# MODELOS LINEALES

# MODELOS LINEALES
## Introducción

**Douglas Rivas**
Universidad de Los Andes

**Clases**

**PART I**

# INTRODUCCIÓN

## Capítulo 1

# 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 $\sigma^2$. 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 $p < n$.

## 1.1 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 $\beta_i$. A continuación algunos ejemplos

**EJEMPLO 1.1**

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 1.2**

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 $\beta_1$ sea igual a $\beta_2$, ignora los valores de los $\beta_i$ restantes.
2. $\beta_1 = \beta_2$ y $\beta_3 = \beta_4$, una prueba de que $\beta_1$ sea igual a $\beta_2$ y $\beta_3$ sea igual a $\beta_4$ ignorando el valor de $\beta_0$
3. $\beta_1 = \beta_2 = 6$; esta es una prueba diferente a la 1, aquí la prueba es que $\beta_1$ y $\beta_2$ son ambos iguales a 6; en la parte 1 la prueba es que $\beta_1$ y $\beta_2$ 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 $\beta_j$ son simultaneamente iguales a cero.
5. $\beta_1 = \beta_2 = \beta_3 = \beta_4$; una prueba de que todos los $\beta_j$ (excepto $\beta_0$) 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 $\beta_j$.

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

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

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 1.2. 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 1.1** *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 $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2\mathbf{I})$. 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 **??**.

### 1.1.1 Cálculo del Estadístico de Prueba

El cálculo del estadístico de prueba se puede realizar a través de dos metodologías, bien sea usando la prueba de la razón de verosimilitud generalizada o lo que algunos autores han llamado el modelo completo-reducido.

***1.1.1.1 Razón de Verosimilitud Generalizada*** En esta metodología se compara el valor máximo de la función de verosimilitud, $L(\boldsymbol{\beta}, \sigma^2)$, restringido por la hipótesis, $H_0 : \boldsymbol{H\beta} = \boldsymbol{h}$, 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:

$$v(y) = \frac{\max\limits_{(\boldsymbol{\beta},\sigma^2)\in\omega} [L(\boldsymbol{\beta}, \sigma^2 : \boldsymbol{y})]}{\max\limits_{(\boldsymbol{\beta},\sigma^2)\in\Omega} [L(\boldsymbol{\beta}, \sigma^2 : \boldsymbol{y})]} = \frac{L(\hat{\omega})}{L(\hat{\Omega})} \tag{1.1}$$

donde los espacios muestrales $\Omega$ y $\omega$ están dados por:

$$\begin{aligned} \Omega &= \{(\boldsymbol{\beta}, \sigma^2) : \boldsymbol{\beta} \in E_p; \sigma^2 > 0\} \\ \omega &= \{(\boldsymbol{\beta}, \sigma^2) : \boldsymbol{\beta} \in E_p; \boldsymbol{H\beta} = \boldsymbol{h}; \sigma^2 > 0\} \end{aligned} \tag{1.2}$$

Es decir, $\Omega$ es el espacio de los parámetros bajo la hipótesis alternativa (sin restricciones), y $\omega$ 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, $\boldsymbol{y} = (y_1, ..., y_n)'$ 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 $L(\boldsymbol{\beta}, \sigma^2 : \boldsymbol{y})$, está dada por

$$L(\boldsymbol{\beta}, \sigma^2 : \boldsymbol{y}) = (2\pi)^{-n/2}(\sigma^2)^{-n/2}e^{-(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta})/2\sigma^2} \tag{1.3}$$

Por lo tanto para hallar el estadístico de prueba debemos hallar $L(\hat{\Omega})$ y $L(\hat{\omega})$. En el caso de $L(\hat{\Omega})$ la cuestión es fácil, ya que como no existen restricciones la función de verosimilitud alcanza el máximo cuando esta es evaluada en los estimadores de máxima verosimilitud de $\boldsymbol{\beta}$ y $\sigma^2$, los cuales fueron calculados en el capítulo anterior y están dados por $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ y $\hat{\sigma}^2 = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$, que en este caso se van a denotar como $\hat{\boldsymbol{\beta}}_{\boldsymbol{\Omega}}$ y $\hat{\sigma}^2_{\Omega}$,respectivamente. Por lo tanto sustituyendo estas dos ecuaciones en la ecuación 1.3 se obtiene que

$$L(\hat{\Omega}) = (2\pi)^{-n/2}(\hat{\sigma}^2_{\Omega})^{-n/2}e^{-n/2} \tag{1.4}$$

Para hallar $L(\hat{\omega})$ se debe encontrar el máximo de la función de verosimilitud $L(\boldsymbol{\beta}, \sigma^2 : \boldsymbol{y}; X)$ con respecto a $\boldsymbol{\beta}$ y $\sigma^2$ cuando estos parámetros pueden variar en $\omega$, es decir cuando tienen que cumplir la restricción impuesta por la hipótesis. Esto puede hacerse por dos métodos, bien sea por el método de sustitución o por el método de los multiplicadores de Lagrange. A continuación se desarrollan ambos casos.

1. **Método de Sustitución.** Este método consiste en sustituir la restricción, $\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}$, en el modelo $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, 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 así

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

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 $\boldsymbol{H}\boldsymbol{G}' = 0$, 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

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

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

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

lo cual es equivalente a

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

Haciendo $\boldsymbol{Z} = \boldsymbol{Y} - \boldsymbol{XH}^{-}\boldsymbol{h}$, $\boldsymbol{B} = \boldsymbol{XG}^{-}$ y $\boldsymbol{\gamma} = \boldsymbol{G\beta}$, el modelo se escribe como $\boldsymbol{Z} = \boldsymbol{B\gamma} + \boldsymbol{\varepsilon}$, el cual es el modelo $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ 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 $\sigma^2$ están dados también en las ecuaciones **??** y **??** con $\boldsymbol{B}$ reemplazando a $\boldsymbol{X}$ y $\boldsymbol{Z}$ reemplazando a $\boldsymbol{Y}$. Ellos son denotados por $\hat{\boldsymbol{\gamma}}_\omega$ y $\hat{\sigma}^2_\omega$, y están dados por

$$\hat{\boldsymbol{\gamma}}_\omega = \boldsymbol{B}^{-}\boldsymbol{Z} \tag{1.5}$$

$$\hat{\sigma}^2_\omega = \frac{1}{n}\boldsymbol{Z}'(\mathbf{I} - \boldsymbol{BB}^{-})\boldsymbol{Z} \tag{1.6}$$

Entonces,

$$L(\hat{\omega}) = (2\pi)^{-n/2}(\hat{\sigma}^2_\omega)^{-n/2}e^{-n/2} \tag{1.7}$$

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

$$V(y) = V = \left(\frac{\hat{\sigma}^2_\Omega}{\hat{\sigma}^2_\omega}\right)^{n/2} \tag{1.8}$$

**Nota 1.2** *Si $q = p$, entonces $\boldsymbol{B} = 0$ y el modelo reducido es $\boldsymbol{Z} = \varepsilon$, donde $\boldsymbol{Z} = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{H}^{-1}\boldsymbol{h}$ y el espacio del parámetro $\omega$ es $\omega = \{(\boldsymbol{\beta}, \sigma^2) : \boldsymbol{\beta} = \boldsymbol{H}^1\boldsymbol{h} = b; \sigma^2 > 0\}$. Además $L(\hat{\omega}) = (2\pi)^{-n/2}(\hat{\sigma}_\omega^2)^{-n/2}e^{-n/2}$ donde $\hat{\sigma}_\omega^2 = \frac{1}{n}\boldsymbol{Z}'\boldsymbol{Z} = \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{X}b)'(\boldsymbol{Y} - \boldsymbol{X}b)$.*

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:

$$W = \left(\frac{n-p}{q}\right)\left(\frac{\hat{\sigma}_\omega^2 - \hat{\sigma}_\Omega^2}{\hat{\sigma}_\Omega^2}\right) = \left(\frac{n-p}{q}\right)\left[\frac{\boldsymbol{Z}'(\mathbf{I} - \boldsymbol{B}\boldsymbol{B}^-)\boldsymbol{Z} - \boldsymbol{Y}'(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Y}}{\boldsymbol{Y}'(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Y}}\right] \tag{1.9}$$

**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

$$\begin{aligned}
\frac{\partial L}{\partial \boldsymbol{\beta}} &= L(\boldsymbol{\beta}, \sigma^2 : \boldsymbol{y}; \boldsymbol{X})\left[-(2\sigma^2)^{-1}(-\boldsymbol{y}'\boldsymbol{X} - \boldsymbol{X}'\boldsymbol{y} + \boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X})\right] - \boldsymbol{H}'\boldsymbol{\lambda} \\
&= L(\boldsymbol{\beta}, \sigma^2 : \boldsymbol{y}; \boldsymbol{X})\left[-(2\sigma^2)^{-1}(-2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta})\right] - \boldsymbol{H}'\boldsymbol{\lambda} \\
&= L(\boldsymbol{\beta}, \sigma^2 : \boldsymbol{y}; \boldsymbol{X})\left[-(2\sigma^2)^{-1}(2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} - 2\boldsymbol{X}'\boldsymbol{y})\right] - \boldsymbol{H}'\boldsymbol{\lambda} && (1.10) \\
\frac{\partial L}{\partial \sigma^2} &= -\frac{n}{2}(2\pi\sigma^2)^{-n/2-1}\exp\left[\frac{-1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right] \\
&\quad + L(\boldsymbol{\beta}, \sigma^2 : \boldsymbol{y}; \boldsymbol{X})(2\sigma^4)^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\
&= -n\pi(2\pi\sigma^2)^{-1}L(\boldsymbol{\beta}, \sigma^2) + (2\sigma^4)^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})L(\boldsymbol{\beta}, \sigma^2) && (1.11) \\
\frac{\partial L}{\partial \boldsymbol{\lambda}} &= -(\boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{h}) && (1.12)
\end{aligned}$$

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

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} & \boldsymbol{H}' \\ \boldsymbol{H} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\beta}}_\omega \\ \boldsymbol{\lambda}^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{y} \\ \boldsymbol{h} \end{bmatrix} \tag{1.13}$$

$$\tilde{\sigma}^2_\omega = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}_\omega)'(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}_\omega) \tag{1.14}$$

Resolviendo estas ecuaciones se tiene que

$$\boldsymbol{\lambda}^* = [\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\tilde{\boldsymbol{\beta}}_\Omega - \boldsymbol{h}) \tag{1.15}$$

$$\tilde{\boldsymbol{\beta}}_\omega = \hat{\boldsymbol{\beta}}_{\boldsymbol{\Omega}} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\tilde{\boldsymbol{\beta}}_\Omega - \boldsymbol{h}) \tag{1.16}$$

$$\tilde{\sigma}^2_\omega = \tilde{\sigma}^2_\Omega + \left(\frac{1}{n}\right)(\boldsymbol{H}\tilde{\boldsymbol{\beta}}_\Omega - \boldsymbol{h})[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\tilde{\boldsymbol{\beta}}_\Omega - \boldsymbol{h}) \tag{1.17}$$

donde $\hat{\boldsymbol{\beta}}_{\boldsymbol{\Omega}} = \boldsymbol{X}^{-}\boldsymbol{Y}$ es el Estimador máximo verosímil de $\boldsymbol{\beta}$ en el modelo completo, $\tilde{\sigma}^2_\Omega$ es el estimador máximo verosímil de $\sigma^2$ en el modelo completo. Si se sustituye $\tilde{\boldsymbol{\beta}}_\omega$ y $\tilde{\sigma}^2_\omega$ en la ecuación 1.9 y realizando ciertas operaciones algebraicas se obtiene la siguiente expresión para $W$.

$$W = \left(\frac{n-p}{q}\right)\left\{\frac{(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})}{\boldsymbol{Y}'(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^{-})\boldsymbol{Y}}\right\} \tag{1.18}$$

***1.1.1.2 Modelo Completo-Reducido*** La idea se basa en la comparación de 2 modelos, el **modelo completo** el cual es el modelo $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, y el **modelo reducido** el cual es el modelo $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ sujeto a la hipótesis $\boldsymbol{H\beta} = \boldsymbol{h}$. Por ejemplo si el modelo completo es

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i$$

y la hipótesis es $H_0 : \beta_1 = \beta_2$, entonces el modelo reducido es

$$y_i = \beta_0 + \beta_1(x_{i1} + x_{i2}) + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i$$

Ahora bien, la comparación de los 2 modelos se realiza comparando las sumas de cuadrados de regresión (SCR) entre ambos modelos. Hemos visto que la SCR aumenta a medida que ingresamos más variables en el modelo, por lo tanto, bajo $H_0$ cierta, se esperaría que la SCR del modelo completo sea mayor que la SCR del modelo reducido. De manera equivalente, en lugar de comparar las SCR de los dos modelos se pueden comparar las sumas de cuadrados del error de dichos modelos, en este caso si la $H_0 : \boldsymbol{H\beta} = \boldsymbol{h}$ es cierta se esperaría que la SCE del modelo completo sea mayor que la SCE del modelo reducido.

A continuación se deriva una relación existente entre las SCE de ambos modelos, para ello usaremos la siguiente notación:

- $\hat{\boldsymbol{\beta}}$ es el estimador de $\boldsymbol{\beta}$ en el modelo completo, el cual está dado por $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$
- $\hat{\boldsymbol{\beta}}_r$ es el estimador de $\boldsymbol{\beta}$ en el modelo reducido, es decir es el estimador de $\boldsymbol{\beta}$ bajo la hipótesis nula, que como se vio en la sección anterior, en el método de Lagrange, este esta dado por $\tilde{\boldsymbol{\beta}}_\omega$

Por lo tanto la SCE del modelo reducido está dada por

$$\begin{aligned}
SCE(reducido) &= (\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_r)'(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_r)\\
&= (\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_r)'(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_r)\\
&= (\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})+(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_r)\\
&+ (\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_r)'(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})+(\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_r)'(\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_r)\\
&= SCE(completo)+(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})'\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_r)+(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_r)'\boldsymbol{X}'(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})\\
&+ (\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_r)\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_r)\\
&= SCE(completo)+(\boldsymbol{X}'\boldsymbol{y}-\boldsymbol{X}\boldsymbol{X}\hat{\boldsymbol{\beta}})'(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_r)+(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_r)'(\boldsymbol{X}'\boldsymbol{y}-\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}})\\
&+ (\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_r)\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_r)\\
&= SCE(completo)\\
&+ \{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h})\}'\boldsymbol{X}'\boldsymbol{X}\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h})\}\\
&= SCE(completo)\\
&+ (\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h})'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h})\\
&= SCE(completo)+(\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h})'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h}) \qquad (1.19)
\end{aligned}$$

Entonces

$$SCE(reducido)-SCE(completo) = (\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h})'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h}) \quad (1.20)$$

A dicha diferencia se le conoce como la suma de cuadrados debido a la hipótesis y se representa por SCH. Entonces la cantidad SCH nos dice cuanto se incrementa el error al considerar el modelo reducido en vez del modelo completo, es decir de no rechazar $H_0$. Intuitivamente si esta cantidad es grande implica el rechazo de $H_0 : \boldsymbol{H\beta} = \boldsymbol{h}$. Ahora la pregunta es ¿cuándo consideramos que SCH es grande?. Para ello comparamos el valor de SCH con el valor de SCE del modelo completo, lo cual se resume en el siguiente teorema.

**Teorema 1.1** *Considere el modelo lineal $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon}\sin N(0, \Sigma^2\boldsymbol{I})$ y sea $\boldsymbol{H}$ una matriz $q \times p$ de rango $q \leq p$, entonces el estadístico de prueba de la hipótesis $\boldsymbol{H\beta} = \boldsymbol{h}$ esta dado por*

$$W = \frac{SCH/q}{SCE/n-p} \quad (1.21)$$

*donde*

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

$$SCE = \boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{y}$$

Se puede notar que el estadístico de prueba obtenido con el método de Lagrange es el mismo que el obtenido con el método del modelo completo-reducido, pero estos a su vez difieren con el obtenido por el método de sustitución. A continuación veremos que aunque ellos difieren a la vista, realmente son equivalentes.

***1.1.1.3 Equivalencia de los 2 estadísticos*** A continuación se demuestra que la fórmula para el estadístico de prueba obtenido con el método de selección, ecuación 1.9, es igual al obtenido con el método de Lagrange y el método del modelo completo-reducido, ecuación 1.18. Para ello sólo basta con probar que

$$\boldsymbol{Z}'(\mathbf{I}-\boldsymbol{B}\boldsymbol{B}^-)\boldsymbol{Z}-\boldsymbol{Y}'(\mathbf{I}-\boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{Y} \quad (1.22)$$

en la ecuación 1.9 sea igual a

$$(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h}) \tag{1.23}$$

en la ecuación 1.18.
Para ello es necesario realizar algunas transformaciones en las dos ecuaciones. Comencemos con la ecuación 1.9, 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

$$\boldsymbol{Z}'(\mathbf{I} - \boldsymbol{B}\boldsymbol{B}^{-})\boldsymbol{Z} - \boldsymbol{Y}'(\mathbf{I} - \boldsymbol{X}\boldsymbol{X}^{-})\boldsymbol{Y} = \boldsymbol{Z}'[\boldsymbol{X}\boldsymbol{X}^{-} - (\boldsymbol{X}\boldsymbol{G}^{-})(\boldsymbol{X}\boldsymbol{G}^{-})^{-}]\boldsymbol{Z} \tag{1.24}$$

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 1.24.
Con respecto a la ecuación 1.23, al sustituir $\hat{\boldsymbol{\beta}} = \boldsymbol{X}^{-}\boldsymbol{Y}$ y realizando algunas simplificaciones se obtiene que

$$(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h}) = \boldsymbol{Z}'(\boldsymbol{H}\boldsymbol{X}^{-})^{-}(\boldsymbol{H}\boldsymbol{X}^{-})\boldsymbol{Z} \tag{1.25}$$

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 simétrica 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 1.9 y 1.18 son los mismos, se debe mostrar que las ecuaciones 1.25 y 1.24 son las mismas, y esto significa que se debe probar que

$$\boldsymbol{X}\boldsymbol{X}^- - (\boldsymbol{X}\boldsymbol{G}^-)(\boldsymbol{X}\boldsymbol{G}^-)^- = (\boldsymbol{H}\boldsymbol{X}^-)^-(\boldsymbol{H}\boldsymbol{X}^-) \tag{1.26}$$

lo cual es equivalente a probar que

$$\boldsymbol{X}\boldsymbol{X}^- = (\boldsymbol{X}\boldsymbol{G}^-)(\boldsymbol{X}\boldsymbol{G}^-)^- + (\boldsymbol{H}\boldsymbol{X}^-)^-(\boldsymbol{H}\boldsymbol{X}^-) \tag{1.27}$$

Haciendo

$$\boldsymbol{N} = (\boldsymbol{X}\boldsymbol{G}^-)(\boldsymbol{X}\boldsymbol{G}^-)^- + (\boldsymbol{H}\boldsymbol{X}^-)^-(\boldsymbol{H}\boldsymbol{X}^-) \tag{1.28}$$

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

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

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

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

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

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

- 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

$$\boldsymbol{A} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'(\boldsymbol{HX}^-)'^- + \boldsymbol{G}^-(\boldsymbol{XG}^-)^- \tag{1.29}$$

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 1.9 es el mismo $W$ en la ecuación 1.18.

### 1.1.2 Distribución del Estadístico de Prueba

En esta sección se determina la distribución del estadístico de prueba. En la sección anterior se obtuvieron dos estadísticos de prueba pero también se demostró que los dos estadísticos son equivalentes lo cual implica que ellos tienen la misma distribución, por lo tanto es suficiente determinar como se distribuye uno de ellos. En el siguiente teorema se enuncia la distribución del estadístico de prueba obtenido por el método de Lagrange (Modelo Completo-Reducido).

**Teorema 1.2** *Sea $\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$ y definamos un estadístico $F$ como sigue:*

$$W = \frac{(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})/q}{\boldsymbol{y}'[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y}/(n-p)} \tag{1.30}$$

*La distribución de $F$ en (1.31) es la siguiente:*

*(i) Si $H_0 : \boldsymbol{H\beta} = \boldsymbol{h}$ es falsa, entonces*

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

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

*(ii) Si $H_0 : \boldsymbol{H\beta} = \boldsymbol{h}$ es cierta, entonces $\lambda = 0$ y*

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

***Prueba:*** *Para probar que el estadístico W tiene una distribución F con los parámetros mencionados, debemos primero hacerle un cambio al estadístico W dado de la siguiente manera:*

$$W = \frac{SCH/q}{SCE/(n-p)} = \frac{(SCH/\sigma^2)/q}{(SCE/\sigma^2)/(n-p)} = \frac{U/q}{V/(n-p)}$$

*y ahora es necesario que se cumplan tres condiciones*

1. $U \sim \chi^2(q, \lambda)$
2. $V \sim \chi^2(n-p)$
3. *$U$ y $V$ son independientes*

*Veamos el cumplimiento de estas condiciones*

1. $U \sim \chi^2(q, \lambda)$ *Para esta parte recordemos que si $\boldsymbol{y} \sim N(\mu, \Sigma)$ entonces $\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} \sim \chi^2(q, \lambda)$, donde $\lambda = \mu'\boldsymbol{A}\mu/2\sigma^2$, si y solo si*

   *a)* *$\boldsymbol{A}\Sigma$ es idempotente*

   *b)* $r(\boldsymbol{A}) = q$

   *Entonces el estadístico*

$$\begin{aligned} U &= \frac{(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})}{\sigma^2} \\ &= (\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})'\boldsymbol{A}(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h}) \end{aligned}$$

*se distribuye* $\chi^2(q,\lambda)$ *si sólo si* $\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h}$ *se distribuye normal y se cumplen las condiciones a) y b).*

*Veamos la normalidad. Se sabe que* $\hat{\boldsymbol{\beta}}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}\sim N(\boldsymbol{\beta},\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$, *y como* $\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h}$ *es una combinación lineal de* $\hat{\boldsymbol{\beta}}$ *entonces* $\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h}$ *también se distribuye normal, cuyos parámetros son:*

- $(E\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h})=\boldsymbol{H}E(\hat{\boldsymbol{\beta}})-\boldsymbol{h}=\boldsymbol{H}\boldsymbol{\beta}-\boldsymbol{h}$
- $Cov(\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h})=\boldsymbol{H}Cov(\hat{\boldsymbol{\beta}})\boldsymbol{H}'=\boldsymbol{H}\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'=\sigma^2\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'$

*Por lo tanto,* $\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h}\sim N(\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h},\sigma^2\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}')$. *Ahora probamos las condiciones a) y b)*

*a)* $\boldsymbol{A}\Sigma=\frac{[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}}{\sigma^2}\sigma^2\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'=\boldsymbol{I}$ *la cual es idempotente.*

*b)* $r(A)$ *no lo he hecho*

*Por lo tanto* $U\sim\chi^2(q,\lambda)$ *donde*

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

2. $V\sim\chi^2(n-p)$. *El estadístico* $V$ *está dado por*

$$V=\frac{SCE}{\sigma^2}=\frac{\boldsymbol{y}'[\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y}}{\sigma^2}=\boldsymbol{y}'\boldsymbol{B}\boldsymbol{y}$$

*Siguiendo la misma idea se tiene que como* $\boldsymbol{y}\sim N(\boldsymbol{X}\boldsymbol{\beta},\sigma^2\boldsymbol{I})$ *solo basta con probar las condiciones a) y b) de 1.*

*a)* $\boldsymbol{B}\Sigma=\frac{[\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']}{\sigma^2}\sigma^2\boldsymbol{I}=\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ *la cual es idempotente.*

*b)* $r(B)=r(\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')=n-p$. *Demostrado anteriormente.*

*Por lo tanto* $V\sim\chi^2(n-p)$ *ya que*

$$\begin{aligned}\lambda &= (\boldsymbol{X}\boldsymbol{\beta})'\frac{[\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']}{\sigma^2}(\boldsymbol{X}\boldsymbol{\beta})/2\\ &= \frac{\boldsymbol{\beta}'\boldsymbol{X}'X\boldsymbol{\beta}-\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'X\boldsymbol{\beta}}{2\sigma^2}\\ &= \frac{\boldsymbol{\beta}'\boldsymbol{X}'X\boldsymbol{\beta}-\boldsymbol{\beta}'\boldsymbol{X}'X\boldsymbol{\beta}}{2\sigma^2}=0\end{aligned}$$

3. $U$ *y* $V$ *son independientes. Existen dos maneras de probar la independencia de* $u$ *y* $V$.

   - *En el tema anterior se probó que* $\hat{\boldsymbol{\beta}}$ *y* $SCE$ *son independientes. Además se sabe que funciones de variables aleatorias independientes también son independientes y como* $SCH$ *es una función de* $\hat{\boldsymbol{\beta}}$, *entonces* $SCH$ *y* $SCE$ *son independientes.*
   - *La otra manera de probar la independencia es expresando* $SCH$ *y* $SCE$ *como formas cuadráticas con el mismo vector aleatorio normal.*

*Escribamos a $SCH$ de esa manera*

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

*Ahora*

$$\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{h} = \boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'[\boldsymbol{y} - \boldsymbol{X}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{h}]$$

*entonces*

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

*Escribamos ahora a $SCE$ de igual manera*

$$\begin{aligned} SCE &= \boldsymbol{y}'[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{B}\boldsymbol{y} \\ &= \boldsymbol{y}'\boldsymbol{B}\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{B}\boldsymbol{X}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{h} - \boldsymbol{h}'(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{X}'\boldsymbol{B}\boldsymbol{y} \\ &+ [\boldsymbol{X}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{h}]'\boldsymbol{B}[\boldsymbol{X}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{h}] \\ &+ \boldsymbol{y}'\boldsymbol{B}\boldsymbol{X}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{h} + \boldsymbol{h}'(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{X}'\boldsymbol{B}\boldsymbol{y} \\ &- [\boldsymbol{X}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{h}]'\boldsymbol{B}[\boldsymbol{X}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{h}] \\ &= [\boldsymbol{y} - \boldsymbol{X}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{h}]'\boldsymbol{B}[\boldsymbol{y} - \boldsymbol{X}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{h}] \end{aligned}$$

*Pues*

$$\begin{aligned} \boldsymbol{y}'\boldsymbol{B}\boldsymbol{X}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{h} &= \boldsymbol{y}'[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{X}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{h} \\ &= \boldsymbol{y}'[\boldsymbol{X}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1}]\boldsymbol{h} \\ &= \boldsymbol{y}'[\boldsymbol{X}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1} - \boldsymbol{X}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1}]\boldsymbol{h} = 0 \end{aligned}$$

$$\begin{aligned} \boldsymbol{h}'(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{X}'\boldsymbol{B}\boldsymbol{y} &= \boldsymbol{h}(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{X}'[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y} \\ &= \boldsymbol{h}'[(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{X}' - (\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y} \\ &= \boldsymbol{h}'[(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{X}' - (\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{X}']\boldsymbol{y} = 0 \end{aligned}$$

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

*Por lo tanto $SCE$ y $SCH$ son formas cuadráticas del vector $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{h}$, el cual como es una función lineal de $\boldsymbol{y}$ también se distribuye normal cuyos parámetros son:*

- $E[\boldsymbol{y} - \boldsymbol{X}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{h}] = \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{h}$

- $Cov[\boldsymbol{y} - \boldsymbol{X}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{h}] = \sigma^2\boldsymbol{I}$

*Así* $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{h} \sim N(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{h}, \sigma 2\boldsymbol{I})$*. Entonces solo basta con demostrar que* $\boldsymbol{AB} = 0$

$$\begin{aligned}
\boldsymbol{AB} &= \{\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\}\{\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\} \\
&= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}( \\
&= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' = 0
\end{aligned}$$

*Con ello concluimos que $SCE$ y $SCH$ son independientes.*

*Por lo tanto el estadístico*

$$W = \frac{(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})/q}{\boldsymbol{y}'[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y}/(n-p)} \tag{1.31}$$

*tiene la siguiente distribución*

*(i) Si* $H_0 : \boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}$ *es falsa, entonces*

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

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

*(ii) Si* $H_0 : \boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}$ *es cierta, entonces* $\lambda = 0$ *y*

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

### 1.1.3 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

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

entonces la potencia de esta prueba está dada por

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

Por lo tanto, para determinar la potencia de la prueba es necesario conocer $P[W > F_{\alpha,q,n-p} | \boldsymbol{H}\boldsymbol{\beta} \neq \boldsymbol{h}]$. 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=$\alpha$, df1=, df2=, lower.tail = FALSE, log.p = FALSE), df1=, df2=, ncp=$\lambda$, lower.tail = TRUE, log.p = FALSE)

## 1.2 Pruebas de Hipótesis Comúnmente Usadas

En esta sección se describen brevemente algunas pruebas de hipótesis que son comúnmente usadas en la práctica. Para el desarrollo de las mismas se usará el resultado obtenido en la sección anterior especificando la matriz $\boldsymbol{H}$ y el vector $\boldsymbol{h}$ dependiendo de la hipótesis a probar.

### 1.2.1 Todos los coeficientes son iguales a cero

Se quiere probar la hipótesis de que todos los coeficientes son iguales a cero, es decir, $H_0 : \beta_0 = \beta_1 = \cdots \beta_k = 0$. Para ello vamos a escribir la hipótesis en la forma $\boldsymbol{H\beta} = \boldsymbol{h}$.

La hipótesis $H_0 : \beta_0 = \beta_1 = \cdots \beta_k = 0$ es equivalente a $H_0 : \beta_0 = 0, \beta_2 = 0, \cdots, \beta_k = 0$, lo cual en forma matricial se puede escribir como

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Por lo tanto,

$$\boldsymbol{H} = \mathbf{I}_{k+1} \qquad \boldsymbol{h} = 0 \qquad q = k+1$$

Dado que el estadístico de prueba para $\boldsymbol{H\beta} = \boldsymbol{h}$ está dado por

$$F = \frac{(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})/q}{\boldsymbol{y}'[\mathbf{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y}/(n-p)}$$

entonces en este caso

- $\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h} = \mathbf{I}\hat{\boldsymbol{\beta}} - \mathbf{0} = \hat{\boldsymbol{\beta}}$
- $[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1} = [\mathbf{I}(\boldsymbol{X}'\boldsymbol{X})^{-1}\mathbf{I}']^{-1} = [(\boldsymbol{X}'\boldsymbol{X})^{-1}]^{-1} = \boldsymbol{X}'\boldsymbol{X}$

Así,

$$F = \frac{\hat{\boldsymbol{\beta}}'(\boldsymbol{X}'\boldsymbol{X})\hat{\boldsymbol{\beta}}/(k+1)}{\boldsymbol{y}'[\mathbf{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y}/(n-p)}$$

### 1.2.2 Los coeficientes son iguales a un valor cualquiera

Se quiere probar la hipótesis $H_0 : \beta_0 = b_0, \beta_1 = b_1, \cdots, \beta_k = b_k$. Procediendo como en la prueba anterior se tiene que la hipótesis en forma matricial se puede escribir como

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{pmatrix}$$

Por lo tanto,

$$\boldsymbol{H} = \mathbf{I}_{k+1} \qquad \boldsymbol{h} = \begin{pmatrix} b_0 & b_1 & \cdots & b_k \end{pmatrix}' = \boldsymbol{b} \qquad q = k+1$$

entonces en este caso

- $\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h} = \mathbf{I}\hat{\boldsymbol{\beta}} - \boldsymbol{b} = \hat{\boldsymbol{\beta}} - \boldsymbol{b}$
- $[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1} = [\mathbf{I}(\boldsymbol{X}'\boldsymbol{X})^{-1}\mathbf{I}']^{-1} = [(\boldsymbol{X}'\boldsymbol{X})^{-1}]^{-1} = \boldsymbol{X}'\boldsymbol{X}$

Así,

$$F = \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{b})'(\boldsymbol{X}'\boldsymbol{X})(\hat{\boldsymbol{\beta}} - \boldsymbol{b})/(k+1)}{\boldsymbol{y}'[\mathbf{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y}/(n-p)}$$

### 1.2.3 Un subconjunto de $\boldsymbol{\beta}$ es cero.

Supongamos que se desea probar la hipótesis $H_0 : \boldsymbol{\beta}_h = 0$, donde $\boldsymbol{\beta}_h$ es un vector que contiene los primeros $h$ elementos del vector $\boldsymbol{\beta}$, $0 < h < p$. Dicha hipótesis se puede escribir en forma matricial de la siguiente manera

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_h \\ \beta_{h+1} \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Donde la Matriz $\boldsymbol{H}$ se puede escribir de manera particionada como

$$\boldsymbol{H} = \left(\begin{array}{ccccc:ccc} 1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \end{array}\right) = \begin{pmatrix} \mathbf{I}_h & \mathbf{0} \end{pmatrix}$$

y la siguiente partición del vector $\boldsymbol{\beta}$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_h \\ \hdashline \beta_{h+1} \\ \vdots \\ \beta_p \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_h \\ \boldsymbol{\beta}_{p-h} \end{pmatrix}$$

Tomaremos $m = p - h$

Entonces

$$\boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{h} = \begin{pmatrix} \mathbf{I}_h & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_h \\ \boldsymbol{\beta}_m \end{pmatrix} = \boldsymbol{\beta}_1$$

lo cual implica que $\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h} = \hat{\boldsymbol{\beta}}_h$

Por otro lado, la partición de $\boldsymbol{\beta}$ genera la siguiente partición de la matriz $\boldsymbol{X}$

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_h & \boldsymbol{X}_m \end{pmatrix}$$

entonces

$$\boldsymbol{X}'\boldsymbol{X} = \begin{matrix} \boldsymbol{X}'_h \\ \boldsymbol{X}'_m \end{matrix} \begin{pmatrix} \boldsymbol{X}_{p-h} & \boldsymbol{X}_m \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}'_h\boldsymbol{X}_h & \boldsymbol{X}'_h\boldsymbol{X}_m \\ \boldsymbol{X}'_m\boldsymbol{X}_h & \boldsymbol{X}'_m\boldsymbol{X}_m \end{pmatrix}$$

y así

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} \boldsymbol{X}'_h\boldsymbol{X}_h & \boldsymbol{X}'_h\boldsymbol{X}_m \\ \boldsymbol{X}'_m\boldsymbol{X}_h & \boldsymbol{X}'_m\boldsymbol{X}_m \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{T}_{hh} & \boldsymbol{T}_{hm} \\ \boldsymbol{T}_{mh} & \boldsymbol{T}_{mm} \end{pmatrix}$$

con lo cual

$$\begin{aligned} \boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}' &= \begin{pmatrix} \mathbf{I}_h & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{T}_{hh} & \boldsymbol{T}_{hm} \\ \boldsymbol{T}_{mh} & \boldsymbol{T}_{mm} \end{pmatrix} \begin{pmatrix} \mathbf{I}_h \\ \mathbf{0} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{T}_{hh} & \boldsymbol{T}_{hm} \end{pmatrix} \begin{pmatrix} \mathbf{I}_h \\ \mathbf{0} \end{pmatrix} = \boldsymbol{T}_{hh} \end{aligned}$$

y por consiguiente

$$[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1} = \boldsymbol{T}_{hh}^{-1}$$

Por lo tanto el estadístico de prueba sería

$$F = \frac{\hat{\boldsymbol{\beta}}'_h \boldsymbol{T}_{hh}^{-1} \hat{\boldsymbol{\beta}}_h / h}{\boldsymbol{y}'[\mathbf{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y}/(n-p)}$$

**Nota:** Se deja como tarea probar la significancia de la regresión, la cual consiste en probar que todos los $\beta_j = 0$ excepto $\beta_0$.

### 1.2.4 Prueba de la hipótesis de una combinación lineal de los $\beta_j$.

Ahora se quiere probar la hipótesis $H_0 : \sum_{j=0}^{k} a_j\beta_j = 0$. Escribiendo la hipótesis en la forma $\boldsymbol{H\beta} = \boldsymbol{h}$ se tiene en este caso $\boldsymbol{H} = \boldsymbol{a}'$ donde **a** es un vector $p \times 1$ y $\boldsymbol{h} = 0$ es un escalar. Entonces

- $\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h} = \boldsymbol{a}'\hat{\boldsymbol{\beta}}$ lo cual es un escalar
- $[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1} = [\boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{a}]^{-1}$ que también es un escalar

Por lo tanto

$$\begin{aligned} F &= \frac{(\boldsymbol{a}'\hat{\boldsymbol{\beta}})'[\boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{a}]^{-1}(\boldsymbol{a}'\hat{\boldsymbol{\beta}})}{\boldsymbol{y}'[\mathbf{I}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y}/(n-p)} \\ &= \frac{(\boldsymbol{a}'\hat{\boldsymbol{\beta}})'(\boldsymbol{a}'\hat{\boldsymbol{\beta}})}{\boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{a}\boldsymbol{y}'[\mathbf{I}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y}/(n-p)} \\ &= \frac{(\boldsymbol{a}'\hat{\boldsymbol{\beta}})^2}{\boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{a}\boldsymbol{y}'[\mathbf{I}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y}/(n-p)} \\ &= \frac{(\boldsymbol{a}'\hat{\boldsymbol{\beta}})^2}{\boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{a}S^2} \end{aligned} \tag{1.32}$$

donde $S^2 = \boldsymbol{y}'[\mathbf{I}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y}/(n-p)$

**Nota:** Se deja como tarea probar la significancia de un parámetro. $H_0 : \beta_j = 0$.

## 1.3 Intervalos de Confianza

En esta sección consideramos una región de confianza para $\boldsymbol{\beta}$, intervalos de confianza para $\beta_j$, $\mathbf{a}'\boldsymbol{\beta}$, individuales y simultáneos, y $\sigma^2$. Asumiremos que $\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma^2\mathbf{I})$ .

### 1.3.1 Región de confianza para $\boldsymbol{\beta}$

Recordemos que el estadístico de prueba para la hipótesis $\boldsymbol{H\beta} = \boldsymbol{h}$ está dado por

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

donde $S^2 = \boldsymbol{y}'[\mathbf{I}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y}/(n-p)$.

Como queremos hallar una región de confianza para $\boldsymbol{\beta}$, entonces $\boldsymbol{H} = \mathbf{I}$ y $\boldsymbol{h} = \boldsymbol{\beta}$. Por lo tanto podemos hacer la siguiente declaración de probabilidad

$$P\left[\frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{qS^2} \le F_{\alpha,q,n-p}\right] = 1-\alpha$$

Desde esta declaración, una región de confianza conjunta de $100(1-\alpha)$ para $\beta_0, \beta_1, \cdots, \beta_k$ en $\boldsymbol{\beta}$ esta definida por todos los vectores de $\boldsymbol{\beta}$ que satisfacen

$$(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \le qS^2F_{\alpha,q,n-p} = 1-\alpha \tag{1.33}$$

Para $q = 2$, esta región puede ser graficada como una elipse en 2 dimensiones. Para $q > 2$, la región elipsoidal es difícil de interpretar.

### 1.3.2 Intervalos de confianza individuales

Existen ocasiones en las que el investigador esta interesado en hallar intervalos de confianza para cada parámetro individualmente o para solo una combinación lineal de ellos, en dicho caso se habla de intervalos de confianza individuales, pues la construcción de dicho intervalo no esta influenciado por los otros parámetros. A continuación veamos la construcción de los intervalos.

***1.3.2.1 Intervalo de confianza para cada $\boldsymbol{\beta_j}$*** En la sección **** se obtuvo que el estadístico de prueba para la hipótesis relacionada con cada $\beta_j$ estaba dada por

$$t_j = \frac{\hat{\beta}_j - \beta_j}{S\sqrt{g_{jj}}} \sim t_{n-p}$$

donde $g_{jj}$ es el $j$ ésimo elemento de la diagonal de $(\boldsymbol{X}'\boldsymbol{X})^{-1}$. Entonces

$$P\left[-t_{alpha/2,n-p} \leq \frac{\hat{\beta}_j - \beta_j}{S\sqrt{g_{jj}}} \leq t_{alpha/2,n-p}\right] = 1-\alpha$$

Resolviendo la desigualdad para $\beta_j$ se obtiene

$$P\left[\hat{\beta}_j - t_{alpha/2,n-p}S\sqrt{g_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{alpha/2,n-p}S\sqrt{g_{jj}}\right] = 1-\alpha$$

Antes de tomar la muestra, la probabilidad de que el intervalo aleatorio contenga a $\beta_j$ es $1-\alpha$. Después de tomar la muestra, el intervalo de confianza de $100(1-\alpha)\%$ para $\beta_j$ es

$$\hat{\beta}_j \mp t_{alpha/2,n-p}S\sqrt{g_{jj}} \tag{1.34}$$

ya no es aleatorio, y así decimos que tenemos $100(1-\alpha)\%$ de confianza de que el intervalo contenga a $\beta_j$.

Note que el coeficiente de confianza $1-\alpha$ se obtiene sólo para un intervalo de confianza para uno de los $\beta_j$. Para intervalos de confianza para todos los $k+1$ elementos de $\boldsymbol{\beta}$ que se obtienen simultáneamente con coeficiente de confianza total se verán más adelante.

***1.3.2.2 Intervalo de confianza para cada $\boldsymbol{a'\beta}$*** Si $\mathbf{a}'\boldsymbol{\beta} \neq 0$ podemos restar $\mathbf{a}'\boldsymbol{\beta}$ de $\mathbf{a}'\hat{\boldsymbol{\beta}}$ en (1.32) para obtener

$$W = \frac{(\boldsymbol{a}'\hat{\boldsymbol{\beta}} - \boldsymbol{a}\boldsymbol{\beta})^2}{\boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{a}S^2} \sim F(1, n-p)$$

lo cual es equivalente a

$$T = \frac{\boldsymbol{a}'\hat{\boldsymbol{\beta}} - \boldsymbol{a}\boldsymbol{\beta}}{S\sqrt{\boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{a}}} \sim t(n-p) \tag{1.35}$$

Por lo tanto, un intervalo de confianza de $100(1-\alpha)\%$ para un sólo valor de $\boldsymbol{a}'\boldsymbol{\beta}$ está dados por

$$\boldsymbol{a}'\hat{\boldsymbol{\beta}} \pm t_{\alpha/2,n-p}S\sqrt{\boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{a}} \tag{1.36}$$

***1.3.2.3 Intervalo de confianza para cada $\sigma^2$*** Hemos demostrado que

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

Por lo tanto,

$$P\left[\chi^2_{1-\alpha/2,n-p} \leq \frac{(n-p)S^2}{\sigma^2} \leq \chi^2_{\alpha/2,n-p}\right] = 1-\alpha \tag{1.37}$$

donde $\chi^2_{\alpha/2,n-p}$ es el punto de la distribución chi-cuadrado cuya área derecha es $\alpha/2$ y $\chi^2_{1-\alpha/2,n-p}$ es el punto de la distribución chi-cuadrado cuya área izquierda es $\alpha/2$. Resolviendo la desigualdad para $\sigma^2$ se obtiene el intervalo de confianza de $100(1-\alpha)\%$

$$\frac{(n-p)S^2}{\chi^2_{\alpha/2,n-p}} \leq \sigma^2 \leq \frac{(n-p)S^2}{\chi^2_{1-\alpha/2,n-p}} \tag{1.38}$$

un intervalo de confianza de $100(1-\alpha)\%$ para $\sigma$ está dado por

$$\sqrt{\frac{(n-p)S^2}{\chi^2_{\alpha/2,n-p}}} \leq \sigma \leq \sqrt{\frac{(n-p)S^2}{\chi^2_{1-\alpha/2,n-p}}} \tag{1.39}$$

### 1.3.3 Intervalos de confianza simultáneos

El objetivo de los intervalos de confianza presentados en la sección anterior es proveer conclusiones con un nivel prefijado de confianza sobre cada uno de los parámetros $\beta_0, \beta_1, ..., \beta_k$ por separado. La dificultad es que éstos no proporcionan el nivel de confianza fijado de que las conclusiones de los $p$ intervalos son correctas. Si las inferencias fueran independientes, la probabilidad de que los $p$ intervalos construidos cada uno con un nivel de confianza de $1-\alpha$, contengan al verdadero parámetro sería $(1-\alpha)^p$, el cual es menor a $1-\alpha$. Sin embargo, las inferencias no son independientes, ya que son calculadas a partir de un mismo conjunto de datos de la muestra, lo que hace que la determinación de la probabilidad de que todas las inferencias sean correctas sea mucho más difícil.

En esta sección propondremos intervalos de confianza de nivel conjunto $1-\alpha$, esto quiere decir que nos gustaría construir una serie de intervalos para los cuales tengamos una garantía sobre la exactitud de todo el conjunto de intervalos de confianza. Al conjunto de intervalos de confianza de interés lo llamaremos familias de intervalos de confianza de nivel conjunto o simultáneo (o regiones de confianza de nivel simultáneo).

Supongamos que la familia se compone de $p$ estimaciones, para $\beta_0, \beta_1, ..., \beta_{p-1}$. Podríamos estar interesados en construir regiones de confianza para una cantidad $g$ entre 1 y $p$ de estos parámetros, con $g$ prefijado. Distingamos entre un intervalo de confianza de nivel $1-\alpha$ para un parámetro, y una familia de intervalos de nivel simultáneo $1-\alpha_c$ para $g$ parámetros. En el primer caso, $1-\alpha$ es la proporción de intervalos construidos con el método en cuestión que cubren al verdadero parámetro de interés cuando se seleccionan repetidamente muestras de la población de interés y se construyen los intervalos de confianza para cada una de ellas. Por otro lado, cuando construimos una familia de regiones o intervalos de confianza de nivel simultáneo $1-\alpha$ para $g$ parámetros, $\theta_1, \theta_2, ..., \theta_g$, el valor $1-\alpha_c$ indica la proporción de familias de $g$ intervalos que están enteramente correctos (cubren a los $g$ parámetros de interés, simultáneamente) cuando se seleccionan repetidamente muestras de la población de interés y se construyen los intervalos de confianza específicos para los $g$ parámetros en cuestión, o sea

$$P(\{\theta_1 \in I_1\} \cap \{\theta_2 \in I_2\} \cap ... \cap \{\theta_g \in I_g\}) = 1 - \alpha_c$$

si $I_1, ..., I_g$ son los $g$ intervalos construidos usando los mismos datos. Luego el nivel simultáneo de una familia de regiones o intervalos de confianza corresponde a la probabilidad, calculada previa al muestreo, de que la familia entera de afirmaciones sea correcta.

En general es sumamente deseable contar con un procedimiento que provea una familia de intervalos de confianza de nivel simultáneo cuando se estiman varios parámetros con una misma muestra de datos, ya que le permite al analista entrelazar varios resultados juntos en un conjunto integrado de conclusiones con la seguridad de que todo el conjunto de inferencias es correcto.

Para obtenerlos hay básicamente dos herramientas estadísticas disponibles. Una de ellas es el estudio matemático en detalle del fenómeno en cuestión, en este caso, estudiar matemáticamente las propiedades de los estimadores $\hat{\beta}_0, ..., \hat{\beta}_{p-1}$ de manera de poder obtener la distribución exacta de alguna medida numérica que los resuma, como el $\max_{\{} 0 \leq k \leq p-1\}|\hat{beta}_k|$ o las descripciones matemáticas del elipsoide $p$ dimensional más pequeño que los contenga, con probabilidad $1 - \alpha_c$. La otra herramienta consiste en construir intervalos de confianza con nivel simultáneo a partir de ajustar el nivel de confianza de cada intervalo individual a un valor más alto, de modo de poder asegurar el nivel simultáneo de la construcción. Esto es lo que se conoce como el método de Bonferroni para la construcción de intervalos de nivel simultáneo.