

ECONOMETRIA I - CURSO 2015

PRACTICO 2

Modelo de Regresión Lineal Clásico – Estimación e Inferencia

EJERCICIO 1: *Modelo de Regresión Lineal con Variable Dependiente Binaria*

Suponga quiere explicar la voluntad de pago (Y_i) de cada uno de los deudores de un banco de plaza, en función de determinadas características “clave” observadas para cada uno de ellos (contenidas en el vector x_i). Los individuos se eligen en forma independiente. Para ello, considera un modelo cuya ecuación es:

$$y_i = x_i' \beta + \varepsilon_i, \forall i = 1, 2, \dots, N, \text{ donde } x_i = \begin{bmatrix} 1 \\ x_2 \\ \dots \\ x_k \end{bmatrix} \text{ y } \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{bmatrix} \text{ y cada variable } Y_i \text{ (toma}$$

valores 1 (si es buen pagador) ó 0 (mal pagador) con probabilidad p_i y $(1 - p_i)$ respectivamente.
($i = 1, 2 \dots N$ es el índice correspondiente a los deudores)

Se pide:

- Obtenga la distribución de ε_i , su esperanza y su varianza.
- Como la ecuación es lineal en los parámetros, alguien sugiere estimar el vector β por Mínimos Cuadrados Ordinarios. ¿En este modelo, el β_{MCO} será el BLUE? ¿Coincidirán β_{MCO} y β_{MV} ? ¿Qué supuesto(s) en los que se basan las propiedades de las estimaciones MCO ya no se cumplen?

EJERCICIO 2: (Mayo 2013, modificación de un ejercicio del Johnston, pág. 123).

Considere el siguiente modelo de regresión (las variables fueron centradas respecto de su media):

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t$$

Una vez estimado el mismo en base a información disponible, se conocen los siguientes resultados:

$$\begin{array}{llll} T = 100 & \sum y_t^2 = 493/3 & \sum x_{1t}^2 = 30 & \sum x_{2t}^2 = 3 \\ & \sum x_{1t} y_t = 30 & \sum x_{2t} y_t = 20 & \sum x_{1t} x_{2t} = 0 \end{array}$$

- Calcular las estimaciones MCO de los parámetros del modelo.
- Hallar el valor del coeficiente de determinación R^2 .
- Estimar la varianza de los errores por MCO.
- Analizar la significación global del modelo.
- ¿Se puede decir que estadísticamente β_2 es menor que 7? Justifique mediante un contraste adecuado.

EJERCICIO 3: (enero 2013)

Se considera la siguiente función de producción Cobb-Douglas:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \text{ (las variables están consideradas en logaritmos).}$$

Y: es el logaritmo de la cantidad producida

X₁: logaritmo del factor trabajo

X₂: logaritmo del factor capital

Se obtiene una muestra de 23 empresas de un mismo sector, a partir de la cual se obtienen los siguientes resultados:

$$X'X = \begin{bmatrix} 12 & 8 \\ 8 & 12 \end{bmatrix} \quad X'y = \begin{bmatrix} 10 \\ 8 \end{bmatrix} \quad y'y = 10$$

(Observe que se trata de datos centrados respecto de la media).

- Estime los coeficientes de los dos factores de producción.
- Previo cálculo de la suma de cuadrados explicados, calcule el coeficiente de determinación del modelo.
- Analice la significación de ambos coeficientes y la significación global del modelo (utilice un nivel $\alpha = 0,05$).
- ¿Hay rendimientos constantes de escala? Justifique mediante el contraste adecuado (utilice un nivel $\alpha = 0,05$).

EJERCICIO 4:

Sea el siguiente modelo que relaciona el gasto en educación (E_i) con la renta disponible (R_i):

$$E_i = \beta_1 + \beta_2 R_i + \varepsilon_i$$

De la información obtenida de una muestra de 10 familias se han obtenido los siguientes resultados:

$$\bar{E} = 7 \quad \bar{R} = 50 \quad \sum_{i=1}^{10} R_i^2 = 30650 \quad \sum_{i=1}^{10} E_i^2 = 622 \quad \sum_{i=1}^{10} R_i E_i = 4345$$

Se pide:

- Obtenga una estimación de β_1 y β_2 .
- Estime la elasticidad gasto en educación-renta para el promedio de las familias de la muestra.
- Descomponga la varianza total del gasto en educación de la muestra en varianza explicada y varianza residual.
- Calcule el coeficiente de determinación.
- Estime la varianza de las perturbaciones
- Contraste si la renta disponible tiene o no una influencia significativa sobre el gasto en educación.

EJERCICIO 5:

Para un estudio sobre la demanda de alimentos se especifica el siguiente modelo:

$$Cons_t = \alpha_0 + \alpha_1 YD_t + \alpha_2 PRECIO_t + \varepsilon_t$$

Se dispone de 200 observaciones de las siguientes variables:

$Cons_t$ = Gasto de consumo en alimentos.

YD_t = Ingreso disponible en términos reales.

$PRECIO_t$ = Índice de precios de los alimentos.

ε_t = Término de perturbación, normalmente distribuido con media cero y varianza constante.

Medias y desvios de las variables:

	Cons	YD	PRECIO
Mean	19.34425	5.377356	12.38180
Std. Dev.	6.416843	2.353914	4.202572

Covarianzas muestrales de las variables:

	Cons	YD	PRECIO
Cons	40.97000	12.82490	23.26235
YD		5.513205	5.110159
PRECIO			17.57330

Matriz de varianzas y covarianzas de los estimadores:

	Cons	YD	PRECIO
Cons	0.056367	-0.002367	-0.003091
YD		0.001332	-0.000387
PRECIO			0.000418

Se pide:

1. Estime los coeficientes del modelo sabiendo que la Suma de Cuadrados Residuales es 211.3503.
2. Realice las pruebas de significación individuales para cada coeficiente e indique explícitamente los respectivos intervalos de confianza para todo coeficiente que resulte estadísticamente significativo.
3. ¿Se podría afirmar que la influencia del ingreso en la demanda es el doble de la influencia de los precios? Justifique.

EJERCICIO 6:

De un estudio sobre el consumo de naftas en EEUU, para el período 1960–1986, se obtuvo información sobre:

G: gasto de consumo total de naftas a valores constantes,

IPOIL: índice de precio de las naftas,

IGRDISP: ingreso real disponible per cápita,

IPNCAR: índice de precio de los autos nuevos,

IPUCAR: índice de precios de los autos usados,

IPTPUB: índice de precios del transporte público.

Se presentan a continuación los resultados de dos regresiones múltiples de *G* sobre las otras variables explicativas, utilizando el programa E-VIEWS. Se supuso que las perturbaciones seguían los supuestos clásicos:

Primera regresión:

Variable	Coeficient	Std. Error	t-Statistic	Prob.
C	-101.8148	18.03465	-5.645506	0.0000
IPOIL	-14.94979	3.195233	-4.678780	0.0001
IGRDISP	0.039956	0.002656	15.04225	0.0000
R-squared	0.944302	Mean dependent var		207.0200
Adjusted R-squared	0.939458	S.D. dependent var		44.66605
S.E. of regression	10.99018	Akaike info criterion		7.740048
Sum squared resid	2778.034	Schwarz criterion		7.885213

Log likelihood	-97.62063	F-statistic	194.9692
Durbin-Watson stat	1.361141	Prob(F-statistic)	0.000000

Segunda regresión:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-84.11627	18.35777	-4.582051	0.0002
IPOIL	-24.74372	6.026845	-4.105585	0.0005
IGRDISP	0.037249	0.002802	13.29161	0.0000
IPNCAR	0.001222	0.005640	0.216586	0.8307
IPUCAR	-32.54306	14.43326	-2.254727	0.0355
IPTPUB	43.26793	16.94567	2.553332	0.0189
R-squared	0.958646	Mean dependent var	207.0200	
Adjusted R-squared	0.948308	S.D. dependent var	44.66605	
S.E. of regression	10.15521	Akaike info criterion	7.673024	
Sum squared resid	2062.564	Schwarz criterion	7.963354	
Log likelihood	-93.74932	F-statistic	92.72696	
Durbin-Watson stat	1.302286	Prob(F-statistic)	0.000000	

Se pide:

1. Con las salidas proporcionadas, recalculer el R^2 y el R^2 ajustado, la estimación de σ^2 y las varianzas estimadas de los coeficientes, ¿Es posible calcular los términos de la diagonal de la matriz $(X'X)^{-1}$?
2. Probar la significación de los coeficientes a un nivel de significación del 0.05.
3. Construir intervalos de confianza al 95% para los parámetros estimados.
4. Interpretar económicamente los coeficientes de cada una de las ecuaciones.

Aplicación en *Gretl*:

Considere el archivo de datos *naftasGreene.gdt*. Se trata de información referente al período 1960 – 1995 de los USA, referente a las variables consideradas en el ejercicio.

- a) Restrinja las observaciones al período 1961 – 1986.
- b) Realice la primera regresión (consumo de naftas en función del índice de precio de gasolinas y del ingreso real disponible per cápita).
- c) Indique el valor del R^2 y del R^2 corregido, el valor de la estimación de la varianza de los errores y la diagonal de la matriz $(X'X)^{-1}$.
- d) Analice la significación de los coeficientes para un $\alpha = 0.05$.
- e) Construya intervalos de confianza para los coeficientes.
- f) Interprete económicamente los coeficientes estimados.

EJERCICIO 7:

Dada la siguiente muestra:

T	1	2	3	4	5	6	7	8
Y_t	19	16	20	22	18	15	12	13
X_{1t}	4	5	7	7	9	10	11	12
X_{2t}	2	2	8	8	6	6	4	4

Se pide:

1. Estime la regresión: $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \varepsilon_t$
2. Estimar la matriz de varianzas y covarianzas de los estimadores, suponiendo que las ε_t (término de perturbación) son normales independientes e idénticamente distribuidas, $n(0; \sigma^2)$.
3. Calcule el R^2 y R^2 ajustado.

4. Pruebe la significación de cada uno de los parámetros y del modelo en su conjunto al 5% de nivel de significación.
5. Probar la hipótesis nula $H_0) \beta_1 + \beta_2 = 0$ al 5% de nivel de significación.

Aplicación en Gretl:

Construya un archivo de datos *ejercicio7.gdt* con los datos del cuadro.

- a) Estime la regresión definida en la parte 1.
- b) Estime la varianza de los errores.
- c) Escriba la matriz de varianzas y covarianzas estimada de los estimadores.
- d) Calcule el R^2 y el R^2 ajustado.
- e) Analice la significación de cada uno de los parámetros y del modelo en su conjunto a un nivel del 5%.
- f) Contraste la $H_0) \beta_1 + \beta_2 = 0$ al 5% de significación.

Fac. CC. Económicas y de Administración - Univ. de la República