

Prácticas Tema 3: El modelo lineal básico

Ana J. López y Rigoberto Pérez

Departamento de Economía Aplicada. Universidad de Oviedo

PRACTICA 3.1- Se desea estimar un modelo para explicar la tasa de paro (Y , en %) en función del crecimiento económico (X_2 , tasa de variación del PIB en %) y la inflación (X_3 , tasa de inflación, en %). Se dispone únicamente de la información correspondiente a 6 países, resumida en la tabla siguiente:

Paro, Y	Crecimiento, X_2	Inflación, X_3
6	6	1
12	2	3
11	2	2,5
9	4	2,5
18	0,5	4,5
7	5	3

- a) Estimar un modelo econométrico lineal que explique la tasa de paro en función del crecimiento económico y la inflación, estudiando su bondad.
- b) Analizar la significación de las variables explicativas
- c) Según el modelo estimado ¿Cuál sería la tasa de paro prevista en un país si se espera que el PIB crezca al 3 % y la inflación alcance el 2 %?

SOLUCIÓN:

- a) Estimar un modelo econométrico lineal que explique la tasa de paro en función del crecimiento económico y la inflación, estudiando su bondad.

Con la información disponible se obtiene la matriz de datos $\mathbf{X} = \begin{pmatrix} 1 & 6 & 1 \\ 1 & 2 & 3 \\ 1 & 2 & 2,5 \\ 1 & 4 & 2,5 \\ 1 & 0,5 & 4,5 \\ 1 & 5 & 3 \end{pmatrix}$

que conduce a $\mathbf{X}'\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 6 & 2 & 2 & 4 & 0,5 & 5 \\ 1 & 3 & 2,5 & 2,5 & 4,5 & 3 \end{pmatrix} \begin{pmatrix} 1 & 6 & 1 \\ 1 & 2 & 3 \\ 1 & 2 & 2,5 \\ 1 & 4 & 2,5 \\ 1 & 0,5 & 4,5 \\ 1 & 5 & 3 \end{pmatrix} = \begin{pmatrix} 6 & 19,5 & 16,5 \\ 19,5 & 85,25 & 44,25 \\ 16,5 & 44,25 & 51,75 \end{pmatrix}$

cuyo determinante adopta el valor $|\mathbf{X}'\mathbf{X}| = 309,375$, a partir del cual se obtiene la matriz inversa

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 7,9309 & -0,9018 & -1,75758 \\ -0,9018 & 0,12364 & 0,18182 \\ -1,75758 & 0,18182 & 0,42424 \end{pmatrix}$$

Por su parte se obtiene el vector $\mathbf{X}'\mathbf{y}$

$$\mathbf{X}'\mathbf{y} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 6 & 2 & 2 & 4 & 0,5 & 5 \\ 1 & 3 & 2,5 & 2,5 & 4,5 & 3 \end{pmatrix} \begin{pmatrix} 6 \\ 11 \\ 10 \\ 9 \\ 18 \\ 7 \end{pmatrix} = \begin{pmatrix} 63 \\ 162 \\ 194 \end{pmatrix} \text{ y como producto de}$$

los dos resultados anteriores se obtiene el vector de estimadores mínimo cuadráticos:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{pmatrix} 67,9309 & -0,9018 & -1,75758 \\ -0,9018 & 0,12364 & 0,18182 \\ -1,75758 & 0,18182 & 0,42424 \end{pmatrix} \begin{pmatrix} 63 \\ 162 \\ 194 \end{pmatrix} = \begin{pmatrix} 12,583 \\ -1,513 \\ 1,03 \end{pmatrix}$$

Como se puede observar los signos coinciden con los esperados, de modo que al aumentar un 1 % el PIB, ceteris paribus, el paro se reduce 1,5 % mientras al aumentar 1 % la inflación (ceteris paribus) la tasa de paro aumenta algo más de un punto porcentual.

Los parámetros estimados conducen a los valores teóricos:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} = \begin{pmatrix} 4,537 \\ 12,6485 \\ 12,133 \\ 9,1079 \\ 16,463 \\ 8,1103 \end{pmatrix}$$

y los residuos o errores de estimación recogidos en el vector $\hat{\mathbf{u}}$:

$$\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\beta} = \begin{pmatrix} 6 \\ 12 \\ 10 \\ 9 \\ 18 \\ 8 \end{pmatrix} - \begin{pmatrix} 4,537 \\ 12,6485 \\ 12,133 \\ 9,1079 \\ 16,463 \\ 8,1103 \end{pmatrix} = \begin{pmatrix} 1,463 \\ -0,64806 \\ -1,133 \\ -0,1079 \\ 1,537 \\ -1,1103 \end{pmatrix}$$

Tal y como se sabe, la suma de los residuos es nula.

Para estudiar la bondad del modelo estimado se lleva a cabo el análisis de la varianza cuantificando la variación total de los ingresos y sus componentes (explicada y residual):

$$\begin{cases} VT &= \mathbf{y}'\mathbf{y} - n\bar{y}^2 = 93,5, \text{ siendo: } \bar{y} = 10,5 \\ VE &= \hat{\beta}\mathbf{X}'\mathbf{y} - n\bar{Y}^2 = \hat{\mathbf{y}}'\hat{\mathbf{y}} - n\bar{Y}^2 = 86,0479 \\ VNE &= \mathbf{y}'\mathbf{y} - \hat{\beta}\mathbf{X}'\mathbf{y} = \hat{\mathbf{u}}'\hat{\mathbf{u}} = 7,4521 \end{cases}$$

La bondad del modelo puede ser analizada mediante el coeficiente de determinación:

$$R^2 = 1 - \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\mathbf{y}'\mathbf{y} - n\bar{Y}^2} = 0,9203$$

b) Analizar la significación de las variables explicativas

Los contrastes de significación global e individual pueden ser llevados a cabo mediante los contrastes F y t respectivamente.

Así, con la información anteriormente obtenida la discrepancia F adoptaría un valor:

$$F = \frac{\frac{\hat{\beta}\mathbf{X}'\mathbf{y} - n\bar{Y}^2}{\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-k}}}{\frac{2}{7,4521}} = \frac{93,5}{3} = 15,32$$

Este resultado de la discrepancia F podría ser también obtenido a partir del coeficiente de determinación:

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

El nivel crítico asociado al resultado obtenido es bajo $p = P(F_3^2 > 15,32) \approx 0,022$ con lo cual se rechaza la hipótesis de nulidad de los coeficientes ($H_0 : \beta_2 = \beta_3 = 0$).

Para contrastar la significación individual mediante el contraste t obtenemos la varianza estimada $S^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-k} = 2,1688$ y la matriz de varianzas-covarianzas de los estimadores:

$$\begin{aligned} S^2(\hat{\beta}) = S^2(X'X)^{-1} &= 2,484 \begin{pmatrix} 7,9309 & -0,9018 & -1,75758 \\ -0,9018 & 0,12364 & 0,18182 \\ -1,75758 & 0,18182 & 0,42424 \end{pmatrix} \\ &= \begin{pmatrix} 19,7007 & -2,2402 & -4,3659 \\ -2,2402 & 0,12364 & 0,18182 \\ -4,3659 & 0,18182 & 0,42424 \end{pmatrix} \end{aligned}$$

con lo cual se tienen los resultados resumidos en la tabla:

H_0	$\hat{\beta}_j$	$S(\hat{\beta}_j)$	$\frac{\hat{\beta}_j}{S(\hat{\beta}_j)}$	$p = P(t_3 > d^*)$	Conclusión
$\beta_2 = 0$	-1,513	0,5541	-2,73	0,07	Significativo al 10 %
$\beta_3 = 0$	1,03	1,0265	1,004	0,39	No significativo

Así pues, nuestra conclusión (con un nivel crítico del 7 %) sería rechazar la nulidad del parámetro β_2 y por tanto validar como explicativa la variable crecimiento económico. En cambio no es posible rechazar la nulidad de β_3 con lo cual concluimos que la inflación no parece tener validez como variable explicativa del paro en la muestra de países analizados.

c) Según el modelo estimado ¿Cuál sería la tasa de paro prevista en un país si se espera que el PIB crezca al 3 % y la inflación alcance el 2 %?

El vector de observaciones para el año próximo sería $\mathbf{x}_0 = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}$ a partir del cual se obtiene una predicción para la tasa de paro: $\hat{Y}_0 = \mathbf{x}_0'\hat{\beta} = 10,11$.

Si se quieren obtener bandas de confianza al 95 % para esta predicción vendrán dadas por la expresión $\left[\hat{Y}_0 \pm kS\sqrt{1+\mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}\right]$ donde la constante $k = 3,1824$ se obtiene en tablas de la t de Student para el nivel de confianza pedido y 3 grados de libertad, mientras la información disponible proporciona el resultado $\mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 = 0,48$. Como consecuencia, las bandas al 95 % de confianza para la predicción de paro son $[4,16, 21]$.

PRACTICA 3.2- Una empresa automovilística estudia conjuntamente la evolución de las ventas de uno de sus modelos (Y , en miles de unidades), su precio (X_2 , en miles de euros) y el precio medio de sus competidores más directos (X_3 , en miles de euros), obteniendo la siguiente información durante los últimos 24 años:

$$\begin{aligned}\sum_{i=1}^{24} Y_i &= 2430 & \sum_{i=1}^{24} X_{2i} &= 4,83 & \sum_{i=1}^{24} X_{3i} &= 6,415 \\ \sum_{i=1}^{24} Y_i^2 &= 250418 & \sum_{i=1}^{24} X_{2i}^2 &= 1,042 & \sum_{i=1}^{24} X_{3i}^2 &= 1,8805 \\ \sum_{i=1}^{24} Y_i X_{2i} &= 501,85 & \sum_{i=1}^{24} Y_i X_{3i} &= 671,42 & \sum_{i=1}^{24} X_{2i} X_{3i} &= 1,39235\end{aligned}$$

a) Estimar una función lineal de demanda del vehículo, comentando el resultado.

b) Llevar a cabo contrastes de significación global e individuales sobre el modelo anterior.

c) Se sospecha que el impacto marginal sobre la demanda del precio del vehículo es exactamente el opuesto al del precio de sus competidores. Sabiendo que bajo este supuesto el modelo estimado es $\hat{Y} = 106,0453 - 93,38(X_2 - X_3)$ con $\hat{\mathbf{u}}'_R \hat{\mathbf{u}}_R = 4008,54$, justificar si es admisible su hipótesis.

d) Se contempla fijar un precio de 20000 euros para el vehículo el próximo año, y se estima que el precio de la competencia se situará en 25000 euros. Con esta información ¿Cuál será la demanda prevista el próximo año? Construir una banda de confianza al 95 %.

SOLUCIÓN:

a) Estimar una función lineal de demanda del vehículo, comentando el resultado.

El modelo de demanda planteado será del tipo $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$, (en forma condensada $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$) y para la estimación del vector $\boldsymbol{\beta}$ utilizaremos la expresión mínimo-cuadrática $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$. Teniendo en cuenta que $k = 3$ y $n = 24$, la matriz $\mathbf{X}'\mathbf{X}$ viene dada por la expresión:

$$\begin{aligned}\mathbf{X}'\mathbf{X} &= \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_{2,1} & X_{2,2} & \cdots & X_{2,24} \\ X_{3,1} & X_{3,2} & \cdots & X_{3,24} \end{pmatrix} \begin{pmatrix} 1 & X_{2,1} & X_{3,1} \\ 1 & X_{2,2} & X_{3,2} \\ \vdots & \vdots & \vdots \\ 1 & X_{2,24} & X_{3,24} \end{pmatrix} \\ &= \begin{pmatrix} 24 & \sum_{i=1}^{24} X_{2i} & \sum_{i=1}^{24} X_{3i} \\ \sum_{i=1}^{24} X_{2i} & \sum_{i=1}^{24} X_{2i}^2 & \sum_{i=1}^{24} X_{2i} X_{3i} \\ \sum_{i=1}^{24} X_{3i} & \sum_{i=1}^{24} X_{2i} X_{3i} & \sum_{i=1}^{24} X_{3i}^2 \end{pmatrix}\end{aligned}$$

que con la información disponible conduce al resultado:

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 24 & 4,83 & 6,415 \\ 4,83 & 1,042 & 1,39235 \\ 6,415 & 1,39235 & 1,8805 \end{pmatrix}$$

El determinante de esta matriz es $|\mathbf{X}'\mathbf{X}| = 0,032177$ y su inversa:

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 0,6505 & -4,6931 & 1,2558 \\ -4,6931 & 123,7020 & -75,5801 \\ 1,2558 & -75,5801 & 52,2079 \end{pmatrix}$$

y el producto $\mathbf{X}'\mathbf{y}$ vendría dado por:

$$\mathbf{X}'\mathbf{y} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_{2,1} & X_{2,2} & \cdots & X_{2,24} \\ X_{3,1} & X_{3,2} & \cdots & X_{3,24} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{24} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{24} Y_i \\ \sum_{i=1}^{24} X_{2i} Y_i \\ \sum_{i=1}^{24} X_{3i} Y_i \end{pmatrix}$$

que en nuestro caso proporciona el resultado: $\mathbf{X}'\mathbf{y} = \begin{pmatrix} 2430 \\ 501,85 \\ 671,42 \end{pmatrix}$.

Como consecuencia, podemos calcular el vector de estimadores mínimo cuadrados:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{pmatrix} 68,62 \\ -70,37 \\ 175,05 \end{pmatrix}$$

llegando así al modelo estimado $\hat{Y} = 68,62 - 70,37X_2 + 175,05X_3$ en el que se observan signos coincidentes con los esperados: $\beta_2 = \frac{\partial Y}{\partial X_2} < 0$, $\beta_3 = \frac{\partial Y}{\partial X_3} > 0$.

Para estudiar la bondad de este modelo llevamos a cabo un análisis de varianza, cuantificando la dispersión total de la demanda y de sus componentes explicada y residual:

Variación Total	$\mathbf{y}'\mathbf{y} - n\bar{Y}^2 = 250418 - 24(101,25)^2 = 4380,5$
Variación Explicada	$\hat{\beta}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2 = 248969,71 - 24(101,25)^2 = 2932,2054$
Variación no explicada	$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y} = 250418 - 248969,71 = 1448,2946$

A partir de esta información calculamos el coeficiente de determinación del modelo:

$$R^2 = 1 - \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\mathbf{y}'\mathbf{y} - n\bar{Y}^2} = 1 - \frac{1448,29}{4380,5} = 0,6694$$

que permite afirmar que casi un 67% de las variaciones de la demanda son explicadas mediante el modelo lineal estimado.

b) Llevar a cabo contrastes de significación global e individuales sobre el modelo anterior.

El contraste global se lleva a cabo mediante la F de Snedecor a partir del análisis de varianza anterior.

$$F = \frac{\frac{\beta' \mathbf{X}' \mathbf{y} - n \bar{Y}^2}{\frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{n-k}}}{\frac{2932,21}{1448,29}} = \frac{2932,21}{24-3} = 21,2582$$

y su nivel crítico es $p = P(F_{21}^2 > 21,258) \approx 0$ con lo cual rechazamos la hipótesis nula $H_0 : \beta_2 = \beta_3 = 0$. Otra posibilidad sería calcular F mediante su conexión con el coeficiente de determinación ya calculado:

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

El contraste de significación individual asociado a cada una de las variables consideradas como explicativas en el modelo se lleva a cabo mediante la discrepancia $d_{\hat{\beta}} = \frac{\hat{\beta} - \beta}{S_{\hat{\beta}}} \approx t_{n-k}$, para cuyo cálculo debemos conocer la estimación de la matriz de varianzas-covarianzas.

La estimación de σ^2 se lleva a cabo mediante: $S^2 = \frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{n-k} = \frac{1448,2946}{24-3} = 68,9664$, con lo cual la matriz de varianzas-covarianzas estimada es:

$$S^2(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 44,86 & -323,6663 & 86,6051 \\ -323,6663 & 8531,2817 & -5212,4864 \\ 86,6051 & -5212,4864 & 3600,5904 \end{pmatrix}$$

Esta información y los coeficientes mínimo cuadráticos anteriormente calculados permiten obtener las discrepancias muestrales asociadas a las dos variables explicativas. Así, para contrastar la hipótesis $H_0 : \beta_2 = 0$ se calcula $\frac{\hat{\beta}_2}{S_{\hat{\beta}_2}} = \frac{-70,3664}{\sqrt{8531,2817}} = -0,7618$, valor que lleva asociado el nivel crítico: $p = P(|t_{21}| > |-0,7618|) = 0,45$ que no conduce al rechazo de la hipótesis de nulidad de β_2 .

Siguiendo el mismo razonamiento se contrasta $H_0 : \beta_3 = 0$. En este caso la discrepancia observada es $\frac{\hat{\beta}_3}{S_{\hat{\beta}_3}} = \frac{175,0482}{\sqrt{3600,5904}} = 2,9172$ y el nivel crítico: $p = P(|t_{21}| > |2,9172|) = 0,009$ que por ser suficientemente bajo conduce al rechazo de la hipótesis planteada.

El resultado de estos contrastes aconsejaría algún cambio en la especificación del modelo, sustituyendo la variable *Precio de los vehículos* por otra que resultase significativa. Esta conclusión puede deberse a la presencia de una correlación elevada entre los dos precios (se obtiene $r_{X_2, X_3} = 0,94$) que, como veremos en temas posteriores, nos sitúa ante un problema de multicolinealidad aproximada, que puede haber causado el aumento en la varianza del estimador $S(\hat{\beta}_2)$ y la consiguiente disminución del valor de la t .

c) Se sospecha que el impacto marginal sobre la demanda del precio del vehículo es exactamente el opuesto al del precio de sus competidores. Sabiendo que bajo este supuesto el modelo estimado es $\hat{Y} = 106,0453 - 93,38(X_2 - X_3)$ con $\hat{\mathbf{u}}_R' \hat{\mathbf{u}}_R = 4008,54$, justificar si es admisible su hipótesis.

En este caso debemos contrastar el supuesto planteado $H_0 : \beta_2 = -\beta_3$, que podría ser también expresado como

$$\mathbf{D}\beta = \beta^*, \text{ con } \mathbf{D} = (011), \text{ y } \beta^* = (0)$$

Para ello se comparan los residuos del modelo sometido a esta restricción con los del modelo anteriormente estimado sin restricciones, mediante la expresión:

$$\left(\frac{\hat{u}'_R \hat{u}_R - \hat{u}' \hat{u}}{\hat{u}' \hat{u}} \right) \frac{n-k}{r} \simeq F_{n-k}^r$$

donde r es el número de restricciones ($r=1$ en este caso).

La expresión $\hat{\mathbf{u}}'_R \hat{\mathbf{u}}_R = 4008,54$ se ha obtenido a partir de los errores del modelo restringido, $\hat{Y}_R = 106,0453 - 93,38(X_2 - X_3)$, y la expresión de la discrepancia proporciona el resultado $F^* = 37,123$ que lleva asociado un nivel crítico muy bajo ($p = P(F_{21}^1 > 37,123) \approx 0$) y por tanto conduce al rechazo de la hipótesis planteada.

d) Se contempla fijar un precio de 20000 euros para el vehículo el próximo año, y se estima que el precio de la competencia se situará en 25000 euros. Con esta información ¿Cuál será la demanda prevista el próximo año? Construir una banda de confianza al 95 %.

El vector de observaciones para el año próximo sería $\mathbf{x}_0 = \begin{pmatrix} 1 \\ 20 \\ 25 \end{pmatrix}$ a partir

del cual se obtiene para el valor esperado de la demanda la predicción puntual: $\hat{Y}_0 = \mathbf{x}'_0 \hat{\beta} = 3037,5$.

Por su parte, las bandas de confianza al 95 % para esta predicción vendrán dadas por la expresión $\left[\hat{Y}_0 \pm kS\sqrt{1+\mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0} \right]$ donde la constante $k = 2,0796$ se obtiene en tablas de la t de Student para el nivel de confianza pedido y la información disponible proporciona $S^2 [1 + \mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0] = 441892,734$.

Como consecuencia, el intervalo al 95 % de confianza para la demanda de automóviles conduce al resultado $[1655,08 \ 4419,914]$.

PRACTICA 3.3- A partir de información suministrada por un organismo internacional para 30 países se analiza la inversión en educación (Y , en millones de dólares) en función de la Población (X_2 , en millones de habitantes) y la Renta per cápita (X_3 , en millones de dólares), obteniendo los resultados:

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 12,97 & 0,145 & 0,34 \\ -0,145 & 0,002 & -0,0042 \\ 0,34 & -0,0042 & 0,014 \end{pmatrix}; \mathbf{X}'\mathbf{y} = \begin{pmatrix} 1070 \\ 118116 \\ 9600 \end{pmatrix}; \mathbf{y}'\mathbf{y} = 5212600$$

a) Estimar por mínimos cuadrados un modelo que explique la inversión en educación en función de la población y la renta per cápita. ¿Cuál es su capacidad explicativa?

b) Llevar a cabo los contrastes de significación individual interpretando los resultados.

c) Se especifica un modelo que distingue entre población femenina (X_2^F) y masculina (X_2^M) obteniéndose las siguientes estimaciones:

$$\begin{aligned} \hat{Y} &= 14 + 42X_2^F + 56X_2^M + 2,4X_3 & \hat{\mathbf{u}}'\hat{\mathbf{u}} &= 350000 \\ \hat{Y} &= 15 + 50(X_2^F + X_2^M) + 2,2X_3 & \hat{\mathbf{u}}'_R \hat{\mathbf{u}}_R &= 350000 \end{aligned}$$

Justificar cuál es la restricción impuesta en el segundo modelo y si, con la información estadística disponible, debería ser o no rechazado este supuesto.

SOLUCIÓN:

a) Estimar por mínimos cuadrados un modelo que explique la inversión en educación en función de la población y la renta per cápita. ¿Cuál es su capacidad explicativa?

El modelo planteado será del tipo $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$, (en forma condensada $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$) y para la estimación del vector $\boldsymbol{\beta}$ utilizaremos la expresión mínimo-cuadrática $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, de forma que el vector de estimadores mínimo cuadráticos:

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} 15,08 \\ 40,762 \\ 2,112 \end{pmatrix}$$

Para estudiar la bondad de este modelo llevamos a cabo un análisis de varianza, cuantificando la variación total de la inversión en educación, la variación explicada por el modelo y la no explicada o residual:

Variación Total	$\mathbf{y}'\mathbf{y} - n\bar{Y}^2 = 5212600 - 30(35,67)^2 = 5174436,67$
Variación Explicada	$\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2 = 4851063,87 - 30(35,67)^2 = 4812899,53$
Variación no explicada	$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = 5212600 - 4851062,87 = 361537,13$

A partir de esta información calculamos el coeficiente de determinación del modelo: $R^2 = 1 - \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\mathbf{y}'\mathbf{y} - n\bar{Y}^2} = 0,93$ que permite afirmar que el 93 % de las variaciones de la inversión en educación son explicadas mediante el modelo lineal estimado.

b) Llevar a cabo los contrastes de significación individual interpretando los resultados.

El contraste de significación individual asociado a cada una de las variables considerada como explicativa en el modelo se lleva a cabo mediante la discrepancia $d_{\hat{\beta}} = \frac{\hat{\beta} - \beta}{S_{\hat{\beta}}} \approx t_{n-k}$, para cuyo cálculo debemos conocer la estimación de

la matriz de varianzas-covarianzas, $S^2(\mathbf{X}'\mathbf{X})^{-1}$. La varianza muestral se obtiene mediante la expresión $S^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-k} = 13390,2$.

A partir de los elementos de la diagonal de la matriz de varianzas-covarianzas estimada se obtiene $S_{\hat{\beta}_2} = 26,78$ y $S_{\hat{\beta}_3} = 187,46$. Con esta información se obtienen los siguientes resultados:

H_0	$\hat{\beta}_j$	$S(\hat{\beta}_j)$	$d(\hat{\beta}_j)/H_0 = \frac{\hat{\beta}_j}{S(\hat{\beta}_j)}$	$p = P(t_{27} > d^*)$	Conclusión
$\beta_2 = 0$	40,762	5,175	7,876	0,000	Significativo %
$\beta_3 = 0$	2,112	13,691	0,154	0,879	No significativo

c) Se especifica un modelo que distingue entre población femenina (X_2^F) y masculina (X_2^M) obteniéndose las siguientes estimaciones:

$$\begin{aligned}\hat{Y} &= 14 + 42X_2^F + 56X_2^M + 2,4X_3 & ; \quad \hat{\mathbf{u}}'\hat{\mathbf{u}} &= 350000 \\ \hat{Y} &= 15 + 50(X_2^F + X_2^M) + 2,2X_3 & ; \quad \hat{\mathbf{u}}'_R\hat{\mathbf{u}}_R &= 350000\end{aligned}$$

Justificar cuál es la restricción impuesta en el segundo modelo y si, con la información estadística disponible, debería ser o no rechazado este supuesto.

La restricción impuesta en el modelo es la igualdad del efecto marginal de la población masculina y femenina. Las hipótesis del contraste son:

$$\begin{aligned}H_0 : \beta_2^M &= \beta_2^F \\ H_1 : \beta_2^M &\neq \beta_2^F\end{aligned}$$

Se construye la discrepancia F que viene dada por la siguiente expresión:

$$\left(\frac{\hat{\mathbf{u}}'_R\hat{\mathbf{u}}_R - \hat{\mathbf{u}}'\hat{\mathbf{u}}}{\hat{\mathbf{u}}'\hat{\mathbf{u}}} \right) \left(\frac{n-k}{r} \right) \approx F_{n-k}^r$$

Con la información muestral disponible se obtiene:

$$F^* = \left(\frac{430000 - 350000}{350000} \right) \left(\frac{27}{1} \right) = 6,17$$

que lleva asociado un nivel crítico $p = P(F_{27}^1 > 6,17) = 0,02$ que conduce al rechazo de la hipótesis nula, es decir, de las restricciones incluidas en el modelo.

Esta restricción también puede contrastarse mediante el test de Wald que se basa en la expresión $W = \frac{\hat{\mathbf{u}}^*\hat{\mathbf{u}}^* - \hat{\mathbf{u}}'\hat{\mathbf{u}}}{\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n}} \approx \chi_1^2$.

PRACTICA 3.4- Se han analizado las ventas de una empresa (Y) junto a la evolución del IPC (X_2) y del índice de precios del sector en el que opera (X_3) estimándose los siguientes modelos a partir de una muestra de 50 datos:

Mínimos Cuadrados Ordinarios	$\hat{Y} = -4,85 + 1,53X_2 + 1,04X_3$	$\hat{\mathbf{u}}'\hat{\mathbf{u}} = 10008$
Mínimos Cuadrados Restringidos	$\hat{Y} = 52,59 + 0,52(X_2 + 2X_3)$	$\hat{\mathbf{u}}'_R\hat{\mathbf{u}}_R = 10390$

Describir cómo se llevará a cabo en este caso el contraste de restricciones de parámetros: hipótesis, desarrollo y conclusión.

SOLUCIÓN:

La hipótesis de este contraste vendría dada por:

$$\begin{aligned}H_0 : 2\beta_2 &= \beta_3 \\ H_1 : 2\beta_2 &\neq \beta_3\end{aligned}$$

Es decir, el supuesto de partida es que el efecto marginal que tienen sobre las ventas los precios del sector en el que opera la empresa es el doble del efecto marginal del IPC.

Para realizar el contraste se comparan los residuos del modelo sometido a la restricción con los del modelo estimado sin restricciones, mediante la expresión $\left(\frac{\hat{\mathbf{u}}_{\mathbf{R}}' \hat{\mathbf{u}}_{\mathbf{R}} - \hat{\mathbf{u}}' \hat{\mathbf{u}}}{\hat{\mathbf{u}}' \hat{\mathbf{u}}} \right) \left(\frac{n-k}{r} \right) \approx F_{n-k}^r$. La discrepancia muestral toma un valor de 1,793 que lleva asociado un nivel crítico elevado $p = P(F_{47}^1 \geq 1,793) = 0,187$.

Así pues, sería aconsejable utilizar el modelo restringido para efectuar la previsión de ventas de la empresa ¹

¹A la misma conclusión se llegaría mediante el test de Wald que se basa en la expresión $W = \frac{\hat{\mathbf{u}}^{*'} \hat{\mathbf{u}}^* - \hat{\mathbf{u}}' \hat{\mathbf{u}}}{\frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{n}} \approx \chi_1^2$ cuyo resultado en nuestro caso es $W = \frac{10390 - 10008}{\frac{10008}{50}} = 1,908$ con un nivel crítico $p = P(\chi_1^2 \geq 1,908) \geq 0,1$ que nos llevaría a no rechazar el supuesto planteado (en general se cumple para tamaños elevados de muestra $W = rF$ y en este caso $r = 1$).