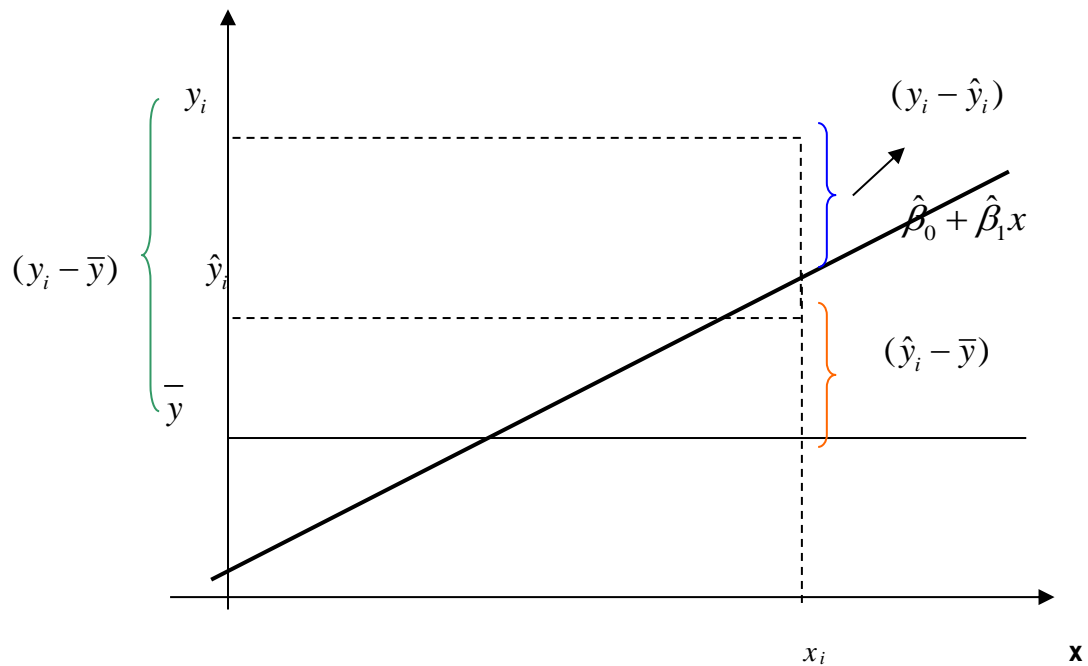


Regresión Múltiple

DESCOMPOSICIÓN DE LA VARIACIÓN TOTAL



¿Qué es la V.T.?

En el gráfico para una observación se tiene:

$$(y_i - \bar{y}) = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$$

Variación total = Variación explicada + Variación no explicada

$$(y_i - \bar{y})^2 = [(\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)]^2$$

Para “n” observaciones:

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

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

$$STC = SCR + SCE$$

2.2 Coeficiente de Determinación Múltiple

Es una medida descriptiva del ajuste global del modelo de regresión lineal múltiple y se obtiene dividiendo a la expresión (α) por STC:

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

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

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

2.3 PROPIEDADES:

1º Es una cantidad no negativa: $0 \leq R^2 \leq 1$

Si $R^2 \rightarrow 0$

Si $R^2 \rightarrow 1$

2º $100(1 - R^2)$ representa el porcentaje de la variabilidad no explicada por la relación lineal múltiple.

3º Se utiliza para comparar varios modelos de regresión.

PRUEBAS DE HIPOTESIS EN EL MRLM

Luego de estimar los parámetros del modelo y si $e_i \sim N(0, \sigma^2)$, podemos elaborar las siguientes tablas:

Tabla básica

ANVA

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Valor F
Regresión	p	$\mathbf{b}'\mathbf{X}'\mathbf{Y}$	$\text{CM}_{\text{Reg}} = \mathbf{b}'\mathbf{X}'\mathbf{Y}/p$	$F = \frac{\text{CM}_{\text{Reg}}}{\text{CM}_{\text{Res}}}$
Residual	$n - p$	$\mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y}$	$\text{CM}_{\text{Res}} = \frac{\mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y}}{n - p}$	
Total (no corregido)	n	$\mathbf{Y}'\mathbf{Y}$		

Tabla básica

ANVA

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Valor F
Regresión	p	$\mathbf{b}'\mathbf{X}'\mathbf{Y} - n\bar{Y}^2$	CM_{Reg}	$F = \frac{\text{CM}_{\text{Reg}}}{\text{CM}_{\text{Res}}}$
Residual	$n - p$	$\mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y}$	CM_{Res}	
Total (corregido)	$n - 1$	$\mathbf{Y}'\mathbf{Y} - n\bar{Y}^2$		

1° Prueba de significancia

Hipótesis:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1: \beta_j \neq 0 \text{ para al menos un valor de "j"}$$

La estadística de prueba bajo la hipótesis nula es:

$$F_0 = \frac{\text{CMR}}{\text{CM Res}} \sim F(k, n - p)$$

TABLA ANVA

Fuente de Variación	SC	g.l	CM	Estadística
Regresión	SCR	k	$CMR = SRC / k$	$F_0 = CMR / CM Res$
Residuales	SCE	$n-k-1$	$CMRes = SCRes / (n-k-1)$	
Total	STC	$n-1$		

Regla de Decisión:

Para un nivel de significancia “ α ”, se rechaza la hipótesis nula si:

$$F_0 > F_{(\alpha, k, n-p)}$$

2º Pruebas sobre los coeficientes individuales

1. Se incluye otra variable regresora
2. Se elimina una o más variables regresoras presentes en el modelo

Se debe decidir si se justifica la inclusión de otra variable regresora en el modelo.

La hipótesis para probar la significancia de cualquier coeficiente de regresión es

Hipótesis:

$$H_0 : \beta_j = 0 \quad j = 1, 2, \dots, k$$

$$H_1 : \beta_j \neq 0$$

Estadística para la prueba:

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 c_{jj}}} \sim t(n-p)$$

Regla de Decisión:

Para un nivel de significación “ α ”, se rechaza la hipótesis nula si:

$$|t_0| > t_{(\alpha/2; n-p)}$$

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	contrastes 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

CONTRIBUCIÓN DE UN SUBCONJUNTO DE LAS VARIABLES REGRESORAS para el modelo.

Considérese el modelo de regresión con k regresores

$$Y = X\beta + \varepsilon,$$

donde

y es un vector $n \times 1$,

X es una matriz $n \times p$,

β es un vector $p \times 1$,

ε es un vector $n \times 1$ y $p = k + 1$.

Se desea determinar si algún subconjunto de $r < k$ regresores contribuyen en forma significativa al modelo de regresión. Se a seccionado como sigue el vector de los coeficientes de regresión:

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

donde β_1 es un vector $(p-r) \times 1$ y β_2 es un vector $r \times 1$. Se desean probar las siguientes hipótesis

$$H_0 : \beta_2 = 0$$

$$H_1 : \beta_2 \neq 0$$

Este modelo se puede escribir como sigue:

$$y = X\beta + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

en el que la matriz X_1 de $n \times (p-r)$ representa a las columnas de X asociadas con β_1 y la matriz X_2 de $n \times r$ representa a las columnas de X asociadas con β_2 . A este se le llama el modelo completo.

Para el modelo completo, se sabe que

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

$$SCR(\beta) = \hat{\beta}'X'Y \quad (p \text{ g.l.})$$

y

$$CMRes = (Y'Y - \hat{\beta}'X'Y)/(n-p)$$

Para determinar la contribución de los términos de β_2 a la regresión se ajusta el modelo suponiendo que es cierta

$H_0 : \beta_2 = 0$

Este es conocido como el modelo reducido y está dado por

$$Y = X_1\beta_1 + \varepsilon,$$

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'y.$$

$$SCR(\beta_1) = \hat{\beta}_1'X_1'y \quad (p-r \text{ g.l.})$$

$$SCR(\beta_2|\beta_1) = SCR(\beta) - SCR(\beta_1) \quad \text{con } p-(p-r) = r \text{ g.l.}$$

Esta suma de cuadrados se llama **suma extra de cuadrados debida a β_2**

$$F_0 = (SCR(\beta_2|\beta_1)/r)/CMRes$$

Si $F_0 > F_{\alpha, r, n-p}$, se rechaza H_0

Tarea:

Resolver caso especial de columnas ortogonales en X