Prácticas Tema 2: El modelo lineal simple

Ana J. López y Rigoberto Pérez

Departamento de Economía Aplicada. Universidad de Oviedo

PRACTICA 2.1- Se han analizado sobre una muestra de 10 familias las variables renta familiar disponible (X) y gasto en ocio (Y), ambas expresadas en miles de euros, obteniendo los siguientes resultados:

Familia	Renta (X)	Gasto (Y)
1	300	20
2	400	40
3	500	80
4	600	120
5	600	90
6	450	100
7	250	20
8	350	60
9	450	90
10	500	120

a) Estimar un modelo lineal que explique los gastos a partir de las rentas familiares ¿Cual sería la elasticidad media?

b) Obtener la distribución de los errores de estimación

c) Analizar la bondad del modelo estimado

d) Llevar a cabo el contraste básico de significación sobre este modelo

e) ¿Cómo se vería afectado el modelo anterior si las variables apareciesen expresadas en dolares? ¿Y si la renta de todas las familias se incrementa en 2000 euros?

SOLUCION:

a) Estimar un modelo lineal que explique los gastos a partir de las rentas familiares ¿Cual sería la elasticidad media?

A partir de la información disponible de las 10 familias de la muestra se obtiene la tabla siguiente

Familia	X	Y	X^2	Y^2	XY
1	300	20	90000	400	6000
2	400	40	160000	1600	16000
3	500	80	250000	6400	40000
4	600	120	360000	14400	72000
5	600	90	360000	8100	54000
6	450	100	202500	10000	45000
7	250	20	62500	400	5000
8	350	60	122500	3600	21000
9	450	90	202500	8100	40500
10	500	120	250000	14400	60000
Sumas	4400	740	2060000	67400	359500
Medias	440	74	206000	6740	35950

a partir de la cual se obtiene:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = 440$$

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} = 74$$

$$S_X^2 = \frac{\sum_{i=1}^{n} X_i^2}{n} - \bar{X}^2 = 206000 - 440^2 = 12400$$

$$S_Y^2 = \frac{\sum_{i=1}^{n} Y_i^2}{n} - \bar{Y}^2 = 6740 - 440^2 = 1264$$

$$S_{XY} = \frac{\sum_{i=1}^{n} X_i Y_i}{n} - \bar{X}\bar{Y} = 35950 - 32560 = 3390$$

pudiendo estimar así los coeficientes mínimo cuadráticos

$$\hat{\beta}_2 = \frac{S_{XY}}{S_X^2} = \frac{3390}{12400} = 0,27$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = -46,29$$

con lo cual el modelo estimado para el gasto en ocio es:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

indicando que por cada mil euros adicionales de renta disponible las familias dedican 270 euros a gasto en ocio.

Por su parte, la elasticidad media se obtendría mediante la expresión:

$$E_{\bar{Y},\bar{X}} = \frac{\delta Y}{\delta X}\frac{\bar{X}}{\bar{Y}} = 0,27\frac{440}{74} = 1,625$$

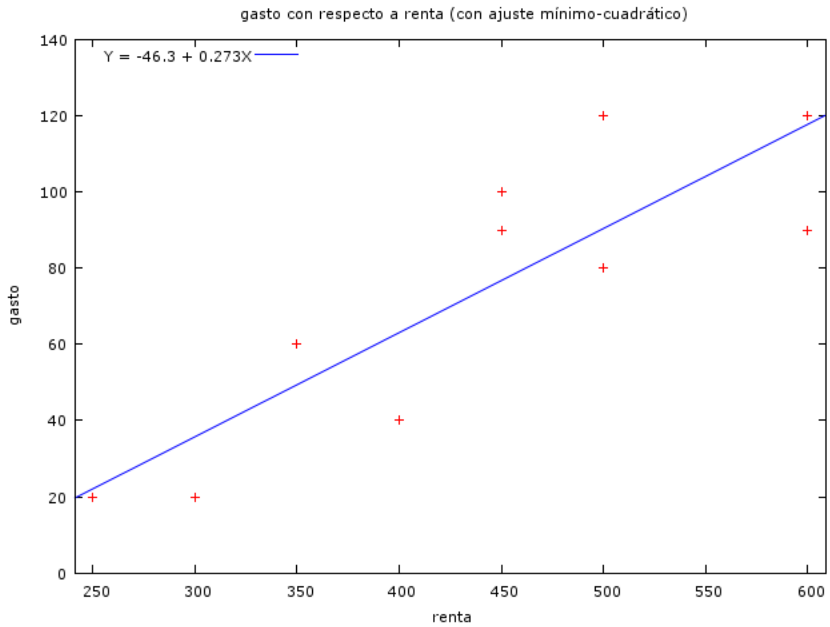
resultado que permite afirmar que, si partiendo del centro de gravedad $(\bar{X}, \bar{Y})$la renta familiar disponible se incrementa un 1 %, la respuesta en el gasto en ocio sería un incremento de 1,625 %.

b) Obtener la distribución de los errores de estimación

A partir del modelo mínimo cuadrático anterior se obtienen los valores estimados y los correspondientes errores de estimación, que aparecen resumidos en la tabla siguiente:

Familia	X	Y	$\hat{Y}$	$\hat{u} = Y - \hat{Y}$
1	300	20	35,73	-15,73
2	400	40	63,06	-23,06
3	500	80	90,4	-10,4
4	600	120	117,74	2,26
5	600	90	117,74	-27,74
6	450	100	76,73	23,27
7	250	20	22,06	-2,06
8	350	60	49,4	10,6
9	450	90	76,73	13,27
10	500	120	90,4	29,6
Sumas	4400	740	740	0
Medias	440	74	74	0

cuya representación gráfica se recoge a continuación:



c) Analizar la bondad del modelo estimado

Para analizar la bondad del modelo calculamos la Variación total y sus componentes explicada y residual:

$$VT = \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2 = 12640$$

$$VE = \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2 = 9267,82$$

$$VNE = \sum_{i=1}^{n} \hat{u}_i^2 = 3372,18$$

que también podría ser calculado como $VNE = VT - VE$

A partir de estos resultados el coeficiente de determinación se obtendría como:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \hat{u}_i^2}{\sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2} = 0,7332$$

Es decir, algo más del 73 % de las variaciones del gasto en ocio se explican mediante las rentas familiares.

Dicho resultado podría también obtenerse utilizando la igualdad $R^2 = r_{XY}^2$ que se cumple por tratarse de un modelo lineal simple.

d) Llevar a cabo el contraste básico de significación sobre este modelo

La hipótesis que contrastamos es la nulidad del coeficiente de la renta, es decir:

$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$

Para llevar a cabo este contraste debemos obtener el valor muestral de la expresión t de student $\frac{\hat{\beta}_2}{S_{\hat{\beta}_2}}$, cuyo denominador se obtiene a partir de la expresión:

$$S_{\hat{\beta}_2}^2 = \frac{S^2}{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2} = \frac{421,522}{124000} = 0,0034$$

donde hemos tenido en cuenta que la varianza residual es $S^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-2} = 421,522$

A partir de los resultados anteriores se obtiene $\frac{\hat{\beta}_2}{S_{\hat{\beta}_2}} = \frac{0,27}{0,0583} = 4,69$, que proporciona el nivel crítico:

$$p = (|t_{n-2}| > 4,69) = 0,0016$$

resultado que es significativo para rechazar la nulidad del coeficiente y por tanto permite concluir que la renta familiar es una variable significativa para explicar el gasto en ocio.

e) ¿Cómo se vería afectado el modelo anterior si las variables apareciesen expresadas en dolares? ¿Y si la renta de todas las familias se incrementa en 2000 euros?

Para analizar el efecto de un cambio de unidad monetaria en las variables se tendría que utilizar el tipo de cambio del dolar respecto al euro que genéricamente denominamos c, con lo cual pasaríamos a estimar un modelo sobre las variables renta y gasto expresadas en dólares: $X^* = cX$ e $Y^* = cY$.

El nuevo modelo estimado sería entonces $\hat{Y}_i^* = \hat{\beta}_1 + \hat{\beta}_2 X_i^*$y basta tener presentes las propiedades de la media, la varianza y la covarianza para obtener:

$$\hat{\beta}_2{}^* = \frac{S_{X^*Y^*}}{S_{X^*}^2} = \frac{c^2 S_{XY}}{c^2 S_X^2} = \hat{\beta}_2 = 0,27$$

$$\hat{\beta}_1{}^* = \bar{Y}^* - \hat{\beta}_2 \bar{X}^* = c\bar{Y} - \hat{\beta}_2 c\bar{X} = c\hat{\beta}_1$$

Así pues, el cambio de euros a dólares (cambio de escala) no altera el efecto marginal de la renta sobre los gastos en ocio, pero en cambio sí altera la estimación de los gastos fijos (que sufrirían el mismo cambio de escala)

De modo similar, si la renta de todas las familias aumentase en 2000 euros nos encontraríamos ante un cambio de origen $X^* = X + 2$ con lo cual se tendría en este caso:

$$\hat{\beta}_2{}^* = \frac{S_{X^*Y}}{S^2_{X^*}} = \frac{S_{XY}}{S^2_X} = \hat{\beta}_2 = 0,27$$

$$\hat{\beta}_1{}^* = \bar{Y} - \hat{\beta}_2 \bar{X}^* = \bar{Y} - \hat{\beta}_2 \left(\bar{X} + 2\right) = \hat{\beta}_1 - 2\hat{\beta}_2 = -46,837$$

PRACTICA 2.2- Se ha recopilado información sobre los tiempos de vuelo (X, en horas) y los consumos de combustible (Y, en miles de libras) para 32 trayectos distintos de un avión:

$$\sum_{i=1}^{n} Y_i = 439,4 \qquad \sum_{i=1}^{n} X_i = 188 \qquad \sum_{i=1}^{n} X_i^2 = 1839$$
$$\sum_{i=1}^{n} Y_i^2 = 9587,8 \qquad \sum_{i=1}^{n} X_i Y_i = 4192,4$$

a) Estimar un modelo lineal mínimo cuadrático para explicar el consumo de combustible

b) Llevar a cabo un análisis de la varianza de Y interpretando los resultados

c) Construir intervalos de confianza al 95 % para los parámetros de regresión

d) Llevar a cabo el contraste de significación del modelo

e) ¿Qué consumo total se estimaría para un programa de 100 vuelos, de los que 20 son de una hora y 80 son de dos horas?

f) ¿Cómo cambiarían los resultados anteriores si gracias a una innovación en el aparato se reduce un 15 % el consumo de todos los vuelos?

SOLUCION: **a)** Estimar un modelo lineal mínimo cuadrático para explicar el consumo de combustible

A partir de la información disponible se tiene:

$$n = 32 \ , \ \bar{Y} = 13,7313 \ , \ \bar{X} = 5,875$$
$$\sum_{i=1}^{n} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) = \sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y} = 1610,925$$
$$\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 = 734,5$$
$$\sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2 = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2 = 3554,2888$$

y con esta información se obtienen los EMC:

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} = 2,1932$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 0,846$$

a partir de los cuales podemos afirmar que cada hora adicional de vuelo conlleva un incremento de 2193,2 libras de combustible mientras el consumo fijo de combustible (asociado al calentamiento, arranque de motores, ... antes de iniciarse el vuelo) se estima en 846 libras.

b) Llevar a cabo un análisis de la varianza de Y interpretando los resultados

Podemos plantear la tabla de variaciones que sigue:

Variación total (VT):	$\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2 = $ $= \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2 = 3554,2888$
Variación explicada (VE):	$\sum_{i=1}^{n}\left(\hat{Y}_i-\bar{Y}\right)^2 = \hat{\beta}_2^2 \sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 =$ $= (2,1932)^2(734,5) = 3533,1237$
Variación no explicada (VNE):	$\sum_{i=1}^{n} \hat{u}_i^2 = 21,1651$

resultados a partir de los que se llega a un coeficiente de determinación del 99,4 %

c) Construir intervalos de confianza al 95 % para los parámetros de regresión

Para construir los intervalos de confianza al 95 % se plantea la necesidad de estimar las varianzas de los estimadores:

$$S^2 = \frac{\sum_{i=1}^{n}\hat{u}_i^2}{n-2} = \frac{021,1651}{30} = 0,7055$$

$$S^2\left(\hat{\beta}_2\right) = \frac{S^2}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2} = 0,0010$$

$$S^2\left(\hat{\beta}_1\right) = \frac{S^2\sum_{i=1}^{n}X_i^2}{n\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2} = 0,0552$$

llegando posteriormente a los intervalos:

$$\begin{aligned}\left[\hat{\beta}_2 - kS\left(\hat{\beta}_2\right), \; ; \; \beta_2 + kS\left(\hat{\beta}_2\right)\right] &= \left[2,1932 - 2,0422\sqrt{0,001} \; ; \; 2,1932 + 2,0422\sqrt{0,001}\right] \\ &= [2,1299; \; 2,2565]\end{aligned}$$

$$\begin{aligned}\left[\hat{\beta}_1 - kS\left(\hat{\beta}_1\right) \; ; \; \hat{\beta}_1 + kS\left(\hat{\beta}_1\right)\right] &= \left[0,846 - 2,0422\sqrt{0,0552} \; ; \; 0,846 + 2,0422\sqrt{0,0552}\right] \\ &= [0,366 \; ; \; 1,3258]\end{aligned}$$

d) Llevar a cabo el contraste de significación del modelo

Para analizar si el modelo puede ser considerado significativo podemos aplicar el contraste de la t de student

$$H_0 : \beta_2 = 0 \; ; \; H_1 : \beta_2 \neq 0$$

para el que se obtiene la discrepancia:

$$d_{\hat{\beta}}/H_o = \frac{\hat{\beta}_2 - 0}{S_{\hat{\beta}}} = \frac{2,1932}{\sqrt{0,001}} = 70,76 \;\; ; \;\; p = P\left(|t_{30}| > 70,76\right) = 0$$

y el resultado es significativo para rechazar la hipótesis.

Otra posibilidad sería aplicar el contraste de la F de Snedecor, que conduce a la misma conclusión que el anterior:

$$F_{30}^{1} = \frac{\hat{\beta}_2^2 \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{S^2} = \frac{2,1932^2(734,5)}{0,7055} = 5007,96 \;\; ; \;\; p = P\left(F_{30}^{1} > 5007,96\right) = 0$$

e) ¿Qué consumo total se estimaría para un programa de 100 vuelos, de los que 20 son de una hora y 80 son de dos horas?

$$\hat{C}_{Total} = 20\hat{Y}_{X=1} + 80\hat{Y}_{X=2} = 20(3,04) + 80(5,23) = 479,39$$

Conviene tener presente que este resultado no coincide con el que se obtendría sumando el total de horas de vuelo (180) y sustituyendo este valor en la recta estimada, ya que esta opción sería equivalente a considerar un vuelo de 180 horas de duración y un único gasto fijo (en cuyo caso el consumo previsto sería notablemente inferior: 395,63 miles de libras).

f) ¿Cómo cambiarían los resultados anteriores si gracias a una innovación en el aparato se reduce un 15 % el consumo de todos los vuelos?

En este caso se produciría un cambio de escala sobre la variable consumo de modo que $Y^* = 0,85Y$. Como consecuencia, puede comprobarse que los estimadores del modelo se verían afectados en la misma proporción:

$$\hat{\beta}_2^{\,*} = \frac{S_{XY^*}}{S_X^2} = \frac{0,85S_{XY}}{S_X^2} = 0,85\hat{\beta}_2 = 1,8642$$

$$\hat{\beta}_1^{\,*} = \bar{Y}^* - \hat{\beta}_2^{\,*}\bar{X} = 0,85\bar{Y} - 0,85\hat{\beta}_2\bar{X} = 0,85\hat{\beta}_1 = 0,7191$$

Como consecuencia las predicciones de consumo también se verían reducidas un 15 % pero en cambio no se verían afectadas las medidas de bondad ni los contrastes de significación.

PRACTICA 2.3- Se ha estimado un modelo simple para la demanda de teléfonos móviles (Q, en miles de unidades) en función de su precio medio (P, en euros). A partir de una muestra de 22 observaciones se ha obtenido:

$$\begin{array}{ll} \hat{\beta}_1 = 63,84 & \hat{\beta}_2 = -0,08 \\ S\left(\hat{\beta}_1\right) = 0,7 & S\left(\hat{\beta}_2\right) = 0,008 \\ r_{Q,P} = -0,91 & \end{array}$$

a) Con la información disponible ¿sería posible defender que la demanda de teléfonos móviles es inelástica respecto al precio?

b) Obtener e interpretar el coeficiente de determinación del modelo estimado

Solución: **a)** Contrastar la inelasticidad de la demanda respecto al precio sería equivalente a llevar a cabo un contraste de significación mediante la t de student:

$$H_0 : \beta_2 = 0 \; ; \; H_1 : \beta_2 \neq 0$$

Obteniéndose en este caso bajo la hipótesis nula la discrepancia:

$$d_{\hat{\beta}}/H_o = \frac{\hat{\beta}_2 - 0}{S_{\hat{\beta}}} = \frac{-0,08}{0,008} = 10 \ ; \ p = P\left(|t_{20}| \text{»} 10\right) = 0$$

Obteniéndose un resultado significativo para rechazar la hipótesis de inelasticidad de la demanda.

b) Para obtener el coeficiente de determinación en este caso bastaría tener en cuenta que, al tratarse de un modelo simple dicho coeficiente coincide con el cuadrado del coeficiente de correlación lineal, por lo que se tiene:

$$R^2 = \left(r_{Q,P}\right)^2 = 0,8281$$

Es decir, el modelo estimado explica más del 82 % de los cambios en la demanda de teléfonos móviles a partir de sus precios.