

Tema 4. Problemas de inferencia estadística en el modelo de regresión lineal múltiple

1. Intervalos de confianza y contrastes para los coeficientes de regresión.

- 1.1. Intervalos de confianza para los coeficientes de regresión.
- 1.2. Contrastes de hipótesis para los coeficientes de regresión.

2. Intervalos de confianza para la varianza.

3. Contraste conjunto sobre los coeficientes.

- 3.1. Descomposición básica y contraste de regresión.
- 3.2. Interpretación de los contrastes.

4. Correlación en regresión múltiple.

- 4.1. El coeficiente de determinación.
- 4.2. El coeficiente de determinación corregido.

1. Intervalos de confianza y contrastes para los coeficientes de regresión.

$$\text{Modelo: } Y = X\beta + U, \quad U \sim NM_n(0, \sigma^2 I)$$

$$\text{Hiperplano de regresión: } \hat{Y} = X\hat{\beta}, \quad \hat{\beta} = (X'X)^{-1}X'Y$$

$$\hat{Y} = \underbrace{X(X'X)^{-1}X'}_H Y = HY$$

1.1. Intervalos de confianza para los coeficientes de regresión.

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 q_{jj}) \Rightarrow \frac{\hat{\beta}_j - \beta_j}{\sigma \sqrt{q_{jj}}} \sim N(0,1)$$

donde q_{jj} es el elemento $(j+1, j+1)$ de la matriz $(X'X)^{-1}$. Puesto que, en general, σ^2 será desconocida, utilizaremos s_R^2 como estimador de la varianza, y entonces:

$$\frac{\hat{\beta}_j - \beta_j}{s_R \sqrt{q_{jj}}} \sim t_{n-k-1}$$

El intervalo de confianza para el parámetro β_j al nivel $(1-\alpha)100\%$ es:

$$IC_{1-\alpha}(\beta_j) = \hat{\beta}_j \mp t_{1-\alpha/2, s_R} \sqrt{q_{jj}}$$

donde $t_{1-\alpha/2}$ es el percentil $(1-\alpha/2)100\%$ de la ley t de Student con $n-k-1$ grados de libertad y q_{jj} es el elemento $(j+1, j+1)$ de la matriz $(X'X)^{-1}$.

Ejemplo (Ejercicio 17): Hallar un I.C. para los coeficientes de regresión β_1 y β_2 del **Ejercicio 14** (consumo de 8 provincias españolas), con un nivel de confianza del 95%.

Debemos calcular (para $j=1,2$):

$$IC_{1-\alpha}(\beta_j) = \hat{\beta}_j \mp t_{1-\alpha/2, s_R} \sqrt{q_{jj}}$$

Modelo estimado: $y = 9.05 + 0.5203x_1 + 0.2397x_2$,

donde $x_1 = n^\circ$ de automóviles por 1000 habitantes, $x_2 = n^\circ$ de teléfonos por 1000 habitantes, $y =$ consumo.

La varianza residual es $s_R^2 = 33.83$ ($\Rightarrow s_R = 5.82$); $n-k-1=5$, $(1-\alpha)=0.95$, $t_{1-\alpha/2} = t_{0.975} = 2.57$.

$$X'X = \begin{pmatrix} 8 & 672 & 1176 \\ 672 & 57822 & 100453 \\ 1176 & 100453 & 176758 \end{pmatrix} \quad (X'X)^{-1} = \begin{pmatrix} 6.35 & -0.03 & -0.02 \\ -0.03 & 1.52 \cdot 10^{-3} & -6.5 \cdot 10^{-4} \\ -0.02 & 6.5 \cdot 10^{-4} & 5.38 \cdot 10^{-4} \end{pmatrix} \Rightarrow \begin{cases} q_{11} = 1.52 \cdot 10^{-3} \\ q_{22} = 5.38 \cdot 10^{-4} \end{cases}$$

Sustituyendo se obtienen los siguientes intervalos para β_1 y β_2 :

$$I.C.(\beta_1) = \left[0.5203 \mp 2.57(5.82)\sqrt{0.0015} \right] = [-0.0590, 1.0996]$$

$$I.C.(\beta_2) = \left[0.2397 \mp 2.57(5.82)\sqrt{0.0005} \right] = [-0.0948, 0.5742]$$

1.2. Contrastes de hipótesis para los coeficientes de regresión (contrastos de hipótesis para un solo coeficiente)

Hipótesis de los contrastes (o de los tests):

$$(a) \begin{cases} H_0: \beta_j = \beta_j^* \\ H_1: \beta_j > \beta_j^* \end{cases} \quad (b) \begin{cases} H_0: \beta_j = \beta_j^* \\ H_1: \beta_j < \beta_j^* \end{cases} \quad (c) \begin{cases} H_0: \beta_j = \beta_j^* \\ H_1: \beta_j \neq \beta_j^* \end{cases}$$

Estadístico de contraste:

$$t_{\text{exp}} = \frac{\hat{\beta}_j - \beta_j^*}{s_R \sqrt{q_{jj}}} \sim t_{n-k-1}$$

donde q_{jj} es el elemento $(j+1, j+1)$ de la matriz $(X'X)^{-1}$.

Para un nivel de significación α fijado, las regiones de rechazo son:

$$(a) \{ t_{\text{exp}} > t_{1-\alpha} \} = (t_{1-\alpha}, +\infty) \quad (b) \{ t_{\text{exp}} < t_{\alpha} \} = (-\infty, t_{\alpha}) = (-\infty, -t_{1-\alpha})$$

$$(c) \{ |t_{\text{exp}}| > t_{1-\alpha/2} \} = (-\infty, -t_{1-\alpha/2}) \cup (t_{1-\alpha/2}, +\infty)$$

donde $t_{1-\alpha}$, t_{α} , $t_{1-\alpha/2}$ son, respectivamente, los percentiles $(1-\alpha)100\%$, $\alpha 100\%$ y $(1-\alpha/2)100\%$ de la ley t de Student con $n-k-1$ grados de libertad.

Contrastes de significación individuales

Hipótesis del contraste:

$$\begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0 \end{cases}$$

Estadístico de contraste:

$$t_{\text{exp}} = \frac{\hat{\beta}_j - 0}{s_R \sqrt{q_{jj}}} \sim t_{n-k-1}$$

donde q_{jj} es el elemento $(j+1, j+1)$ de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$.

Para un nivel de significación α fijado, la región de rechazo es:

$$\left\{ |t_{\text{exp}}| > t_{1-\alpha/2} \right\} = (-\infty, -t_{1-\alpha/2}) \cup (t_{1-\alpha/2}, +\infty)$$

donde $t_{1-\alpha/2}$ es el percentil $(1-\alpha/2)100\%$ de la ley t de Student con $n-k-1$ grados de libertad.

Regla práctica para saber si los coeficientes son significativos

- 1) Calcular el correspondiente estadístico t_{exp} .
- 2) Si $|t_{\text{exp}}| > 2.5$, el coeficiente es significativo ($\neq 0$)

Justificación:

Si $n-k-1 \geq 6$ y $\alpha=0.05$, el percentil $(1-\alpha/2)100\%$ de la ley t de Student con $n-k-1$ grados de libertad toma valores en el intervalo $[1.96, 2.45]$, es decir:

$$t_{1-\alpha/2} \in [1.96, 2.45]$$

Reproducimos parte de la tabla de la distribución t de Student. Observad como **decrece el percentil** 97.5% a medida que **aumentan los grados de libertad**:

g.l.	$t_{0.975}$
6	2.45
7	2.36
8	2.31
9	2.26
10	2.23
...	...
120	1.98
∞	1.96

Ejemplo (Ejercicio 18): Realizar contrastes de significación para los coeficientes de regresión β_1 y β_2 del **Ejercicio 14** (consumo de 8 provincias españolas), con un nivel de significación del 5%.

$$\begin{cases} H_0 : \beta_1 = 0 \Rightarrow \beta_1 \text{ no es significativo individualmente} \\ H_1 : \beta_1 \neq 0 \end{cases}$$

$$t_{\text{exp}} = \frac{\hat{\beta}_1 - 0}{s_R \sqrt{q_{11}}} = \frac{0.5203}{5.82 \sqrt{0.0015}} = 2.31 \begin{cases} < 2.57 = t_{0.975} \Rightarrow \text{no rechazar } H_0 (\alpha = 0.05) \\ > 2.02 = t_{0.95} \Rightarrow \text{rechazar } H_0 (\alpha = 0.10) \end{cases}$$

$$\begin{cases} H_0 : \beta_2 = 0 \Rightarrow \beta_2 \text{ no es significativo individualmente} \\ H_1 : \beta_2 \neq 0 \end{cases}$$

$$t_{\text{exp}} = \frac{\hat{\beta}_2 - 0}{s_R \sqrt{q_{22}}} = \frac{0.2397}{5.82 \sqrt{0.0005}} = 1.84 < 2.57 = t_{0.975} \Rightarrow \begin{matrix} \text{no se puede rechazar} \\ H_0 \text{ para } \alpha = 0.05, 0.10 \end{matrix}$$

Ejercicio 18 (continuación):

Ambos coeficientes considerados **individualmente son no significativos**, es decir, que tanto x_1 (nº de automóviles cada 1000 habitantes) como x_2 (nº de teléfonos cada 1000 habitantes) no parecen explicar la variable dependiente y (índice de consumo global) mediante un modelo lineal.

Atención!

- Tal vez **sí podrían explicar conjuntamente** la variable y .
- Disponemos de muy pocas observaciones ($n=8$ datos).
La situación podría cambiar si aumentáramos el tamaño de la muestra.
- El verdadero modelo podría ser no lineal.

2. Intervalos de confianza para la varianza

Utilizando que:

$$\frac{n-k-1}{\sigma^2} s_R^2 \sim \chi_{n-k-1}^2$$

se construye el intervalo de confianza para σ^2 al nivel $(1-\alpha)100\%$:

$$IC_{1-\alpha}(\sigma^2) = \left(\frac{n-k-1}{x_{1-\alpha/2}} s_R^2, \frac{n-k-1}{x_{\alpha/2}} s_R^2 \right)$$

donde $x_{1-\alpha/2}$ y $x_{\alpha/2}$ son, respectivamente, los percentiles $(1-\alpha/2)100\%$ y $(\alpha/2)100\%$ de la ley chi-cuadrado con $n-k-1$ grados de libertad.

Es muy frecuente construir I.C. **de una sola cola**, del tipo:

$$IC_{1-\alpha}(\sigma^2) = \left(0, \frac{n-k-1}{x_\alpha} s_R^2 \right)$$

donde x_α es el percentil $\alpha 100\%$ de la ley chi-cuadrado con $n-k-1$ grados de libertad.

Ejemplo (Ejercicio 19): Hallar un I.C. para la varianza σ^2 del **Ejercicio 14** (consumo de 8 provincias españolas), con un nivel de confianza del 95%.

La varianza residual es $s_R^2 = 33.83 \Rightarrow s_R = 5.82$,
 $n-k-1=5$, $1-\alpha=0.95$, $x_{1-\alpha/2} = x_{0.975} = 12.8$, $x_{\alpha/2} = x_{0.025} = 0.831$.

El I.C. para la varianza es:

$$IC_{1-\alpha}(\sigma^2) = \left[\frac{5}{12.8} 33.83, \frac{5}{0.831} 33.83 \right] = [13.21, 203.55]$$

El I.C. para la desviación típica es:

$$IC_{1-\alpha}(\sigma) = [3.63, 14.27]$$

El I.C. a una sola cola para la varianza, teniendo en cuenta que $x_\alpha = x_{0.05} = 1.15$, es:

$$IC_{1-\alpha}(\sigma^2) = \left[0, \frac{5}{1.15} 33.83 \right] = [0, 147.09]$$

3. Contraste conjunto sobre los coeficientes (o contraste de regresión, ANOVA)

3.1 Descomposición básica y contraste de regresión

Consideremos el modelo: $y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ki} + u_i$, $1 \leq i \leq n$.

Queremos contrastar las hipótesis

$$\begin{cases} H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0 \\ H_1 : \exists j \text{ tal que } \beta_j \neq 0 \end{cases}$$

Interpretación:

Si H_0 fuera cierta, el modelo quedaría $y_i = \beta_0 + u_i$, lo que implicaría que fueran cuales fuesen los valores de las x_j , el valor de la variable respuesta y sería un valor **constante** β_0 . Esto significa que, tomadas **de forma conjunta**, las variables independientes **no tienen ninguna influencia** (lineal) **sobre la variable respuesta** y el intentar explicar el comportamiento de la variable respuesta a través de un modelo lineal sería un fracaso.

Este contraste se basa en las mismas ideas ya vistas en el Tema 2 sobre el Análisis de la Varianza (ADEVA o ANOVA)

i) Descomposición de la variabilidad:

$$\underbrace{\sum_{i=1}^n (y_i - \bar{y})^2}_{VT} = \underbrace{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}_{VE} + \underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)^2}_{VNE}$$

donde para $i=1, \dots, n$,

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_k x_{ki}$ son los valores ajustados o predichos,

y_i son los valores observados,

$e_i = y_i - \hat{y}_i$ son los residuos.

Matricialmente:

$$\begin{aligned} VT &= \mathbf{Y}'\mathbf{Y} - n\bar{y}^2 \\ VE &= \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} - n\bar{y}^2 \\ VNE &= \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} \end{aligned}$$

donde

$$\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)', \quad \mathbf{Y} = (y_1, y_2, \dots, y_n)',$$

$\mathbf{X}$ es la matriz $n \times (k+1)$ de los regresores.

ii) Estadístico F

Se puede probar que si la hipótesis nula $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$ es cierta, entonces

$$\frac{VT}{\sigma^2} \sim \chi^2_{n-1} \quad \text{y} \quad \frac{VE}{\sigma^2} \sim \chi^2_k$$

Además, sabemos que

$$\frac{VNE}{\sigma^2} = \frac{n-k-1}{\sigma^2} s_R^2 \sim \chi^2_{n-k-1}$$

Y que VE y VNE son independientes. Por tanto,

$$F_{\text{exp}} = \frac{\frac{VE/\sigma^2}{k}}{\frac{VNE/\sigma^2}{n-k-1}} = \frac{VE}{k s_R^2} \sim F_{k, n-k-1}$$

Rechazamos la hipótesis nula de **no significación conjunta**, $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$, cuando $F_{\text{exp}} > F_{k, n-k-1}^\alpha$, donde $F_{k, n-k-1}^\alpha$ es el percentil $(1-\alpha)100\%$ de la ley F de Fisher con k y $n-k-1$ grados de libertad.

iii) Tabla ANOVA (o ADEVA)

Fuente de variación	Suma de Cuadrados (SC)	Grados de libertad (gl)	Cociente o varianza (cuadrados medios)	contraste
Modelo	$VE = \hat{\beta}'\mathbf{X}'\mathbf{Y} - n\bar{y}^2$	k	$s_e^2 = \frac{\hat{\beta}'\mathbf{X}'\mathbf{Y} - n\bar{y}^2}{k}$	$F = \frac{s_e^2}{s_R^2}$
Error	$VNE = \mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y}$	$n-k-1$	$s_R^2 = \frac{\mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y}}{n-k-1}$	
Total	$VT = \mathbf{Y}'\mathbf{Y} - n\bar{y}^2$	$n-1$	$s_y^2 = \frac{\mathbf{Y}'\mathbf{Y} - n\bar{y}^2}{n-1}$	

Fórmulas útiles en el cálculo de la tabla ANOVA (o ADEVA)

Sean $\mathbf{X}$ la matriz de regresores $n \times (k+1)$, $\mathbf{Y}$ el vector de respuestas $n \times 1$.

- Se calculan la matriz $\mathbf{X}'\mathbf{X}$ y el vector $\mathbf{X}'\mathbf{Y}$.
- Se estima el vector de coeficientes:

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

- Se obtiene el vector $\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta}$ de valores predichos.
- Se calculan las sumas de cuadrados:

$$VE = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \hat{\beta}'\mathbf{X}'\mathbf{Y} - n\bar{y}^2$$

$$VNE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y}$$

$$VT = \sum_{i=1}^n (y_i - \bar{y})^2 = \mathbf{Y}'\mathbf{Y} - n\bar{y}^2$$

3.2 Interpretación de los contrastes

Diremos que un contraste de hipótesis es **significativo** cuando se rechaza la hipótesis nula con un nivel de significación menor que α .

casos	contraste conjunto F de Fisher	contrastos individuales t de Student
1	significativo	todos significativos
2	significativo	algunos significativos
3	significativo	ninguno significativo
4	no significativo	todos significativos
5	no significativo	algunos significativos
6	no significativo	ninguno significativo

1. Todas las variables explicativas influyen en la respuesta.
2. Solamente algunas variables explicativas influyen en la respuesta, por lo que deberíamos **eliminar las no significativas**.
3. **Multicolinealidad**: Las **variables explicativas son muy dependientes entre sí** y, por tanto, aunque conjuntamente influyen, los estimadores de los efectos individuales tienen varianzas muy altas y los valores del estadístico $t_{\text{exp}} = \hat{\beta}_j / (s_R \sqrt{q_{jj}})$ son pequeños y no significativos.
4. y 5. **Multicolinealidad especial**: Ocurre cuando **dos variables influyen mucho sobre la respuesta, pero en sentido contrario**, de manera que su efecto conjunto puede ser no significativo, aunque sus efectos individuales sí lo sean. (Casos poco frecuentes).
6. Ninguna de las variables influye. Sólo podemos concluir que **los efectos de las variables explicativas no se detectan en la muestra**. Por ejemplo, si una variable se mantiene aproximadamente constante, sus efectos sobre la respuesta no serán detectables.

Ejemplo (Ejercicio 20): Con los datos del **Ejercicio 14** (consumo de 8 provincias españolas), realizar el contraste de regresión para un nivel de significación del 5%.

$$\begin{cases} H_0 : \beta_1 = \beta_2 = 0 \\ H_1 : \text{algún } \beta_j \neq 0 \text{ para } j = 1, 2 \end{cases}$$

Para un nivel de significación $\alpha=0.05$, la regla de decisión será:

$$\text{rechazar } H_0 \text{ si } F_{exp} > F_{k, n-k-1}^{\alpha}$$

Construimos la tabla ANOVA, teniendo en cuenta que:

$$\left. \begin{aligned} \bar{y} &= \frac{704}{8} = 88, \\ \hat{\beta}' \mathbf{X}' \mathbf{Y} &= 62970.84, \\ \mathbf{Y}' \mathbf{Y} &= 63140, \\ n &= 8, \end{aligned} \right\} \Rightarrow \begin{aligned} VE &= \hat{\beta}' \mathbf{X}' \mathbf{Y} - n\bar{y}^2 = 1018.84, \\ VNE &= \mathbf{Y}' \mathbf{Y} - \hat{\beta}' \mathbf{X}' \mathbf{Y} = 169.16, \\ VT &= \mathbf{Y}' \mathbf{Y} - n\bar{y}^2 = 1188. \end{aligned}$$

$$F_{k, n-k-1}^{\alpha} = F_{2, 5}^{0.05} = 5.79$$

Ejercicio 20 (continuación):

La tabla ANOVA es

Fuente	S.C.	g.l.	varianza	contraste
VE	1018.84	2	509.42	$F=15.06$
VNE	169.16	5	33.832	
VT	1188	7	169.71	

Puesto que $F_{exp}=15.06 > F_{2,5}^{0.05}=5.79$, rechazamos la hipótesis nula H_0 . Es decir, el contraste conjunto **sí es significativo**.

Recordemos que al realizar los contrastes individuales (véase la transparencia 7 del Tema 4) obteníamos que, para un nivel de significación del 5%, **ninguna de las dos variables era significativa, individualmente**. Por tanto, estaríamos ante una situación de Multicolinealidad (caso 3). Veremos como tratar estos problemas en el Tema 5).

4. Correlación en regresión múltiple.

4.1. Coeficiente de Determinación

Se define como la **proporción de la variabilidad total** que queda **explicada por el modelo** de regresión múltiple.

$$R^2 = \frac{VE}{VT} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\hat{\beta}' \mathbf{X}' \mathbf{Y} - n\bar{y}^2}{\mathbf{Y}' \mathbf{Y} - n\bar{y}^2}$$

Inconvenientes:

- R^2 aumenta siempre cuando se aumenta el número de regresores, aunque éstos no sean significativos. Es decir, siempre puede incrementarse artificialmente el valor de R^2 añadiendo nuevas variables.
- R^2 es muy sensible a la formulación del modelo y a la elección de la variable respuesta. Es decir, es posible estimar dos modelos formalmente idénticos y con la misma capacidad predictiva pero que conducen a valores distintos de R^2 .

Puesto que $VT=VE+VNE$, $1-R^2$ es la **proporción de la variabilidad total no explicada por el modelo** de regresión.

Proposición:

$$\frac{s_R^2}{s_y^2} = \frac{n-1}{n-k-1} (1-R^2)$$

(Demostración)

De la relación anterior podemos ver cómo R^2 **aumenta siempre al aumentar k** :

$$\left(\frac{n-k-1}{n-1} \right) \frac{s_R^2}{s_y^2} = 1-R^2 \Rightarrow R^2 = 1 - \left(\frac{n-k-1}{n-1} \right) \frac{s_R^2}{s_y^2}$$

Coeficiente de Correlación Múltiple

Se define como:

$$R = \sqrt{R^2}$$

y mide el **grado de relación lineal** entre los regresores $x_1, \dots, x_k$ y la variable respuesta y .

Propiedades:

- 1) $|R| \leq 1$. Cuando $R=1$ existe una relación funcional exacta entre la respuesta y los regresores.
- 2) R es el coeficiente de correlación lineal (de Pearson) entre $\mathbf{Y}$ e $\hat{\mathbf{Y}}$.

4.2. El coeficiente de determinación corregido (*adjusted* R^2)

Para evitar que el efecto que sufre R^2 cuando se añaden nuevas variables, se define el **coeficiente de determinación corregido** como:

$$\tilde{R}^2 = 1 - \frac{VNE / (n-k-1)}{VT / (n-1)} = 1 - \frac{s_R^2}{s_y^2} = 1 - \frac{\frac{\mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}}{n-k-1}}{\frac{\mathbf{Y}'\mathbf{Y} - n\bar{y}^2}{n-1}}$$

Éste se utiliza para **comparar la eficacia de distintos modelos de regresión**, considerándose mejor modelo aquél con **mayor** coef. de determinación ajustado.

La siguiente proposición pone de manifiesto la relación entre el coef. de determinación y el coef. de determinación ajustado.

Proposición:

$$\tilde{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k-1}$$

(Demostración)

Ejemplo (Ejercicio 21): Con los datos del ejemplo del consumo de 8 provincias españolas, calcular el coeficiente de determinación y el coeficiente de determinación corregido.

El coeficiente de determinación es:

$$R^2 = \frac{VE}{VT} = \frac{1018.84}{1188} = 0.8576 \approx 0.86$$

El coeficiente de determinación sugiere que el **86%** de la variabilidad de la respuesta y queda explicada por el modelo de regresión múltiple; queda sin explicar el **100-86=14%** de la variabilidad de y .

Además, el grado de relación lineal entre $\hat{y}(=9.05+0.5203x_1+0.2397x_2)$ e y es elevado, puesto que $R=(0.8576)^{1/2}=0.9261$.

El coeficiente de determinación corregido es:

$$\tilde{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k-1} = 1 - (1 - 0.8576) \frac{7}{5} = 0.8006$$