

Adenda para comprender más en profundidad aspectos de regresión lineal múltiple

1. Evaluación de coeficientes de un modelo de regresión múltiple.

Dados los datos genéricos que siguen donde **y** es la variable dependiente y hay **p** **variables** independientes

Datos	y	X₁	X₂	X_p
1	y ₁	X ₁₁	X ₂₁	X _{p1}
2	y ₂	X ₁₂	X ₂₂	X _{p2}
3	y ₃	X ₁₃	X ₂₃	X _{p3}
.....
n	y _n	X _{1n}	X _{2n}	X _{pn}

Los coeficientes del modelo de regresión lineal múltiple se estiman como:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

Donde X en la ecuación anterior es (p variables más término independiente, p+1):

X₀	X₁	X₂	X_p
1	X ₁₁	X ₂₁	X _{p1}
1	X ₁₂	X ₂₂	X _{p2}
1	X ₁₃	X ₂₃	X _{p3}
.....
1	X _{1n}	X _{2n}	X _{pn}

\mathbf{X}^T

X₀	1	1	1	1
X₁	X ₁₁	X ₁₂	X ₁₃	X _{1n}
X₂	X ₂₁	X ₂₂	X ₂₃	X _{2n}
.....
X_p	X _{p1}	X _{p2}	X _{p3}	X _{pn}

$$X^T X = \begin{pmatrix} n & \sum_{i=1}^n X_{1i} & \dots & \sum_{i=1}^n X_{pi} \\ \sum_{i=1}^n X_{1i} & \sum_{i=1}^n X_{1i}^2 & \dots & \sum_{i=1}^n X_{1i} X_{pi} \\ \dots & \dots & \dots & \dots \\ \sum_{i=1}^n X_{pi} & \sum_{i=1}^n X_{pi} X_{1i} & \dots & \sum_{i=1}^n X_{pi}^2 \end{pmatrix}$$

$$X^T y = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n X_{1i} y_i \\ \dots \\ \sum_{i=1}^n X_{pi} y_i \end{pmatrix}$$

EJEMPLO DE APLICACIÓN

Se pretenden estimar los gastos en alimentación de una familia en base a la información que proporcionan las variables regresoras '*ingresos mensuales*' y '*número de miembros de la familia*'. Para ello se recoge una muestra aleatoria simple de 15 familias, cuyos resultados se facilitan en la tabla adjunta. (El gasto e ingreso se expresan en cien mil euros).

Gasto Alimentación	Ingresos	Tamaño
0,43	2,1	3
0,31	1,1	4
0,32	0,9	5
0,46	1,6	4
1,25	6,2	4
0,44	2,3	3
0,52	1,8	6
0,29	1	5
1,29	8,9	3
0,35	2,4	2
0,35	1,2	4
0,78	4,7	3
0,43	3,5	2
0,47	2,9	3
0,38	1,4	4

SOLUCIÓN:

$$y = \begin{pmatrix} 0.43 \\ 0.31 \\ 0.32 \\ 0.46 \\ 1.25 \\ 0.44 \\ 0.52 \\ 0.29 \\ 1.29 \\ 0.35 \\ 0.35 \\ 0.78 \\ 0.43 \\ 0.47 \\ 0.38 \end{pmatrix} = X\beta = \begin{pmatrix} 1 & 2.1 & 3 \\ 1 & 1.1 & 4 \\ 1 & 0.9 & 5 \\ 1 & 1.6 & 4 \\ 1 & 6.2 & 4 \\ 1 & 2.3 & 3 \\ 1 & 1.8 & 6 \\ 1 & 1 & 5 \\ 1 & 8.9 & 3 \\ 1 & 2.4 & 3 \\ 1 & 1.2 & 4 \\ 1 & 4.7 & 3 \\ 1 & 3.5 & 2 \\ 1 & 2.9 & 3 \\ 1 & 1.4 & 4 \end{pmatrix} * \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

$$X^T X = \begin{pmatrix} 15 & \sum_{i=1}^{15} X_{1i} & \sum_{i=1}^{15} X_{2i} \\ \sum_{i=1}^{15} X_{1i} & \sum_{i=1}^{15} X_{1i}^2 & \sum_{i=1}^{15} X_{1i}X_{2i} \\ \sum_{i=1}^{15} X_{2i} & \sum_{i=1}^{15} X_{1i}X_{2i} & \sum_{i=1}^{15} X_{2i}^2 \end{pmatrix}$$

X1	X2	X1 ²	X2 ²	X1*X2
Ingresos	Tamaño			
2.1	3	4.41	9	6.3
1.1	4	1.21	16	4.4
0.9	5	0.81	25	4.5
1.6	4	2.56	16	6.4
6.2	4	38.44	16	24.8
2.3	3	5.29	9	6.9
1.8	6	3.24	36	10.8
1	5	1	25	5
8.9	3	79.21	9	26.7
2.4	2	5.76	4	4.8
1.2	4	1.44	16	4.8
4.7	3	22.09	9	14.1
3.5	2	12.25	4	7
2.9	3	8.41	9	8.7
1.4	4	1.96	16	5.6

Σ X1	Σ X1	Σ X1 ²	Σ X2 ²	Σ X1X2
42	55	188.08	219	140.8

$$X^T X = \begin{pmatrix} 15 & 42 & 55 \\ 42 & 188.08 & 140.8 \\ 55 & 140.8 & 219 \end{pmatrix}$$

$$X^T y = \begin{pmatrix} \sum_{i=1}^{15} y_i \\ \sum_{i=1}^{15} X_{1i} y_i \\ \sum_{i=1}^{15} X_{2i} y_i \end{pmatrix}$$

y	X1	X2	y*X1	y*X2
Gasto Alimentación	Ingresos	Tamaño		
0.43	2.1	3	0.903	1.29
0.31	1.1	4	0.341	1.24
0.32	0.9	5	0.288	1.6
0.46	1.6	4	0.736	1.84
1.25	6.2	4	7.75	5
0.44	2.3	3	1.012	1.32
0.52	1.8	6	0.936	3.12
0.29	1	5	0.29	1.45
1.29	8.9	3	11.481	3.87
0.35	2.4	2	0.84	0.7
0.35	1.2	4	0.42	1.4
0.78	4.7	3	3.666	2.34
0.43	3.5	2	1.505	0.86
0.47	2.9	3	1.363	1.41
0.38	1.4	4	0.532	1.52

Σ y
8.07

Σ y*X1	Σ y*X2
32.063	28.96

$$X^T y = \begin{pmatrix} 8.07 \\ 32.063 \\ 28.96 \end{pmatrix}$$

$$(X^T X)^{-1} = \begin{pmatrix} 1.36 & -0.092 & -0.282 \\ -0.092 & 0.016 & 0.013 \\ -0.282 & 0.013 & 0.067 \end{pmatrix}$$

Y de aquí calculamos los coeficientes de regresión:

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 1.36 & -0.092 & -0.282 \\ -0.092 & 0.016 & 0.013 \\ -0.282 & 0.013 & 0.067 \end{pmatrix} * \begin{pmatrix} 8.07 \\ 32.063 \\ 28.96 \end{pmatrix}$$

Que resultan ser los siguientes:

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} -0,16 \\ 0,149 \\ 0,077 \end{pmatrix} \Rightarrow Y = -0,16 + 0,149 X_1 + 0,077 X_2 + \text{Residuo} \quad (\text{Modelo regresión lineal})$$

2. Obtención de residuos. Análisis de varianza de los residuos

Estimación de gasto de alimentación para el primer caso de los 15 cuyos valores son:
 $y_1 = 0,43$, $X_{11} = 2,1$ y $X_{21} = 3$ se obtiene

$$\text{estimación: } \hat{y}_1 = -0.16 + 0.149 \cdot 2.1 + 0.077 \cdot 3 = 0.3839$$

$$\text{error cometido: } e_1 = y_1 - \hat{y}_1 = 0.0461$$

Haciendo lo mismo para el resto de casos se tiene lo siguiente:

y	X1	X2	\hat{y}	e	e^2
Gasto Alimentación	Ingresos	Tamaño	Predicción	error	error al cuadrado
0.43	2.1	3	0.3839	0.0461	0.00213
0.31	1.1	4	0.3119	-0.0019	0.00000
0.32	0.9	5	0.3591	-0.0391	0.00153
0.46	1.6	4	0.3864	0.0736	0.00542
1.25	6.2	4	1.0718	0.1782	0.03176
0.44	2.3	3	0.4137	0.0263	0.00069
0.52	1.8	6	0.5702	-0.0502	0.00252
0.29	1	5	0.374	-0.084	0.00706
1.29	8.9	3	1.3971	-0.1071	0.01147
0.35	2.4	2	0.3516	-0.0016	0.00000
0.35	1.2	4	0.3268	0.0232	0.00054
0.78	4.7	3	0.7713	0.0087	0.00008
0.43	3.5	2	0.5155	-0.0855	0.00731
0.47	2.9	3	0.5031	-0.0331	0.00110
0.38	1.4	4	0.3566	0.0234	0.00055

Variación residual del modelo (**SCR**) es:

$$SCR = \sum_{i=1}^{15} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{15} e_i^2 = 0.0721$$

Variación explicada del modelo (**SCE**) es:

$$SCE = \sum_{i=1}^{15} (\hat{y}_i - \bar{y})^2$$

La media de la variable y (gasto de alimentación) es $\bar{y} = 0.538$

y	X1	X2	\hat{y}	$\hat{y} - \bar{y}$	$(\hat{y} - \bar{y})^2$
Gasto Alimentación	Ingresos	Tamaño	Predicción		
0.43	2.1	3	0.3839	-0.1541	0.02375
0.31	1.1	4	0.3119	-0.2261	0.05112
0.32	0.9	5	0.3591	-0.1789	0.03201
0.46	1.6	4	0.3864	-0.1516	0.02298
1.25	6.2	4	1.0718	0.5338	0.28494
0.44	2.3	3	0.4137	-0.1243	0.01545
0.52	1.8	6	0.5702	0.0322	0.00104
0.29	1	5	0.374	-0.164	0.02690
1.29	8.9	3	1.3971	0.8591	0.73805
0.35	2.4	2	0.3516	-0.1864	0.03474
0.35	1.2	4	0.3268	-0.2112	0.04461
0.78	4.7	3	0.7713	0.2333	0.05443
0.43	3.5	2	0.5155	-0.0225	0.00051
0.47	2.9	3	0.5031	-0.0349	0.00122
0.38	1.4	4	0.3566	-0.1814	0.03291

1.36464

SCE=1.36464

También SCE se puede expresar como: $SCE = \hat{\beta}^t X^t Y - n \bar{Y}^2$

$$SCE = \sum_{i=1}^n (\hat{y}_i - \bar{y}) \rightarrow SCE = \hat{Y}' \hat{Y} - n \bar{Y}^2 :$$

Variación total del modelo (**SCT**) es:

$$SCT = \sum_{i=1}^{15} (y_i - \bar{y})^2$$

Recuérdese de nuevo que la media de la variable y (gasto de alimentación) es $\bar{y} = 0.538$

y	X1	X2	y- \bar{y}	(y- \bar{y}) ²
Gasto Alimentación	Ingresos	Tamaño	error	error al cuadrado
0.43	2.1	3	-0.108	0.01166
0.31	1.1	4	-0.228	0.05198
0.32	0.9	5	-0.218	0.04752
0.46	1.6	4	-0.078	0.00608
1.25	6.2	4	0.712	0.50694
0.44	2.3	3	-0.098	0.00960
0.52	1.8	6	-0.018	0.00032
0.29	1	5	-0.248	0.06150
1.29	8.9	3	0.752	0.56550
0.35	2.4	2	-0.188	0.03534
0.35	1.2	4	-0.188	0.03534
0.78	4.7	3	0.242	0.05856
0.43	3.5	2	-0.108	0.01166
0.47	2.9	3	-0.068	0.00462
0.38	1.4	4	-0.158	0.02496
				1.43164

SCT=1.43164

Obsérvese que se da la siguiente relación: **SCT=SCE+SCR**

Demostración:

DESCOMPOSICIÓN DE LA VARIABILIDAD

$$SCT = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n \left[(Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y}) \right]^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + 2 \sum_{i=1}^n (Y_i - \hat{Y}_i) \cdot (\hat{Y}_i - \bar{Y})$$

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

suma cuadrados total (n-1) grados libertad suma cuadrados explicada k grados libertad suma cuadrados residual (n-k-1) grados libertad

Veamos por qué $2 \sum_{i=1}^n (Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) = 0$

$$\sum_{i=1}^n e_i (\hat{Y}_i - \bar{Y}) = \sum_{i=1}^n e_i \hat{Y}_i - \bar{Y} \sum_{i=1}^n e_i$$

Proposición A. Los residuos y las variables explicativas del modelo de regresión lineal múltiple son ortogonales. Demostración:

Al minimizar para encontrar β pasamos por la ecuación $X^T(y - X\beta) = 0$

$$X^T(y - X\beta) = X^T e = 0$$

Proposición B. Si el modelo de regresión lineal múltiple tiene un término constante, la suma de los residuos es igual a cero

Demostración:

$$X^T e = \begin{pmatrix} 1 & 1 & \dots & 1 \\ X_{11} & X_{12} & \dots & X_{1n} \\ \dots & \dots & \dots & \dots \\ X_{p1} & X_{p2} & \dots & X_{pn} \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n e_i \\ \sum_{i=1}^1 X_{1i} e_i \\ \dots \\ \sum_{i=1}^n X_{pi} e_i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \end{pmatrix}$$

En la primera fila se ve que la suma de los residuos del modelo ha de ser cero

$$\sum_{i=1}^n e_i = 0 \text{ y por lo tanto: } \bar{Y} \sum_{i=1}^n e_i = 0$$

Proposición C. Los residuos y los valores predichos de la variable dependiente son ortogonales. Demostración:

Al minimizar encontramos $\hat{y}^T e = \beta^T X^T e = 0$

Y por tanto

$$\sum_{i=1}^n e_i \hat{Y}_i = 0$$

Análisis de varianza

Tabla ANOVA			
Suma de cuadrados	g.l.	Varianza	Estadístico
$SCE = \sum_i (\hat{y}_i - \bar{y})^2$	k	$\frac{SCE}{k}$	$F = \frac{SCE/k}{SCR/(n-k-1)}$
$SCR = \sum_i (y_i - \hat{y}_i)^2$	$n - k - 1$	$S_R^2 = \frac{SCR}{n-k-1}$	
$SCT = \sum_i (y_i - \bar{y})^2$	$n - 1$		

En nuestro ejemplo $k=2$ (2 variables explicativas), $n=15$ el número de casos de la muestra.

La varianza residual S^2_{Res} es =

$$S^2_{\text{R}} = \frac{\text{SCR}}{15-2-1} = \frac{0,0721}{12} = 0,006$$

La desviación es

$$S_{\text{R}} = \sqrt{0,006} = 0,0775$$

El RMSE es $\text{sqrt}(0.0721)/\text{sqrt}(15) = 0.0693$

Coeficiente de determinación

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

$$R^2 = \frac{SCE}{SCT}$$

Aquí la suma de errores al cuadrado es:

$$\text{SCR} = \sum_{i=1}^{15} (Y_i - \hat{Y}_i)^2 = 0,0721$$

Y respecto a la media

$$\text{SCT} = \sum_{i=1}^{15} (Y_i - \bar{Y})^2 = 1,4316$$

Luego $R^2 = 1 - (0.0721/1.4316) = \mathbf{0.9496}$

Coeficiente de determinación corregido

$$R^2_{\text{AJU}} = 1 - \left(1 - R^2\right) \frac{n-1}{n-p-1}$$

$R^2_{\text{AJU}} = 1 - (1 - 0.9496) * (15-1)/(15-2-1) = \mathbf{0.9412}$

3. Contraste de hipótesis de un modelo de regresión lineal múltiple

La cuestión a verificar: ¿Es el modelo significativo?

Vamos a ver si se rechaza la hipótesis nula o no

Hipótesis nula $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ el modelo no es explicativo

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

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

F-Snedecor - COEFICIENTE DE DETERMINACIÓN

El coeficiente de determinación se define: $R^2 = \frac{SCE}{SCT} = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$

De otra parte, la distribución F-Snedecor:

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

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

En nuestro ejemplo:

Variación	Suma de Cuadrados	Gr. libertad	Media cuadrática	F-Snedecor
Explicada	$SCE = \sum_{i=1}^{15} (\hat{Y}_i - \bar{Y})^2 = 1,3595$	2	$\frac{SCE}{2} = 0,6797$	$F = \frac{SCE/2}{SCR/(15-2-1)} = 113,28$
Residual	$SCR = \sum_{i=1}^{15} (Y_i - \hat{Y}_i)^2 = 0,0721$	15 - 2 - 1	$\frac{SCR}{15-2-1} = 0,006$	
Total	$SCT = \sum_{i=1}^{15} (Y_i - \bar{Y})^2 = 1,4316$	15 - 1	$\frac{SCT}{15-1} = 0,1023$	

$$F = 0.6797/0.006 = 113.28$$

$$F_{0.5, 2, 12} = 3.8853$$

$$F \geq F_{0.5, 2, 12}$$

Se rechaza la hipótesis nula y se concluye que el modelo es significativo.

4. Varianza de los estimadores del modelo de regresión lineal múltiple

Distribución de los estimadores:

- Las estimaciones de los parámetros vienen dada por la expresión $\hat{\beta} = [X'X]^{-1} X'Y$ (siendo X' la matriz transpuesta del diseño).
- El vector de observaciones Y se distribuye según una normal multivariante de media $X\beta$ y de matriz de varianzas y covarianzas $\sigma^2 I$, es decir, $Y \in N(X\beta, \sigma^2 I)$.
- $\hat{\beta}$ 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:

- $E(\hat{\beta}) = E([X'X]^{-1} X'Y) = [X'X]^{-1} X'E(Y) = [X'X]^{-1} X'X\beta = \beta \Rightarrow \hat{\beta}$ es un estimador insesgado de β
- $Var(\hat{\beta}) = Var([X'X]^{-1} X'Y) = [X'X]^{-1} X' [Var(Y)] X [X'X]^{-1} = [X'X]^{-1} X' \sigma^2 X [X'X]^{-1} = \sigma^2 [X'X]^{-1}$

de donde, $\hat{\beta} \in N(\beta, \sigma^2 [X'X]^{-1})$

Entonces:

$$Var(\hat{\beta}) = (X^T X)^{-1} \sigma^2$$

Solo se desea las varianzas (diagonal de esa matriz inversa) y no las covarianzas

Recordemos que:

$$(X^T X)^{-1} = \begin{pmatrix} 1.36 & -0.092 & -0.282 \\ -0.092 & 0.016 & 0.013 \\ -0.282 & 0.013 & 0.067 \end{pmatrix}$$

σ^2 es la varianza residual = $SCR/(15-2-1) = 0.006$.

$$Var(\hat{\beta}_i) = \sigma^2 [X'X]^{-1} \approx S_R^2 [X'X]^{-1} = (0,006) \begin{matrix} q_{i+1, i+1} \equiv \text{elemento de } [X'X]^{-1} \\ \begin{pmatrix} 1,36 & & \\ & 0,016 & \\ & & 0,067 \end{pmatrix} \end{matrix} = \begin{matrix} S_R^2 q_{i+1, i+1} \\ \begin{pmatrix} 0,00816 & & \\ & 0,000096 & \\ & & 0,0004 \end{pmatrix} \end{matrix}$$

$$\text{de donde se deduce, } \begin{cases} Var(\beta_0) = 0,00816 \Rightarrow \sigma_{\beta_0} = \sqrt{0,00816} = 0,0903 \\ Var(\beta_1) = 0,000096 \Rightarrow \sigma_{\beta_1} = \sqrt{0,000096} = 0,0098 \\ Var(\beta_2) = 0,0004 \Rightarrow \sigma_{\beta_2} = \sqrt{0,0004} = 0,02 \end{cases}$$

- Intervalo de confianza para los parámetros: $IC_{1-\alpha}(\beta_i) = \left[\hat{\beta}_i \pm t_{\alpha/2, (n-k-1)} S_R \sqrt{q_{i+1, i+1}} \right]$

$$\hat{\beta}_0 = -0,160 \quad \hat{\beta}_1 = 0,149 \quad \hat{\beta}_2 = 0,077 \quad t_{0,05, 12} = 1,782$$

$$IC_{1-\alpha}(\beta_0) = \left[-0,160 \pm (1,782) \sqrt{0,00816} \right] = \left[-0,321; 0,001 \right]$$

$$IC_{1-\alpha}(\beta_1) = \left[0,149 \pm (1,782) \sqrt{0,000096} \right] = \left[0,1315; 0,1665 \right] \quad (\text{Ingreso})$$

$$IC_{1-\alpha}(\beta_2) = \left[0,077 \pm (1,782) \sqrt{0,0004} \right] = \left[0,0414; 0,1126 \right] \quad (\text{Tamaño})$$

RESUMEN DE CONTRASTES

Contraste Conjunto F-Snedecor	Contrastes Individuales t-Student	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

5. Contraste de hipótesis sobre importancia de las variables explicativas

Nos planteamos si la variable X_i influye sobre la variable de respuesta Y . En otras palabras, si el valor del parámetro en la población es cero o no.

Para ello, se establece la hipótesis nula $H_0: \beta_i = 0$ frente a la hipótesis alternativa $H_1: \beta_i \neq 0$.

El estadístico observado $t = \frac{\hat{\beta}_i - \beta_i}{S_R \sqrt{q_{i+1, i+1}}}$, bajo la hipótesis nula resulta, $t = \frac{\hat{\beta}_i}{S_R \sqrt{q_{i+1, i+1}}}$

Se acepta la hipótesis nula H_0 cuando $\underbrace{\frac{\hat{\beta}_i}{S_R \sqrt{q_{i+1, i+1}}}}_{\text{estadístico experimental}} \leq \underbrace{t_{\frac{\alpha}{2}; (n-k-1)}}_{\text{estadístico teórico}}$. En caso contrario, se rechaza.

Ejemplo: ¿Influye la variable tamaño familiar?

Nos planteamos si la variable X_2 (tamaño) influye sobre la variable de respuesta Y (gastos). En otras palabras, si el valor del parámetro en la población es cero o no.

Para ello, se establece la hipótesis nula $H_0: \beta_2 = 0$ frente a la hipótesis alternativa $H_1: \beta_2 \neq 0$.

El estadístico observado $t = \frac{\hat{\beta}_2 - \beta_2}{S_R \sqrt{q_{33}}}$, bajo la hipótesis nula resulta: $t = \frac{\hat{\beta}_2}{S_R \sqrt{q_{33}}}$

Por tanto,

$$\hat{\beta}_2 = 0,077 \quad S_R q_{33} = (0,0775) \sqrt{0,0004} = 0,00155 \quad t_{0,05,12} = 1,782$$

El estadístico experimental $t = \frac{\hat{\beta}_2}{S_R \sqrt{q_{33}}} = \frac{0,077}{0,00155} = 49,67$

Siendo $|t| > t_{0,05,12}$ se rechaza la hipótesis nula, afirmando, con un 90% de fiabilidad, que el número de miembros de la familia influye en los gastos de alimentación.