Observemos que en el gráfico la recta ajustada *está encerrada* entre 2 curvas interiores y 2 exteriores. Las exteriores corresponden al intervalo de predicción de nivel 0.95 y las interiores a los intervalos de confianza de nivel 0.95 para la media.

Notemos que **el nivel de confianza 0.95 se aplica a cada punto y no es global**

Supongamos que queremos plantear un test de nivel $\alpha$ para

$$\boxed{H_o : \mathbf{C}\boldsymbol{\beta} = \boldsymbol{\delta} \text{ vs. } H_1 : \mathbf{C}\boldsymbol{\beta} \neq \boldsymbol{\delta}}$$

siendo $\text{rg}(\mathbf{C}) = q$ , $\mathbf{C} \in \Re^{q\times p}$.

Sea $\Psi = \mathbf{C}'\boldsymbol{\beta}$. Sabemos que $\widehat{\Psi} \sim N_q(\Psi, \sigma^2 \mathbf{A}^* \mathbf{A}^{*'}) = N_q(\Psi, \sigma^2 \mathbf{B})$. Por lo tanto, tenemos que

$$(1): \quad Q = \frac{1}{q}(\widehat{\Psi} - \boldsymbol{\delta})'\mathbf{B}^{-1}(\widehat{\Psi} - \boldsymbol{\delta})$$

es independiente de

$$(2): \quad s^2 = \frac{\|\mathbf{Y} - \widehat{\mathbf{Y}}\|^2}{n-r}$$

Veremos que

$$E(Q) = \sigma^2 + \eta^2$$

y que $\eta^2 = 0$ sólo cuando $H_o$ es cierta.

Bajo $H_o$, (1) y (2) son estimadores insesgados de $\sigma^2$, es decir que bajo $H_o$ esperamos que

$$\frac{(1)}{(2)} \simeq 1,$$

pero si $H_o$ no es cierta, esperamos que

$$\frac{(1)}{(2)} > 1.$$

Luego, el cociente $\dfrac{(\widehat{\Psi} - \boldsymbol{\delta})'\mathbf{B}^{-1}(\widehat{\Psi} - \boldsymbol{\delta})}{qs^2}$ nos dará una idea de la *veracidad de* $H_O$, de manera que rechazaremos $H_o$ si el cociente es grande.

¿ Cuán grande?

Bajo $H_o$

$$\frac{(\widehat{\Psi} - \boldsymbol{\delta})'\mathbf{B}^{-1}(\widehat{\Psi} - \boldsymbol{\delta})}{\sigma^2} \sim \chi^2_q$$

independiente de

$$\frac{(n-r)s^2}{\sigma^2} \sim \chi^2_{n-r}$$

En consecuencia:

$$F = \frac{(\widehat{\Psi} - \boldsymbol{\delta})'\mathbf{B}^{-1}(\widehat{\Psi} - \boldsymbol{\delta})}{qs^2} \sim \mathcal{F}_{q,n-r}$$

Rechazaremos $H_o$ si

$$F > \mathcal{F}_{q,n-r,\alpha}$$

Veamos dos situaciones frecuentes para el caso de rango completo.

## 1. Una hipótesis simple.

$\mathbf{C} = \mathbf{c}$ consiste en una sola fila, de manera que $\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}$ es un escalar, con lo cual el estadístico resulta

$$F = \frac{(\mathbf{c}'\widehat{\boldsymbol{\beta}} - \delta)^2}{s^2\,\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}$$

que bajo $H_o$ tiene distribución $F_{1,n-p}$

En función de la relación entre las distribuciones $t$ y $F$ podríamos utilizar la distribución $t$ de Student y

$$\text{rechazamos } H_o \text{ si } \left|\frac{c'\widehat{\boldsymbol{\beta}} - \delta}{s\,\sqrt{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}}\right| > t_{n-p,\alpha/2}$$

## 2. Tests para $k$ coeficientes iguales a 0.

$H_o : \Psi = \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, donde $\mathbf{C} = \begin{pmatrix} e_{i_1} \\ \cdot \\ \cdot \\ e_{i_k} \end{pmatrix}$, para $i_1 \leq 1 < \ldots < i_k \leq p$.

El numerador será:

$$(\mathbf{C}\widehat{\boldsymbol{\beta}})'(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\widehat{\boldsymbol{\beta}})$$

donde $\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'$ es una submatriz de $\mathbf{D} = (\mathbf{X}'\mathbf{X})^{-1}$ que sólo involucra los coeficientes correspondientes a aquellos $\beta_i$ presentes en la hipótesis a testear.

Así supongamos que tenemos 5 coeficientes $\beta_1, \ldots, \beta_5$ y queremos testear

$$\begin{aligned} H_o : \beta_1 &= 0 \\ \beta_3 &= 0 \\ \beta_5 &= 0 \end{aligned}$$

luego,

$$\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}' = \begin{pmatrix} d_{11} & d_{13} & d_{15} \\ d_{31} & d_{33} & d_{35} \\ d_{51} & d_{35} & d_{55} \end{pmatrix}$$

y en el numerador tendremos

$$(\widehat{\beta_1}, \widehat{\beta_3}, \widehat{\beta_5}) \begin{pmatrix} d_{11} & d_{13} & d_{15} \\ d_{31} & d_{33} & d_{35} \\ d_{51} & d_{35} & d_{55} \end{pmatrix}^{-1} \begin{pmatrix} \widehat{\beta_1} \\ \widehat{\beta_3} \\ \widehat{\beta_5} \end{pmatrix}$$

## Test de Cociente de Verosimilitud

El test de $F$ también puede motivarse como test de cociente de verosimilitud. Ses $\Omega$ es el conjunto de supuestos generales y supongamos bajo este modelo testeamos la hipótesis $H$, llamemos $\omega = \Omega \cap H$. Así, por ejemplo, si

$$\Omega : \quad \mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}) \boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})'$$

y

$$H : \beta_1 = \ldots = \beta_{p-1} = 0$$

entonces

$$\mathbf{Y} \sim N_n(\beta_0, \sigma^2 \mathbf{I})$$

Si $p(\mathbf{y})$ es la función de densidad o de probabilidad de $\mathbf{Y}$ definimos $\lambda$ el estadístico del cociente de verosimilitud como

$$\lambda = \frac{\text{máx}_\omega \, p(\mathbf{y})}{\text{máx}_\Omega \, p(\mathbf{y})}$$

Notemos que $0 \leq \lambda \leq 1$ ya que $\omega \in \Omega$ y por lo tanto $\text{máx}_\omega \, p(\mathbf{y}) \leq \text{máx}_\Omega \, p(\mathbf{y})$.

$H$ será rechazada cuando $\max_\omega p(\mathbf{y})$ es mucho más chico que $\max_\Omega p(\mathbf{y})$, por lo tanto rechazaremos $H$ si $\lambda < \lambda_\alpha$.

Existen dos formas equivalentes de plantear las hipótesis:

- 1)

$$\begin{aligned} \Omega &: \mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \mathrm{rg} = r \\ H &: \Psi_1 = \Psi_2 = \ldots = \Psi_q = 0 \end{aligned}$$

donde $\{\Psi_i\}$ son l.i. funciones estimables

- 2)

$$\begin{aligned} \Omega &: \mathbf{Y} \sim N_n(\boldsymbol{\eta}, \sigma^2\mathbf{I}) \quad \boldsymbol{\eta} \in \mathcal{V}_r \\ H &: \boldsymbol{\eta} \in \mathcal{V}_{r-q} \end{aligned}$$

donde $\mathcal{V}_r$ es un subespacio de dimensión $r$ en $\Re^n$ y $\mathcal{V}_{r-q}$ es un subespacio de dimensión $r-q$ en $\mathcal{V}_r$.

$\mathcal{V}_r$ es el espacio generado por las columnas de $\mathbf{X}$ y $\mathcal{V}_{r-q}$ es el espacio al cual está restringido $\boldsymbol{\eta}$ a yacer al imponerle las restricciones $\Psi_1 = \Psi_2 = \ldots = \Psi_q = 0$.

Probaremos que • 1) $\Longrightarrow$ • 2).

Tenemos que $\mathbf{Y} \sim N_n(\boldsymbol{\eta}, \sigma^2 \mathbf{I}) \quad \boldsymbol{\eta} \in \mathcal{V}_r$. Llamemos $\mathbf{C}$ a la matriz tal que $\Psi = \mathbf{C}\boldsymbol{\beta}$. Luego:

$$\begin{aligned} V_\omega &= \{\mathbf{v} = \mathbf{X}\boldsymbol{\beta} \text{ tal que } \mathbf{C}\boldsymbol{\beta} = \mathbf{0}\} = \{\mathbf{v} = \mathbf{X}\boldsymbol{\beta} \text{ tal que } \mathbf{A}^*\mathbf{X}\boldsymbol{\beta} = \mathbf{0}\} \\ &= \{v = \mathbf{X}\boldsymbol{\beta} \text{ tal que } \mathbf{A}^*\mathbf{v} = \mathbf{0}\} \end{aligned}$$

$\mathbf{A}^* = \begin{pmatrix} \mathbf{a}_1^* \\ . \\ . \\ . \\ \mathbf{a}_q^* \end{pmatrix}$. Es decir, $\mathbf{v} \perp \mathbf{a}_i^* \quad 1 \leq i \leq q$.

Como rg $= q$ entonces $\mathbf{a}_1^*, \ldots, \mathbf{a}_q^*$ son l.i. Por lo tanto, $\mathbf{v} \in \mathcal{V}^{\perp}_{<\mathbf{a}_1^* \ldots \mathbf{a}_q^*>}$ : complemento ortogonal de $\mathcal{V}_{<\mathbf{a}_1^* \ldots \mathbf{a}_q^*>}$ en $\mathcal{V}_r$.

Además, tenemos que

$$r = \dim(\mathcal{V}_{<\mathbf{a}_1^* \ldots \mathbf{a}_q^*>}) + \dim(\mathcal{V}^{\perp}_{<\mathbf{a}_1^* \ldots \mathbf{a}_q^*>})$$

por lo tanto,

$$\dim(\mathcal{V}^{\perp}_{<\mathbf{a}_1^* \ldots \mathbf{a}_q^*>}) = dim(\mathcal{V}_\omega) = r - q$$

Calculemos $\lambda$. Para ello deberemos calcular el máximo de de $p(\mathbf{y})$ en c/u de los subespacios.

Veremos que $\lambda = \left(\frac{\|\mathbf{y} - \widehat{\boldsymbol{\eta}}\|^2}{\|\mathbf{y} - \widehat{\boldsymbol{\eta}}_\omega\|^2}\right)^{n/2}$ y por lo tanto rechazamos $H_0$ si

$$\lambda = \left(\frac{\|\mathbf{y} - \widehat{\boldsymbol{\eta}}\|^2}{\|\mathbf{y} - \widehat{\boldsymbol{\eta}}_\omega\|^2}\right)^{n/2} < k_\alpha$$

Si aplicamos a este cociente la función $g(t) = \frac{n-r}{q}\left(t^{-2/n} - 1\right)$, resulta

$$\begin{aligned} F &= \frac{n-r}{q} \, \frac{\|\mathbf{y} - \widehat{\boldsymbol{\eta}}_\omega\|^2 - \|\mathbf{y} - \widehat{\boldsymbol{\eta}}\|^2}{\|\mathbf{y} - \widehat{\boldsymbol{\eta}}\|^2} \\ &= \frac{1}{q} \, \frac{\|\mathbf{y} - \widehat{\boldsymbol{\eta}}_\omega\|^2 - \|\mathbf{y} - \widehat{\boldsymbol{\eta}}\|^2}{s^2} \end{aligned}$$

Como veremos

$$F = \frac{1}{q} \frac{\|\widehat{\boldsymbol{\eta}}_\omega - \widehat{\boldsymbol{\eta}}\|^2}{s^2}$$

Luego, rechazaremos $H$ si

$$\frac{1}{q} \frac{\|\widehat{\boldsymbol{\eta}}_\omega - \widehat{\boldsymbol{\eta}}\|^2}{s^2} > \lambda_\alpha$$

Una interpretación intuitiva para este test es que $\|\mathbf{y} - \widehat{\boldsymbol{\eta}}_\omega\|^2$ y $\|\mathbf{y} - \widehat{\boldsymbol{\eta}}\|^2$ miden cuan bien ajustan los modelos $\omega$ y $\Omega$, respectivamente. Por lo tanto, su cociente compara el ajuste de $\omega$ con el de $\Omega$ y rechazamos $H$ si este cociente es grande:

$$F > \lambda_\alpha$$

## ¿ Qué distribución tiene el cociente $F$?

Tenemos que $\mathcal{V}_{r-q} \in \mathcal{V}_r \in \Re^n$. Tomemos una base ortonormal de $\mathcal{V}_{r-q}$ : $\{\boldsymbol{\alpha}_{q+1}, \ldots \boldsymbol{\alpha}_r\}$ y la extendemos a una base ortonormal de $\mathcal{V}_r : \{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_q, \boldsymbol{\alpha}_{q+1}, \ldots, \boldsymbol{\alpha}_r\}$ y finalmente a una de $\Re^n$: $\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r, \boldsymbol{\alpha}_{r+1}, \ldots, \boldsymbol{\alpha}_n\}$.

$$\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_q, \boldsymbol{\alpha}_{q+1}, \dots, \boldsymbol{\alpha}_r, \boldsymbol{\alpha}_{r+1}, \dots, \boldsymbol{\alpha}_n$$

Por lo tanto,

$$\mathbf{y} \in \Re^n : \mathbf{y} = \sum_{j=1}^{n} z_j \boldsymbol{\alpha}_j$$

y tenemos que

$$\boldsymbol{\alpha}_i'\mathbf{y} = \sum_{j=1}^{n} z_j \boldsymbol{\alpha}_i' \boldsymbol{\alpha}_j = z_i \boldsymbol{\alpha}_i' \boldsymbol{\alpha}_i = z_i \quad \forall i = 1, \dots, n$$

Luego, si definimos a $\mathbf{T}$ como la matriz que tiene filas $\boldsymbol{\alpha}_i'$, entonces

$$\mathbf{z} = \mathbf{T}\mathbf{y}$$

Observemos que bajo el modelo $\Omega$

$$E(z_i) = \begin{cases} \boldsymbol{\alpha}_i'\boldsymbol{\eta} = \xi_i & \text{si } 1 \leq i \leq r \\ 0 & \text{si } r+1 \leq i \leq n \end{cases}$$

$$\Sigma_{\mathbf{z}} = \mathbf{T}\Sigma_{\mathbf{y}}\mathbf{T}' = \sigma^2 \mathbf{I}$$

Bajo el modelo $\omega$, tenemos que $\boldsymbol{\eta} = E(\mathbf{Y}) \in \mathcal{V}_{r-q}$, es decir $\boldsymbol{\alpha}_i'\boldsymbol{\eta} = 0$ para $i = 1, \ldots, q$.

$$E(z_i) = \begin{cases} 0 & \text{si } 1 \leq i \leq q \\ \xi_i & \text{si } q+1 \leq i \leq r \\ 0 & \text{si } r+1 \leq i \leq n \end{cases}$$

Entonces podemos escribir:

$$\begin{aligned} \Omega &: \mathbf{z} \sim N_n(\boldsymbol{\xi}, \sigma^2\mathbf{I}) \quad \xi_i = 0 \quad i \geq r+1 \\ \omega &: \mathbf{z} \sim N_n(\boldsymbol{\xi}, \sigma^2\mathbf{I}) \quad \xi_i = 0 \quad 1 \leq i \leq q \text{ y } i \geq r+1 \end{aligned}$$

Utilizando la notación de Scheffé tendremos

$$\begin{aligned} \mathcal{S}_\Omega &= \|\mathbf{y} - \widehat{\boldsymbol{\eta}}\|^2 = \sum_{i=r+1}^{n} z_i^2 \\ \mathcal{S}_\omega &= \|\mathbf{y} - \widehat{\boldsymbol{\eta}}_\omega\|^2 = \sum_{i=1}^{q} z_i^2 + \sum_{i=r+1}^{n} z_i^2 \end{aligned}$$

y además

$$\mathcal{S}_\omega - \mathcal{S}_\Omega = \sum_{i=1}^{q} z_i^2$$

$$\|\mathbf{y}-\widehat{\boldsymbol{\eta}}_\omega\|^2-\|\mathbf{y}-\widehat{\boldsymbol{\eta}}\|^2=\|\widehat{\boldsymbol{\eta}}-\widehat{\boldsymbol{\eta}}_\omega\|^2=\sum_{i=1}^{q} z_i^2$$

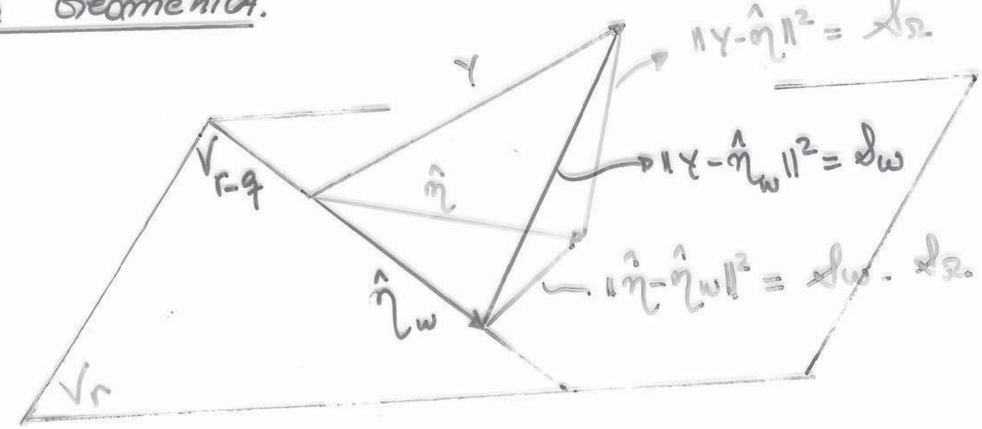
Además, bajo $H$ tenemos que $\frac{\mathcal{S}_\omega-\mathcal{S}_\Omega}{\sigma^2}\sim\chi_q$ independiente de $s^2$ y en consecuencia

$$\frac{1}{q}\frac{\mathcal{S}_\omega-\mathcal{S}_\Omega}{s^2}\sim F_{q,n-r}$$

$$\text{Rechazamos } H \text{ si } \frac{1}{q}\frac{\mathcal{S}_\omega-\mathcal{S}_\Omega}{s^2}>F_{q,n-r,\alpha}$$

**Observación**: Puede demostrarse que este test es equivalente al tests de $F$ ya visto.

## Interpretación Geométrica.



## Ejemplo:

## Significación de la Regresión. Tabla de Análisis de la Varianza

Supongamos que tenemos el modelo con intercept dado por

$$E(\mathbf{Y}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_{p-1} x_{p-1}$$

y queremos testear

$$H_o : \beta_1 = \beta_2 = \ldots = \beta_{p-1} = 0$$

de manera que $\omega = \Omega \cap H$. $H$ impone $p-1$ restricciones l.i. Trataremos el caso en que $\mathrm{rg}(\mathbf{X}) = p$

¿Quién es $\mathcal{V}_\omega$?

$dim(\mathcal{V}_\omega) = r - (p-1) = p - (p-1) = 1$ y tenemos que $\mathcal{V}_1 \in \mathcal{V}_p$

¿Quién es $\widehat{\boldsymbol{\eta}}_\omega$?

Bajo $\omega$, $\beta_1 = \ldots = \beta_{p-1} = 0, \quad E(\mathbf{Y}) = \beta_0$.

Tenemos que:

$$\mathbf{X}_\omega = \begin{pmatrix} 1 \\ 1 \\ . \\ . \\ 1 \end{pmatrix} \rightarrow \widehat{\beta}_0 = (\mathbf{X}'_\omega \mathbf{X}_\omega)^{-1} \mathbf{X}'_\omega \mathbf{Y} = \frac{1}{n} \sum_{i=1}^{n} y_i = \overline{\mathbf{Y}}.$$

Luego:

$$\widehat{\boldsymbol{\eta}}_\omega = \mathbf{X}'_\omega \widehat{\beta}_0 = \begin{pmatrix} \overline{\mathbf{Y}} \\ \overline{\mathbf{Y}} \\ . \\ . \\ \overline{\mathbf{Y}} \end{pmatrix}$$

Además:

$$\|\mathbf{Y}\|^2 = \|\mathbf{Y} - \widehat{\boldsymbol{\eta}}\|^2 + \|\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{\eta}}_\omega\|^2 + \|\widehat{\boldsymbol{\eta}}_\omega\|^2$$

Bajo $\Omega$ si $\mathrm{rg}(\mathbf{X}) = p$

$$\widehat{\beta}_\Omega = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \rightarrow \widehat{\boldsymbol{\eta}} = \mathbf{P}\mathbf{Y} \text{ donde } \mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

En efecto, $\widehat{\boldsymbol{\eta}}_\omega$ es la proyección ortogonal de $\widehat{\boldsymbol{\eta}}$ sobre $\mathcal{V}_\omega = \mathcal{V}_1$. Si fuera así, entonce $\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{\eta}}_\omega \perp \widehat{\boldsymbol{\eta}}_\omega$.

$$\widehat{\boldsymbol{\eta}} = \mathbf{PY} \text{ y } \widehat{\boldsymbol{\eta}}_\omega = \mathbf{1}(\mathbf{1}'\mathbf{1})^{-\mathbf{1}}\mathbf{1}'\mathbf{Y} = \mathbf{P_1Y}$$

luego,

$$(\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{\eta}}_\omega)'\widehat{\boldsymbol{\eta}}_\omega = \mathbf{Y}'(\mathbf{P} - \mathbf{P}_\omega)\mathbf{P}_\omega\mathbf{Y} = \mathbf{Y}'(\mathbf{P}\mathbf{P}_\omega - \mathbf{P}_\omega'\mathbf{P}_\omega)\mathbf{Y} = \mathbf{Y}'(\mathbf{P} - \mathbf{I})'\mathbf{P}_\omega\mathbf{Y} = \mathbf{0}$$

pues $(\mathbf{I} - \mathbf{P})$ proyecta sobre el ortogonal de $\mathcal{V}_r$, mientras que $\mathbf{P}_\omega$ proyecta sobre $\mathcal{V}_1 \in \mathcal{V}_r$

$$\|\mathbf{Y}\|^2 = \|\mathbf{Y} - \widehat{\boldsymbol{\eta}}\|^2 + \|\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{\eta}}_\omega\|^2 + \|\widehat{\boldsymbol{\eta}}_\omega\|^2$$

Llamaremos

$\|\mathbf{Y}\|^2$: suma de cuadrados total

$\|\mathbf{Y} - \widehat{\boldsymbol{\eta}}\|^2$: suma de cuadrados residual

$\|\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{\eta}}_\omega\|^2$: suma de cuadrados de la regresión

$\|\mathbf{Y} - \widehat{\boldsymbol{\eta}}_\omega\|^2$: suma de cuadrados total corregida

Tenemos las siguientes igualdades

$$\begin{aligned}
\|\mathbf{Y}\|^2 &= \mathbf{Y}'\mathbf{Y} && \text{g.l}=n\\
\|\mathbf{Y} - \widehat{\boldsymbol{\eta}}\|^2 &= \mathbf{Y}'\mathbf{Y} - \widehat{\boldsymbol{\beta}}'\mathbf{X}\mathbf{Y} && \text{g.l}=n-p\\
\|\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{\eta}}_\omega\|^2 &= \widehat{\boldsymbol{\beta}}'\mathbf{X}\mathbf{Y} - n(\overline{\mathbf{Y}})^2 = \|\widehat{\boldsymbol{\eta}}\|^2 - n(\overline{\mathbf{Y}})^2 && \text{g.l}=p-1\\
\|\mathbf{Y} - \widehat{\boldsymbol{\eta}}_\omega\|^2 &= n(\overline{\mathbf{Y}})^2 && \text{g.l}=1
\end{aligned}$$

Si quisiéramos verificar la significación de la regresión, haríamos

$$F = \frac{\|\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{\eta}}_\omega\|^2 / p - 1}{\|\mathbf{Y} - \widehat{\boldsymbol{\eta}}\|^2 / n - p}$$

Muchos programas ofrecen en su salida una tabla como la que sigue

| Fuente | | g.l. | M.S. | F | p-valor |
|---|---|---|---|---|---|
| Regresión | $\|\widehat{\boldsymbol{\eta}}\|^2 - n(\overline{\mathbf{Y}})^2$ | $p-1$ | $(1) = \frac{\|\widehat{\boldsymbol{\eta}}\|^2 - n(\overline{\mathbf{Y}})^2}{p-1}$ | | |
| Residual | $\|\mathbf{Y} - \widehat{\boldsymbol{\eta}}\|^2$ | $n-p$ | $(2) = \frac{\|\mathbf{Y} - \widehat{\boldsymbol{\eta}}\|^2}{n-p}$ | $(1)/(2)$ | |
| Tot. Cor. | $\|\mathbf{Y}\|^2 - n(\overline{\mathbf{Y}})^2$ | $n-1$ | | | |

Cuadro 1: Tabla de ANOVA

## Datos de Biomasa

Producción de biomasa en el estuario de Cape Fear: los datos corresponden a un estudio de la Universidad de North Carolina en el que se muestrearon 3 tipos de vegetación en tres localidades. En cada una se muestreó al azar 5 lugares con un total de 45 observaciones. Analizaremos las variables del sustrato:

$x_1$=SAL: Salinidad

$x_2$=pH: Acidez

$x_3$= K: Potasio

$x_4$=Naa: Sodio

$x_5$=Zn: Zinc

$y$: Biomasa Aérea

En esta etapa nos concentraremos en identificar aquellas variables que muestran mayor relación con $y$. Ajustaremos el modelo

$$E(y) = \beta_0 + \beta_1 SAL + \beta_2 pH + \beta_3 K + \beta_4 Naa + \beta_5 Zn$$

```
SALIDA DE S-PLUS

DATOS DE BIOMASA

> sal.lm

Call:
lm(formula = BIO ~ ., data = bio)

Coefficients:
 (Intercept)          K          NAA       PH       SAL        ZN
    1252.589 -0.2853166 -0.008662343 305.4821 -30.28808 -20.67844

Degrees of freedom: 45 total; 39 residual
Residual standard error: 398.2671


> summary(sal.lm)

Call: lm(formula = BIO ~ ., data = bio)
Residuals:
    Min     1Q Median    3Q  Max
 -748.1 -223.7 -85.22 139.1 1072
```

```
Coefficients:
                 Value Std. Error   t value   Pr(>|t|)
(Intercept)  1252.5895  1234.7294    1.0145     0.3166
          K    -0.2853     0.3483   -0.8191     0.4177
        NAA    -0.0087     0.0159   -0.5438     0.5897
         PH   305.4821    87.8831    3.4760     0.0013
        SAL   -30.2881    24.0298   -1.2604     0.2150
         ZN   -20.6784    15.0544   -1.3736     0.1774

Residual standard error: 398.3 on 39 degrees of freedom
Multiple R-Squared: 0.6773
F-statistic: 16.37 on 5 and 39 degrees of freedom, the p-value is 1.082e-008


Correlation of Coefficients:
    (Intercept)       K     NAA      PH     SAL
  K -0.3122
NAA  0.3767     -0.8103
 PH -0.8406      0.1212 -0.2442
SAL -0.9180      0.3047 -0.4324  0.6045
 ZN -0.8809      0.1908 -0.3386  0.8350  0.7113
```

```
STATISTIX 4.0                                            BIOMASS, 04/24/96, 16:24

UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF BIO

PREDICTOR
VARIABLES     COEFFICIENT    STD ERROR     STUDENT'S T        P        VIF
---------     -----------    ---------     -----------     ------     -----
CONSTANT         1252.55      1234.71          1.01        0.3166
SAL             -30.2875      24.0296         -1.26        0.2150      2.2
PH               305.485      87.8818          3.48        0.0013      3.3
K               -0.28532      0.34832         -0.82        0.4177      3.0
NA              -0.00866      0.01593         -0.54        0.5897      3.3
ZN              -20.6778      15.0541         -1.37        0.1774      4.3

R-SQUARED              0.6773      RESID. MEAN SQUARE (MSE)  1.586E+05
ADJUSTED R-SQUARED     0.6360      STANDARD DEVIATION          398.267

SOURCE         DF       SS            MS          F        P
----------     ---   ----------   ----------    -----    ------
REGRESSION      5    1.298E+07    2.597E+06     16.37    0.0000
RESIDUAL       39    6.186E+06    1.586E+05
TOTAL          44    1.917E+07

CASES INCLUDED 45    MISSING CASES 0
```

## Hipótesis Anidadas

La interpretación del test de F en términos de las hipótesis anidadas.

Supongamos que tenemos $H_1, H_2, \ldots, H_k$ un conjunto de hipótesis que imponen $q_1, q_2, \ldots, q_k$ restricciones independientes, respectivamente. Luego, las $q_1+q_2+\ldots+q_k$ funciones estimables son l.i. La secuencia de hipótesis anidadas estará dada por

$$\Omega, \omega_1 = \Omega \cap H_1, \omega_2 = \Omega \cap H_1 \cap H_2, \ldots, \omega_k = \Omega \cap H_1 \cap H_2 \ldots \cap H_k$$

Si lamamos $\mathcal{V}^{(j)}$ a los espacios asociados cada uno de dimensión $r - q_1 - q_2 - \ldots - q_j$

$$\mathcal{V}^{(r)} \supset \mathcal{V}^{(r-q_1)} \supset \ldots \mathcal{V}^{(r-q_1-q_2-\ldots-q_k)}$$

Sea $\widehat{\boldsymbol{\eta}}_{\omega_j}$ la proyección ortogonal de $Y$ sobre $\mathcal{V}^{(j)}$, por lo tanto tenemos que

$$Y = Y - \widehat{\boldsymbol{\eta}} + \widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{\eta}}_{\omega_1} + \widehat{\boldsymbol{\eta}}_{\omega_1} - \widehat{\boldsymbol{\eta}}_{\omega_2} + \ldots + \widehat{\boldsymbol{\eta}}_{\omega_{k-1}} - \widehat{\boldsymbol{\eta}}_{\omega_k} + \widehat{\boldsymbol{\eta}}_{\omega_k}$$

y en consecuencia

$$\|Y\|^2 = \|Y - \widehat{\boldsymbol{\eta}}\|^2 + \|\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{\eta}}_{\omega_1}\|^2 + \|\widehat{\boldsymbol{\eta}}_{\omega_1} - \widehat{\boldsymbol{\eta}}_{\omega_2}\|^2 + \ldots + \|\widehat{\boldsymbol{\eta}}_{\omega_{k-1}} - \widehat{\boldsymbol{\eta}}_{\omega_k}\|^2 + \|\widehat{\boldsymbol{\eta}}_{\omega_k}\|^2$$

donde cada suma tiene una distribución $\chi^2$ no central bajo $\Omega$ con $n-r$, $q_1$, $q_2, \ldots, q_k$, $r - q_1 - q_2 - \ldots - q_k$ grados de libertad.