Universidad Autónoma de Madrid
Estadística Ciencias Ambientales
Facultad de Ciencias
Departamento de Matemáticas
Profesor: Santiago de la Fuente Fernández

REGRESIÓN LINEAL MÚLTIPLE

REGRESIÓN LINEAL MÚLTIPLE

Las técnicas de regresión lineal múltiple parten de k+1 variables cuantitativas, siendo Y la variable de respuesta y (x_1, ... , x_k) las variables explicativas.

Se trata de extender a las 'k' variables las técnicas de la regresión lineal simple. En esta línea, la variable Y se puede expresar mediante una función lineal de las variables (x_1, ... , x_k)

$$\boxed{\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \cdots + \beta_\mathbf{k} \mathbf{x}_\mathbf{k}}$$

Para ello, dispondremos de un modelo de probabilidad (la Normal).

El experimentador fija los valores de las X_{ki} y obtiene 'al azar' los correspondientes Y_i

$$\boxed{\text{Modelo: } \mathbf{Y} = \beta_0 + \beta_1 \mathbf{X}_1 + \beta_2 \mathbf{X}_2 + \cdots + \beta_\mathbf{k} \mathbf{X}_\mathbf{k} + \mathbf{U}}$$

REGRESIÓN LINEAL MÚLTIPLE - MUESTRA ALEATORIA

Modelo: $\mathbf{Y} = \beta_0 + \beta_1 \mathbf{X}_1 + \beta_2 \mathbf{X}_2 + \cdots + \beta_{\mathbf{k}} \mathbf{X}_{\mathbf{k}} + \mathbf{U}$

Sea la muestra aleatoria:

$$\mathbf{Y}_{\mathbf{i}} = \beta_0 + \beta_1 \mathbf{X}_{1\mathbf{i}} + \beta_2 \mathbf{X}_{2\mathbf{i}} + \cdots + \beta_{\mathbf{k}} \mathbf{X}_{\mathbf{ki}} + \mathbf{u}_{\mathbf{i}} \quad (\mathbf{i} = 1, 2, \cdots, \mathbf{n})$$

$\mathrm{Y_i} \; \varepsilon \; \mathrm{N}(\beta_0 + \beta_1 \mathrm{X}_1 + \cdots + \beta_{\mathrm{k}} \mathrm{X}_{\mathrm{k}}, \sigma^2)$ independientes, (i=1, ..., n)

$\mathbf{u}_{\mathrm{i}} \; \varepsilon \; \mathrm{N}(0, \sigma^2)$ independientes, (i=1, ..., n)

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_{\mathbf{n}} \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{X}_{11} & \cdots & \mathbf{X}_{\mathbf{k}1} \\ 1 & \mathbf{X}_{12} & \cdots & \mathbf{X}_{\mathbf{k}2} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & \mathbf{X}_{1\mathbf{n}} & \cdots & \mathbf{X}_{\mathbf{kn}} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{\mathbf{k}} \end{pmatrix} + \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_{\mathbf{n}} \end{pmatrix}$$

$\mathbf{Y} = \mathbf{X}\beta + \mathbf{U}$ X = matriz del diseño

Hipótesis comunes entre regresiones lineal y múltiple:

- Normalidad
- Homocedasticidad
- Linealidad
- Independencia

Requesitos adicionales de la regresión múltiple:

- $n > k+1$

El modelo depende de $(k+2)$ parámetros. Para que la regresión tenga significado debe haber un número suficiente de datos.

- **Ninguna X es combinación lineal de las otras (colinealidad)**

Si alguna de las X_i es combinación lineal exactade algunas de las otras X_i, el modelo puede simplificarse con menos variables explicativas. También hay que considerar si alguna de las X_i está fuertemente correlacionada con otras.

DATOS - ESTIMACIÓN DE LOS PARÁMETROS

$$\mathbf{Y_i} = \beta_0 + \beta_1 \mathbf{X_{1i}} + \beta_2 \mathbf{X_{2i}} + \cdots + \beta_k \mathbf{X_{ki}} + \mathbf{u_i}$$

$\boxed{\mathbf{Y} = \mathbf{X}\beta + \mathbf{U}}$ donde $\mathbf{X}$ = matriz del diseño

Datos	Y	X_1	$\cdots$	X_k
1	y_1	x_{11}	$\cdots$	x_{k1}
2	y_2	x_{12}	$\cdots$	x_{k2}
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
n	y_n	x_{1n}	$\cdots$	x_{kn}

La nube de puntos está en un espacio de dimensión (k+1).

Difícil de visualizar para k>2

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

X' = matriz transpuesta

$$\mathbf{e_i} = \mathbf{y_i} - \hat{\mathbf{y}}_i = \mathbf{y_i} - \left(\hat{\beta}_0 + \hat{\beta}_1 \mathbf{x_{1i}} + \hat{\beta}_2 \mathbf{x_{2i}} + \cdots + \hat{\beta}_k \mathbf{x_{ki}}\right)$$

$$\hat{\sigma}_r^2 = S_R^2 = \frac{1}{n-k-1}\sum_{i=1}^{n} e_i^2$$ varianza residual

NOTA.- Al igual que en regresión simple, necesitamos estimar la varianza σ^2 del error aleatorio U.

Un estimador razonable, en un principio, es la varianza de los errores de predicción (también denominados como residuos del modelo).

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2$$

Sin embargo, este estimador es sesgado para σ^2, lo que significa que $E\left(\hat{\sigma}^2\right) \neq \sigma^2$

El sesgo se define como la diferencia entre la media del estimador y el verdadero valor del parámetro que se quiere estimar.

Utilizamos, por tanto, la varianza residual para estimar σ^2, que sí es un estimador insesgado de σ^2, es decir, centrado en torno a σ^2

$$\hat{\sigma}_r^2 = S_R^2 = \frac{1}{n-k-1}\sum_{i=1}^{n} e_i^2$$

RELACIONES ENTRE LAS VARIABLES

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

$$\hat{Y} = X\hat{\beta} = \boxed{X\left(X'X\right)^{-1} X'}\,Y = H\,Y$$

H matriz de la proyección

La matriz H cumple las propiedades
- Idempotente: H.H = H
- Simétrica: H' = H
- Rango: rango (X) = rango H = k+1

El error de predicción se puede escribir en forma matricial:

$$e = Y - \hat{Y} = Y - HY = (I - H)Y$$

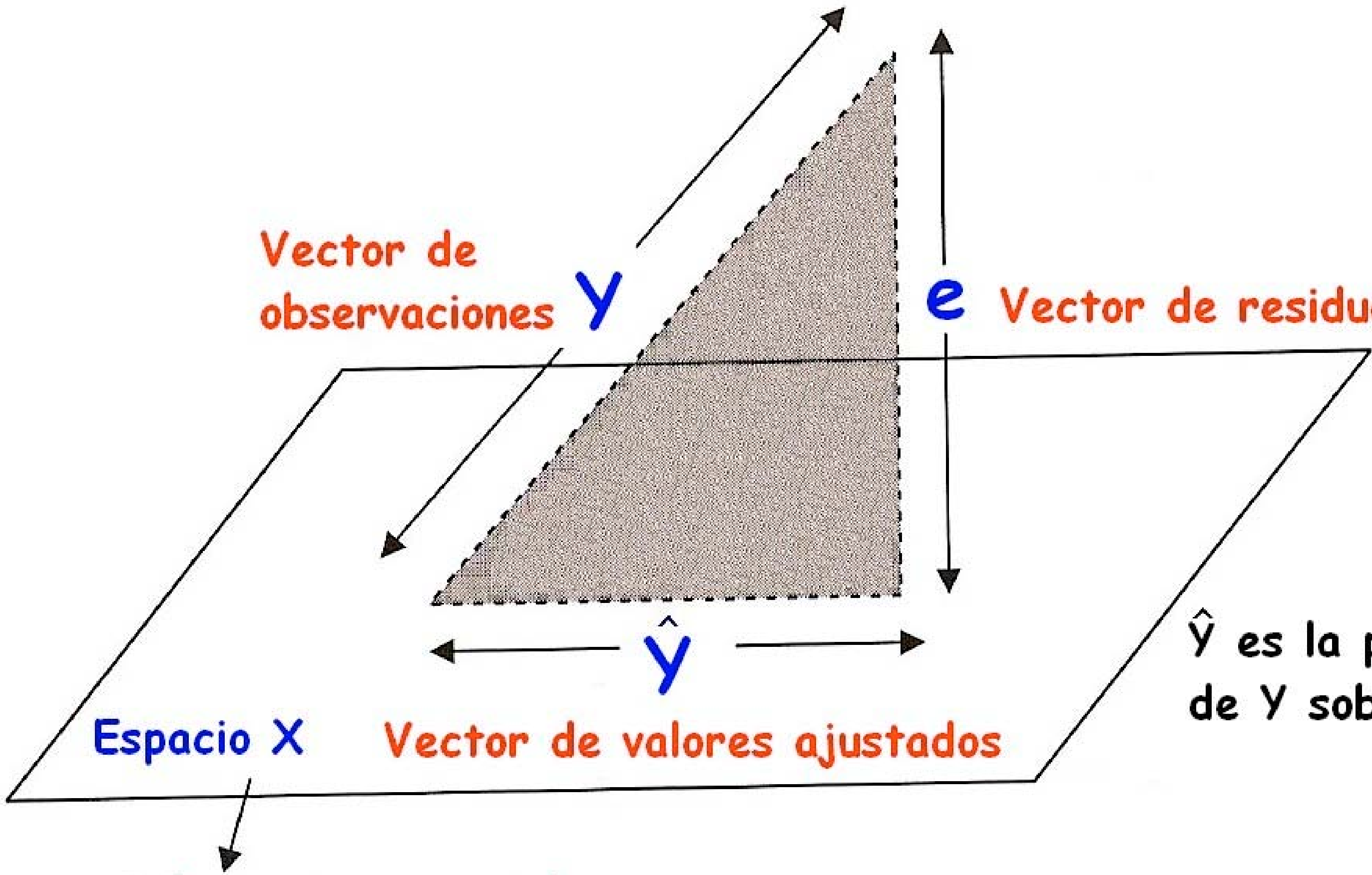
La expresión $\hat{Y} = H\,Y$ indica que la matriz H (idempotente) transforma el vector de las observaciones Y en el vector de valores ajustados (predicción) $\hat{Y}$, siendo la regresión una proyección.

El modelo regresión $\hat{Y} = H\,Y$ proyecta el valor de observaciones sobre el subespacio vectorial de las columnas de la matriz X (subespacio de las variables independientes).

El vector de residuos es perpendicular a cada columna de X y al vector de predicción $\hat{Y}$

INTERPRETACIÓN GEOMÉTRICA

El método de los mínimos cuadrados, en el espacio formado por las variables, es equivalente a encontrar un vector en dicho espacio que se encuentre lo más cercano posible al vector de las observaciones.



$\hat{Y}$ es la proyección ortogonal de Y sobre dicho espacio

Subespacio vectorial
generado por la columna X (vectores columna variables explicativas)

Demostración $\hat{\beta} = \left(X'X\right)^{-1} X'Y$

$$Y = X\beta + U$$

El correspondiente modelo ajustado será: $\hat{Y} = X\hat{\beta}$

$$\hat{U} = Y - \hat{Y} = Y - X\hat{\beta}$$

Denominando S a la suma de los cuadrados de los residuos:

$$S = \hat{U}'\hat{U} = \left[\hat{u}_{1,}\hat{u}_2, \ \cdots \ , \hat{u}_n\right] \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix} = \sum_{i=1}^{n} \hat{u}_i$$

(A' = matriz transpuesta A)

Un escalar es igual a su transpuesto:

$\hat{\beta}'X'Y = Y'X\hat{\beta}$

$$S = \left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right) =$$

$$= Y'Y - \hat{\beta}'X'Y - Y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta} =$$

$$= Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

$$\mathbf{S} = \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)^{'}\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) = \mathbf{Y}'\mathbf{Y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$$

Para minimizar S se aplica el criterio mínimo-cuadrático, derivando respecto $\hat{\beta}$

$$\frac{\vartheta \mathbf{S}}{\vartheta \hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = 0 \Rightarrow \underline{\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}}$$

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

$$\Rightarrow \underline{\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{'}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}}$$

DISTRIBUCIÓN de $\hat{\beta}$

Hemos visto que

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

$$\hat{Y} = X\hat{\beta} = \underbrace{\boxed{X\left(X'X\right)^{-1} X'}}_{\text{matriz de proyección}} Y = HY$$

- El vector de observaciones Y se distribuye según una normal multivariante de media $X\beta$ y de matriz de varianzas covarianzas $\sigma^2 I$

$$Y \in N(X\beta, \sigma^2 I)$$

- $\hat{\beta}$ es combinación lineal de las componentes del vector Y, por lo que se distribuye según una variable aleatoria normal, donde su media y matriz de varianzas y covarianzas será:

$$E\left(\hat{\beta}\right) = E\left[(X'X)^{-1} X'Y\right] = (X'X)^{-1} X'E(Y) =$$

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

$\hat{\beta}$ es un estimador centrado de β

$$\mathrm{Var}(\hat{\beta}) = \mathrm{Var}\left[(X'X)^{-1} X'Y\right] = (X'X)^{-1} X' \,\mathrm{Var}(Y)\, X(X'X)^{-1} =$$

$$= (X'X)^{-1} X' \sigma^2 X(X'X)^{-1} = \sigma^2 (X'X)^{-1}$$

$$\hat{\beta} \,\varepsilon\, N\left(\beta, \sigma^2 (X'X)^{-1}\right)$$

- Con el ajuste de mínimos cuadrados:

$$e_i = y_i - \hat{y}_i = y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_k x_{ki}\right)$$

parámetros estimados del modelo

$$\hat{\beta}_i \,\varepsilon\, N\left(\beta_i, \sigma^2 q_{ii}\right)$$

q_{ii} es el elemento i-ésimo de la diagonal principal $(X'X)^{-1}$

La estimación de σ^2 se hacía mediante la varianza residual $\hat{S}_R^2 = \dfrac{1}{n-k-1}\sum_{i=1}^{n} e_i^2$

De forma que estimaremos la varianza de $\hat{\beta}_i \,\varepsilon\, N\left(\beta_i, \sigma^2 q_{ii}\right)$ mediante $\hat{S}_R^2 q_{ii}$

El error estándar de $\hat{\beta}_i$ viene dado por: $SE(\hat{\beta}_i) = \sqrt{\hat{S}_R^2 q_{ii}} = \hat{S}_R \sqrt{q_{ii}}$

Se demuestra que $\dfrac{(n-k-1)\hat{S}_R^2}{\sigma^2} \approx \chi^2_{n-k-1}$

Contraste t - Student

Los parámetros del modelo estimado $\hat{\beta}_i \,\varepsilon\, N\left(\beta_i,\ \sigma^2 q_{ii}\right)$ siendo q_{ii} el elemento i-ésimo de la diagonal principal $(X'X)^{-1}$

De forma estándar, se obtiene $\dfrac{\hat{\beta}_i - \beta_i}{\sigma\sqrt{q_{ii}}} \approx N(0,1)$

Una variable t de Student con k-grados de libertad, se define $t_k = \dfrac{N(0,1)}{\sqrt{\dfrac{1}{k}\chi_k^2}}$

Considerando que $\dfrac{(n-k-1)\hat{S}_R^2}{\sigma^2} \approx \chi_{n-k-1}^2$ resulta:

$$t = \frac{\dfrac{\hat{\beta}_i - \beta_i}{\sigma\sqrt{q_{ii}}}}{\sqrt{\dfrac{1}{n-k-1}\dfrac{(n-k-1)\hat{S}_R^2}{\sigma^2}}} = \frac{\hat{\beta}_i - \beta_i}{\hat{S}_R\sqrt{q_{ii}}} \approx t_{n-k-1}$$

Planteando la hipótesis nula H_o: $\beta_i = 0$ frente a la hipótesis alternativa H_a: $\beta_i \neq 0$ es decir, si el valor del parámetro en la población es cero o no. En otras palabras, si la variable X_i influye sobre la variable respuesta Y.

Siendo $\mathbf{t} = \dfrac{\hat{\beta}_i - \beta_i}{\hat{S}_R \sqrt{\mathbf{q}_{ii}}} \approx \mathbf{t}_{n-k-1}$ Al establecer la hipótesis nula H_0: $\beta_i = 0$ es decir, el valor del parámetro en la población es cero. Esto es, la variable X_i no influiría sobre la variable respuesta Y.

Bajo la hipótesis nula, siendo $\beta_i = 0$, resulta: $\mathbf{t} = \dfrac{\hat{\beta}_i}{\hat{S}_R \sqrt{\mathbf{q}_{ii}}} \approx \mathbf{t}_{n-k-1}$

- En definitiva, siendo la hipótesis nula H_0: $\beta_i = 0$

Aceptamos la hipótesis nula cuando el valor del estadístico t proviene de una t_{n-k-1}

Para $n > 30$ se acepta la hipótesis nula H_0 cuando $|t| \leq 2$

En caso contrario, se acepta la hipótesis alternativa H_a y se concluye que la variable i-ésima influye en la respuesta.

INTERVALOS DE CONFIANZA para β_i

Estimación de los parámetros: $\hat{\beta} = (X'X)^{-1} X'Y$ X' = matriz transpuesta

$$IC_{1-\alpha}(\beta_i) = \left(\hat{\beta}_i \pm t_{(n-k-1);\alpha/2} \boxed{S_r \sqrt{q_{i+1,i+1}}} \right)$$

Desviación típica (error típico) estimada de $\hat{\beta}_i$

$q_{i+1,i+1} \equiv$ elementos de la diagonal principal $(X'X)^{-1}$

$$S_r^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-k-1}$$

INTERVALOS DE CONFIANZA para β_i

Estimación de los parámetros: $\hat{\beta} = (X'X)^{-1} X'Y$ X' = matriz transpuesta

$$IC_{1-\alpha}(\boldsymbol{\beta}_i) = \left(\hat{\beta}_i \pm t_{(n-k-1);\alpha/2} \boxed{S_r \sqrt{q_{i+1,i+1}}} \right)$$

$q_{i+1,\,i+1} \equiv$ elementos de la diagonal principal $(X'X)^{-1}$ Desviación típica estándar (error típico)

CONTRASTES DE HIPÓTESIS

Hipótesis nula $H_0 : \boldsymbol{\beta}_i = 0$ X_i no influye en Y

Hipótesis alternativa $H_1 : \boldsymbol{\beta}_i \neq 0$ X_i influye en Y

Se acepta la hipótesis nula H_0, con un nivel de confianza $(1 - \alpha)$ cuando el 0 se encuentra en el intervalo de confianza.

En caso contrario, cuando el 0 no cae en el intervalo de confianza, se acepta la hipótesis alternativa H_1, y en consecuencia, X_i influye en Y

Este contraste es equivalente al contraste de la t-Student para cada β_i

DESCOMPOSICIÓN DE LA VARIABILIDAD

$$\mathrm{SCT} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}\left[(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})^2\right] =$$

$$= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \overbrace{2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y})}^{0}$$

$$\underbrace{\sum_{i=1}^{n}(y_i - \bar{y})^2}_{\mathrm{SCT}} = \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}_{\mathrm{SCE}} + \underbrace{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}_{\mathrm{SCR}}$$

Suma cuadrados total — Suma cuadrados explicada — Suma cuadrados residual

Coeficiente de determinación $R^2 = \dfrac{\mathrm{SCE}}{\mathrm{SCT}} = \dfrac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$

ANÁLISIS de la VARIANZA - TABLA ANOVA

$$\underbrace{\sum_{i=1}^{n}(y_i-\bar{y})^2}_{SCT} = \underbrace{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}_{SCE} + \underbrace{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}_{SCR}$$

(n-1) grados libertad   k grados libertad   (n-k-1) grados libertad

Variación	Suma cuadrados	g.l	Media cuadrática	F-Snedecor
Explicada	$SCE=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$	k	SCE/k	$F=\dfrac{SCE/k}{SCR/(n-k-1)}$
Residual	$SCR=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$	n-k-1	$SCR/(n-k-1)$	
Total	$SCT=\sum_{i=1}^{n}(y_i-\bar{y})^2$	n-1		

Se establecen las hipótesis:

$H_0: \beta_1=\beta_2=\cdots=\beta_k=0$ el modelo no es explicativo

$H_1:$ al menos un $\beta_i \neq 0$ el modelo es explicativo

A un nivel de confianza (1-α) se rechaza H_0 si $F \geq F_{k,(n-k-1);\alpha}$

COEFICIENTES de DETERMINACIÓN - F de Snedecor

$$R^2 = \frac{SCE}{SCT} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$F = \frac{SCE/k}{SCR/(n-k-1)} = \frac{SCE}{SCT}\,\frac{SCT}{SCR}\,\frac{n-k-1}{k} = R^2\,\frac{1}{\dfrac{SCR}{SCT}}\,\frac{n-k-1}{k} =$$

$$= R^2\,\frac{1}{\dfrac{SCT-SCE}{SCT}}\,\frac{n-k-1}{k} = R^2\,\frac{1}{1-R^2}\,\frac{n-k-1}{k} = \frac{R^2}{1-R^2}\,\frac{n-k-1}{k}$$

$$\boxed{F = \frac{R^2}{1-R^2}\,\frac{n-k-1}{k}}$$

RESUMEN de los CONTRASTES

Contraste Conjunto F-Snedecor	Contrastes Individuales t-Sudent	Conclusión
Modelo explicativo	Todas las X_i son explicativas	Tomamos todas las X_i
Modelo explicativo	Algunas X_i son explicativas	Nos quedamos con las X_i explicativas
Modelo explicativo	Ninguna X_i es explicativa	Posible Multicolinealidad revisar el modelo
Modelo no explicativo	Todas las X_i son explicativas	Posible Multicolinealidad revisar el modelo
Modelo no explicativo	Algunas X_i son explicativas	Posible Multicolinealidad revisar el modelo
Modelo no explicativo	Ninguna X_i es explicativa	El modelo no explica Y

- Estimación de la respuesta media de Y para los valores $(x_{10}, x_{20}, \ldots, x_{k0})$ de las variables explicativas.

$$\hat{\mathbf{y}}_0 = \hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_1 \mathbf{x}_{10} + \hat{\boldsymbol{\beta}}_2 \mathbf{x}_{20} + \cdots + \hat{\boldsymbol{\beta}}_{\mathbf{k}} \mathbf{x}_{\mathbf{k}0}$$

$$\mathbf{IC}_{1-\alpha} = \left(\hat{\mathbf{y}}_0 \pm \mathbf{t}_{(\mathbf{n}-\mathbf{k}-1);\, \alpha/2}\, \mathbf{S}_{\mathbf{r}} \sqrt{\begin{pmatrix} 1 & x_{10} & x_{20} & \cdots & x_{k0} \end{pmatrix} (X'X)^{-1} \begin{pmatrix} 1 \\ x_{10} \\ x_{20} \\ \vdots \\ x_{k0} \end{pmatrix}} \right)$$

- Predicción de un nuevo valor de Y para los valores $(x_{10}, x_{20}, \ldots, x_{k0})$ de las variables explicativas.

$$\mathbf{IP}_{1-\alpha} = \left(\hat{\mathbf{y}}_0 \pm \mathbf{t}_{(\mathbf{n}-\mathbf{k}-1);\, \alpha/2}\, \mathbf{S}_{\mathbf{r}} \sqrt{1 + \begin{pmatrix} 1 & x_{10} & x_{20} & \cdots & x_{k0} \end{pmatrix} (X'X)^{-1} \begin{pmatrix} 1 \\ x_{10} \\ x_{20} \\ \vdots \\ x_{k0} \end{pmatrix}} \right)$$