boldsymbol{\beta}_1^*$ y $\boldsymbol{\varepsilon}^*$, ya que en el modelo reducido, $\boldsymbol{\beta}_1^*$ y $\boldsymbol{\varepsilon}^*$ serán típicamente diferentes a $\boldsymbol{\beta}_1$ y $\boldsymbol{\varepsilon}$ en el modelo completo (a menos que $\boldsymbol{X}_1$ y $\boldsymbol{X}_2$ sean ortogonales, ver teorema 7.9a y su corolario, Rencher). El estimador de $\boldsymbol{\beta}_1^*$ en el modelo reducido (3.41) es $\hat{\boldsymbol{\beta}}_1^* = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{y}$, el cual es, en general, no el mismo como los primeros $k-h+1$ elementos de $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ del modelo completo (3.40)(a menos que $\boldsymbol{X}_1$ y $\boldsymbol{X}_2$ sean ortogonales, ver teorema 7.10, Rencher)

Para comparar el ajuste del modelo completo (3.40) con el ajuste del modelo reducido (3.41), sumamos y restamos $\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}$ y $\hat{\boldsymbol{\beta}}_1^*\boldsymbol{X}_1'\boldsymbol{y}$ a la suma de cuadrados total corregida $\sum_{i=1}^n (y_i - \bar{y})^2 = \boldsymbol{y}'\boldsymbol{y} - n\bar{y}^2$ de manera de obtener la partición

$$\boldsymbol{y}'\boldsymbol{y} - n\bar{y}^2 = (\boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}) + (\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}_1^{*\prime}\boldsymbol{X}_1'\boldsymbol{y}) + (\hat{\boldsymbol{\beta}}_1^{*\prime}\boldsymbol{X}_1'\boldsymbol{y} - n\bar{y}^2) \tag{3.42}$$

o

$$SCT = SCE + SC(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) + SC(reducido) \tag{3.43}$$

donde $SC(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}_1^{*\prime}\boldsymbol{X}_1'\boldsymbol{y}$ es la suma de cuadrados de regresión "extra" debido a $\boldsymbol{\beta}_2$ luego de ajustar por $\boldsymbol{\beta}_1$. Note que $SC(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)$ puede también ser expresada como

$$\begin{aligned} SC(\boldsymbol{\beta}_1|\boldsymbol{\beta}_1) &= \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - n\bar{y}^2 - (\hat{\boldsymbol{\beta}}_1^{*\prime}\boldsymbol{X}_1'\boldsymbol{y} - n\bar{y}^2) \\ &= SCR(completo) - SCR(reducido) \end{aligned}$$

lo cual es la diferencia entre la suma de cuadrados de regresión para el modelo completo y la suma de cuadrados de regresión para el modelo reducido.

Si H_0 es verdadera, debemos esperar que $SC(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)$ sea pequeña de manera que SCT en 3.43 este compuesta en su mayor parte de $SCR(reducido)$ y SCE. Si $\boldsymbol{\beta}_2 \neq 0$ esperamos que $SC(\boldsymbol{\beta}_2)$ sea grande y represente mas a la SCT. Así estamos probando $H_0 : \boldsymbol{\beta}_2 = 0$ en el modelo en el cual no hay restricciones sobre $\boldsymbol{\beta}_1$. No estamos ignorando $\boldsymbol{\beta}_1$ (asumiendo $\boldsymbol{\beta}_1 = 0$) sino que estamos probando $H_0 : \boldsymbol{\beta}_2 = 0$ en la presencia de $\boldsymbol{\beta}_1$, esto es, mucho más allá de de lo que $\boldsymbol{\beta}_1$ contribuye a la SCT.

Para desarrollar un estadístico de prueba basado en $SC(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)$, escribimos primero 3.42 en términos de formas cuadráticas en $\boldsymbol{y}$. Usando $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ y $\hat{\boldsymbol{\beta}}_1^* = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{y}$, la ecuación (3.42) queda como

$$\begin{aligned}
\boldsymbol{y}'\left(\mathbf{I}-\frac{1}{n}\boldsymbol{J}\right)\boldsymbol{y} &= \boldsymbol{y}'\boldsymbol{y}-\boldsymbol{y}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}+\boldsymbol{y}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}\\
&\quad -\boldsymbol{y}'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{y}+\boldsymbol{y}'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{y}-\boldsymbol{y}'\frac{1}{n}\boldsymbol{J}\boldsymbol{y}\\
&= \boldsymbol{y}'[\mathbf{I}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y}+\boldsymbol{y}'[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'-\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1']\boldsymbol{y}\\
&\quad +[\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'-\frac{1}{n}\boldsymbol{J}]\boldsymbol{y}\\
&= \boldsymbol{y}'(\mathbf{I}-\boldsymbol{H})\boldsymbol{y}+\boldsymbol{y}'(\boldsymbol{H}-\boldsymbol{H}_1)\boldsymbol{y}+\boldsymbol{y}'\left(\boldsymbol{H}_1-\frac{1}{n}\boldsymbol{J}\right)\boldsymbol{y}
\end{aligned}$$

donde $\boldsymbol{H}=\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}$ y $\boldsymbol{H}_1=\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1$. La matriz $I-\boldsymbol{H}$ se demostró que es idempotente con rango $n-k-1$, donde $k+1$ es el rango de $\boldsymbol{X}$ ($k+1$ es también el número de elementos en $\boldsymbol{\beta}$). La matriz $\boldsymbol{H}-\boldsymbol{H}_1$ se demuestra a ser idempotente en el siguiente teorema.

Teorema 3.6 *La matriz $\boldsymbol{H}-\boldsymbol{H}_1=\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'-\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'$ es idempotente con rango h, donde h es el número de elementos en $\boldsymbol{\beta}_1$.*

Prueba: *Premultiplicando $\boldsymbol{X}$ por $\boldsymbol{H}$, obtenemos*

$$\boldsymbol{H}\boldsymbol{X}=\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}=\boldsymbol{X}$$

o

$$\boldsymbol{X}=[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{X} \tag{3.44}$$

Particionando $\boldsymbol{X}$ en el lado izquierdo de (3.44) y la última $\boldsymbol{X}$ del lado derecho, obtenemos

$$\begin{aligned}
\begin{pmatrix}X_1 & X_2\end{pmatrix} &= [\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\begin{pmatrix}X_1 & X_2\end{pmatrix}\\
&= \begin{bmatrix}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}_1 & \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}_2\end{bmatrix}
\end{aligned}$$

Por lo tanto,

$$\begin{aligned}
\boldsymbol{X}_1 &= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}_1\\
\boldsymbol{X}_1 &= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}_1
\end{aligned} \tag{3.45}$$

Simplificando $\boldsymbol{H}\boldsymbol{H}_1$ y $\boldsymbol{H}_1\boldsymbol{H}$ por (3.45) y su transpuesta, obtenemos

$$\boldsymbol{H}\boldsymbol{H}_1=\boldsymbol{H}_1 \qquad \boldsymbol{H}_1\boldsymbol{H}=\boldsymbol{H}_1 \tag{3.46}$$

Las matrices $\boldsymbol{H}$ y $\boldsymbol{H}_1$ son idempotentes. Por lo tanto,

$$\begin{aligned}
(\boldsymbol{H}-\boldsymbol{H}_1)^2 &= \boldsymbol{H}^2-\boldsymbol{H}\boldsymbol{H}_1-\boldsymbol{H}_1\boldsymbol{H}+\boldsymbol{H}_1^2\\
&= \boldsymbol{H}-\boldsymbol{H}_1-\boldsymbol{H}_1+\boldsymbol{H}_1\\
&= \boldsymbol{H}-\boldsymbol{H}_1,
\end{aligned}$$

y $\boldsymbol{H}-\boldsymbol{H}_1$ es idempotente. Para el rango de $\boldsymbol{H}-\boldsymbol{H}_1$, tenemos que

$$\begin{aligned}
r(\boldsymbol{H}-\boldsymbol{H}_1) &= tr(\boldsymbol{H}-\boldsymbol{H}_1)\\
&= tr(\boldsymbol{H})-tr(\boldsymbol{H}_1)\\
&= tr[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']-tr[\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1]\\
&= tr[\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}]-tr[\boldsymbol{X}_1'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}]\\
&= tr(\boldsymbol{I}_{k+1})-tr(\boldsymbol{I}_{k-h+1})=k+1-(k-h+1)=h
\end{aligned}$$

Hallamos ahora las distribuciones de $\boldsymbol{y}'(\mathbf{I}-\boldsymbol{H})\boldsymbol{y}$ y $\boldsymbol{y}'(\boldsymbol{H}-\boldsymbol{H}_1)\boldsymbol{y}$ en (**??**) y demostramos que ellos son independientes.

Teorema 3.7 *Si $\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$ y $\boldsymbol{H}$ y $\boldsymbol{H}_1$ están definidos como en (**??**) y (**??**), entonces.*

1. *$\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{y}/\sigma^2$ es $\chi^2(n-k-1)$*
2. *$\boldsymbol{y}'(\boldsymbol{H}-\boldsymbol{H}_1)\boldsymbol{y}/\sigma^2$ es $\chi^2(h\lambda_1)$, donde*

$$\lambda_1 = \boldsymbol{\beta}_2'[\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2]\boldsymbol{\beta}_2/2\sigma^2$$

3. *$\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{y}$ y $\boldsymbol{y}'(\boldsymbol{H}-\boldsymbol{H}_1)\boldsymbol{y}$ son independientes*

Prueba: *Agregando $\boldsymbol{y}'(1/n)\boldsymbol{y}/\sigma^2$ a ambos lados de (**??**), obtenemos la descomposición $\boldsymbol{y}'\boldsymbol{y} = \boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{y} + \boldsymbol{y}'(\boldsymbol{H}-\boldsymbol{H}_1)\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{H}_1\boldsymbol{y}$. Las matrices $\boldsymbol{I}-\boldsymbol{H}$, $\boldsymbol{H}-\boldsymbol{H}_1$ y $\boldsymbol{H}_1$ son idempotentes. Por lo tanto los item se obtienen de manera de directa. Solo faltaría la derivación de λ_1.*

Si $\lambda_1 = 0$ en el teorema 3.7(ii), entonces $\boldsymbol{y}'(\boldsymbol{H}-\boldsymbol{H}_1)\boldsymbol{y}$ tiene la distribución chi-cuadrado central $\chi-^2(h)$. Ya que $\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2$ es definida positiva, $\lambda_1 = 0$ si y sólo si $\boldsymbol{\beta}_2 = 0$.

Una prueba F para $H_0 : \boldsymbol{\beta}_2 = 0$ versus $H_1 : \boldsymbol{\beta}_2 \neq 0$ esta dada en el siguiente teorema.

Teorema 3.8 *Sea $\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$ y definamos un estadístico F como sigue:*

$$F = \frac{\boldsymbol{y}'(\boldsymbol{H}-\boldsymbol{H}_1)\boldsymbol{y}/h}{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{H})/(n-k-1)} = \frac{SC(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)/h}{SCE/(n-k-1)} \tag{3.47}$$

$$= \frac{(\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}_1^{*\prime}\boldsymbol{X}_1'\boldsymbol{y})/h}{(\boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y})/(n-k-1)} \tag{3.48}$$

donde $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ es del modelo completo $\boldsymbol{y} = \boldsymbol{X\beta}+\boldsymbol{\varepsilon}$ y $\hat{\boldsymbol{\beta}}_1^* = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{y}$ es del modelo reducido $\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1^* + \boldsymbol{\varepsilon}$. La distribución de F en (3.48) es la siguiente:

(i) Si $H_0 : \boldsymbol{\beta}_2 = 0$ es falsa, entonces

$$F \sim F(h, n-k-1, \lambda_1)$$

donde $\lambda_1 = \boldsymbol{\beta}_2'[\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2]\boldsymbol{\beta}_2/2\sigma^2$

(ii) Si $H_0 : \boldsymbol{\beta}_2 = 0$ es cierta, entonces $\lambda_1 = 0$ y

$$F \sim F(h, n-k-1)$$

Prueba: *Hacerlo en clase.*

La prueba de la hipótesis $H_0 : \boldsymbol{\beta}_2 = 0$ es llevada a cabo de la siguiente manera: Rechazar H_0 si $F > F_{\alpha,h,n-k-1}$ donde $F_{\alpha,h,n-k-1}$ es el punto de la distribución F central cuya área derecha es α. Alternativamente, rechazamos H_0 si $p < \alpha$, donde p es el p-valor. Ya que $\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2$ es definida positiva, $\lambda_1 > 0$ si $H_0 : \boldsymbol{\beta}_2 = 0$ es falso. Esto justifica el rechazo de H_0 para valores grandes de F.

Los resultados para esta prueba se resumen en la tabla anova (TTTTT), donde $\boldsymbol{\beta}_1$ es $(k-h+1)\times 1$, $\boldsymbol{\beta}_2$ es $h\times 1$, $\boldsymbol{X}_1$ es $n\times(k-h+1)$, $\boldsymbol{\beta}_2$ es $n\times h$.

Las entradas en la columna para los cuadrados medios esperados son $E[SC(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)/h]$ y $E[SCE/(n-k-1)]$. Note que si H_0 es cierta, ambos cuadrados medios esperados son iguales a σ^2, y si H_0 es falsa, $E[SC(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)/h] > E[SCE/(n-k-1)]$ ya que $\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2$ es definida positiva. Esta desigualdad provee otra justificación para rechazar H_0 para valores grandes de F.

Aquí viene la tabla

EJEMPLO 3.4

Considere la variable dependiente y_2 en la data de una reacción química en la tabla 7.4 (Rencher). Para chequear el uso de terminos de segundo orden en la predicción de y_2, usamos un modelo completo, $y_2 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1^2 + \beta_5 x_2^2 + \beta_6 x_3^2 + \beta_7 x_1 x_2 + \beta_8 x_1 x_3 + \beta_9 x_2 x_3 + \varepsilon$ y la prueba de $H_0 : \beta_4 = \beta_5 = ... = \beta_9 = 0$. Para el modelo completo, obtenemos $\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - n\bar{y}^2 = 339.7888$, y para el modelo reducido $y_2 = \beta_0^* + \beta_1^* x_1 + \beta_2^* x_2 + \beta_3^* x_3 + \varepsilon^*$, tenemos $\hat{\boldsymbol{\beta}}_1^{*\prime}\boldsymbol{X}_1'\boldsymbol{y} - n\bar{y}^2 = 151.0022$. La diferencia es $\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}_1^{*\prime}\boldsymbol{X}_1'\boldsymbol{y} = 188.7866$. La suma de cuadrados del error es $SCE = 60.6755$, y el estadístico F sería

$$F = frac188.7866/660.6755/9 = 4.6671$$

el cual tiene un $p-valor$ de 0.0198. Por lo tanto los términos de segundo orden son usados en la predicción de y_2. De hecho, la F general del modelo reducido en (**??**) es 3.27 com $p = 0.0623$, de manera que x_1, x_2 y x_3 son inadecuados para predecir y_2. La F general para el modelo completo es 5.600 con $p = 0.0086$.

En el siguiente teorema, expresamos $SC(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)$ como una forma cuadratica en $\hat{\boldsymbol{\beta}}_2$ que corresponde a λ_1 en el teorema 3.7(ii).

Teorema 3.9 *Si el modelo es particionado como en (3.40), entonces*

$$SC(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}_1^{*\prime}\boldsymbol{X}_1'\boldsymbol{y}$$

se puede escribir como

$$SC(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = \hat{\boldsymbol{\beta}}_2'[\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2]\hat{\boldsymbol{\beta}}_2 \tag{3.49}$$

donde $\hat{\boldsymbol{\beta}}_2$ es de la partición de $\hat{\boldsymbol{\beta}}$ en el modelo completo

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \tag{3.50}$$