Universidad de la República, Facultad de Ciencias Económicas y Administración.

ECONOMETRIA I - CURSO 2015

PRACTICO 4

Modelo de Regresión Lineal Clásico – Predicción y Multicolinealidad

EJERCICIO 1:

Dado el siguiente modelo estimado con 10 observaciones: $\hat{Y}_t = -0{,}06 + 1{,}44\, X_{2t} - 0{,}48\, X_{3t}$, donde se asume el cumplimiento de los supuestos clásicos del MRLG, y se conocen:

$$X'X = \begin{bmatrix} 10 & 8 & 11 \\ & 598 & 791 \\ & & 1128 \end{bmatrix},\ (X'X)^{-1} = \begin{bmatrix} 0{,}1011 & -0.0007 & -0{,}0005 \\ & 0{,}0231 & -0{,}0162 \\ & & 0{,}0122 \end{bmatrix},\ X'y = \begin{bmatrix} 6 \\ 506 \\ 632 \end{bmatrix} \text{ y } y'y = 444$$

Se pide:

- **a)** contraste la hipótesis nula $\beta_2 + 2\beta_3 = 1$
- **b)** dado el valor $Y_0 = 1$ (predicción para $t = 0$), verifique si puede haber sido generado por el modelo estimado (se sabe que $c(1) = 1$ y que $c(2) = 2$, es decir: $c' = [1 \quad 1 \quad 2]$).

EJERCICIO 2:

Un analista desea hacer una primera aproximación a los determinantes de la inversión. Para ello dispone de información para 168 países en el año 2010, y cuenta con las siguientes variables:

tinv % de Inversión sobre PBI (en paridad de poderes de compra)
cln Indicador del clima de negocios del país
aper Tasa de apertura del país [(X+M)/PBI]

Se resuelve estimar por MCO el siguiente modelo: $tinv_i = \beta_1 + \beta_2 cln_i + \beta_3 aper_i + \varepsilon_i$

Tras hacer las operaciones correspondientes se obtienen los siguientes productos cruzados:

	cln	**aper**	**tinv**
cln	183,67	122,54	3369,53
aper		178,42	3629,48
tinv			102398,5

También se dispone de las medias de las variables:

$\overline{cln} = 0{,}93$, $\overline{aper} = 0{,}891$ y $\overline{tinv} = 23{,}145$

Se sabe también que: $SCRes$ = 11.590,00 y SCT = 12.402,4.

Nota: Trabaje con redondeos de 2 decimales, excepto si precisa obtener la matriz $(X'X)^{-1}$ en la que se sugiere que redondee a 4 decimales.

Se pide:

1. Plantee el modelo centrado y obtenga el vector de coeficientes estimados en dicho modelo (recuerde que $\sum_{i=1}^{i=n}(X_i - \overline{X})(Y_i - \overline{Y}) = \sum_{i=1}^{i=n} X_i Y_i - n\overline{X}\,\overline{Y}$).
2. Calcule la estimación de las varianzas de los coeficientes estimados en el punto anterior.
3. Se desea saber si las variables *cln* y *aper* tienen efectos de igual magnitud pero de signo opuesto sobre la tasa de inversión. Realice la prueba de hipótesis con el estadístico F, plantee el estadístico de prueba (detallando todos sus elementos), su distribución, la región crítica, el criterio de decisión y concluya.
4. Si se informa que Uruguay no fue incluido en la muestra de países utilizada, y se sabe que los valores no centrados de Uruguay en las variables consideradas son: cln_{uy} = 1,22 y $aper_{uy}$ = 0,52. Plantee la predicción puntual para la tasa de inversión de Uruguay (Sugerencia: obtenga previamente la estimación de la constante del modelo). Plantee la forma que tendría un intervalo de aleatorio al 90% de confianza para la predicción (no es necesario calcular el intervalo de confianza).

EJERCICIO 3:

Se dispone de datos anuales del consumo total de gasolina en USA (variable *G*) que corresponden al período 1960– 1995.[1] Para explicar dicho consumo se consideraron las variables:

- *Y*: renta disponible (en dólares)
- *Pg*: índice de precio de la gasolina
- *Pnc*: índice de precio de los autos nuevos
- Pd: índice agregado de precios de bienes de consumo durables
- Pnd: índice agregado de precios de bienes de consumo no durables
- Ps: índice agregado de precios por consumo de servicios
- *Pop*: población total de US en millones.

Se construyen dos nuevas variables, que son el consumo de gasolina per cápita (*Gp*) y el ingreso disponible per cápita (*Yp*).

En una primera instancia, se aplican técnicas de MCO a la ecuación:

$$\ln Gp_t = l\_G_t = \beta_1 + \beta_2 Yp_t + \beta_3 \mathrm{Pg}_t + \beta_4 Pnc_t + \beta_5 Pd_t + \beta_6 Pnd_t + \beta_7 Ps_t + \varepsilon_t$$

Los resultados de dicha estimación son los siguientes (salida del programa *Gretl*):

Modelo 1: MCO, usando las observaciones 1960-1995 (T = 36)
Variable dependiente: l_G

	Coeficiente	*Desv. Típica*	*Estadístico t*	*Valor p*
const	3,1263	0,181902	17,1867	<0,00001
Yp	0,0548499	0,00546175	10,0426	<0,00001
Pg	-0,0818418	0,0344337	-2,3768	0,02428
Pnc	-0,37738	0,181718	-2,0767	0,04679
Pd	0,52238	0,413857	1,2622	0,21693
Pn	0,430281	0,432213	0,9955	0,32771
Ps	0,0873854	0,249616	0,3501	0,72881

[1] Datos del libro "*Econometric Analysis*" – Greene, William.

Media de la vble. dep.	5,392989	D.T. de la vble. dep.	0,248779
Suma de cuad. residuos	0,031732	D.T. de la regresión	0,033079
R-cuadrado	0,985351	R-cuadrado corregido	0,982321
F(6, 29)	325,1178	Valor p (de F)	3,15e-25

Coeficientes de correlación muestrales, usando las observaciones 1960 - 1995

Yp	***Pg***	***Pnc***	***Pd***	***Pn***	***Ps***	
1,0000	0,7957	0,8122	0,8610	0,8522	0,8160	***Yp***
	1,0000	0,9242	0,9551	0,9431	0,8984	***Pg***
		1,0000	0,9902	0,9936	0,9942	***Pnc***
			1,0000	0,9957	0,9800	***Pd***
				1,0000	0,9910	***Pn***
					1,0000	***Ps***

<u>Modelo 2</u>: MCO, usando las observaciones 1960-1995 (T = 36)
Variable dependiente: *l_G*

	Coeficiente	***Desv. Típica***	***Estadístico t***	***Valor p***
const	2,69003	0,0920939	29,2096	<0,00001
Pg	-0,0414146	0,0128206	-3,2303	0,00286
Pnc	0,0596165	0,0262941	2,2673	0,03026
Yp	0,0654875	0,0026845	24,3947	<0,00001

Media de la vble. dep.	5,392989	D.T. de la vble. dep.	0,248779
Suma de cuad. residuos	0,040341	D.T. de la regresión	0,035506
R-cuadrado	0,981377	R-cuadrado corregido	0,979631
F(3, 32)	562,1002	Valor p (de F)	9,58e-28

Coeficientes de correlación muestrales, usando las observaciones 1960 - 1995

Pg	**Pnc**	**Yp**	
1,0000	0,9242	0,7957	**Pg**
	1,0000	0,8122	**Pnc**
		1,0000	**Yp**

a. Analice la significación estadística de los coeficientes asociados a los índices de precio agregados del <u>modelo 1</u>.

b. Desarrolle una prueba de hipótesis que analice <u>simultáneamente</u> si los tres índices de precio agregados incluidos en el <u>modelo 1</u> son irrelevantes en la determinación del logaritmo de la demanda de gasolina.

c. Comente la siguiente afirmación: «Si las tres variables agregadas resultan no significativas para un mismo nivel de significación α en **b.** es lo mismo de que resulten no significativas para el mismo nivel α al realizar el contraste planteado en **a.**». Explique.

d. Considere el modelo 2. Escriba la ecuación asociada al mismo. ¿Se puede afirmar que el efecto sobre el logaritmo del consumo de gasolina es igual en el caso de las variables *Pnc* e *Yp*? Justifique mediante el contraste adecuado.

e. Interprete económicamente el valor del coeficiente estimado asociado a la variable *Pg* (índice de precio de la gasolina) en el modelo 2.

f. Calcule una predicción puntual del consumo de gasolina per cápita (Gp) para el año 1996, suponiendo que los valores para el índice de precios de la gasolina, el índice de precios de los autos nuevos y el ingreso disponible per cápita son 3,81 ; 2,82 ; 45,63, respectivamente (confianza del 95%).

EJERCICIO 4:

En el análisis de la oferta de determinado sector de la industria, se busca estimar su función de producción. Para ello se dispone de los datos de toneladas producidas (*Q*), la utilización del factor trabajo en horas (*L*) y la utilización del factor capital en horas-máquina (K) para las 15 firmas que operaron en el período.

Se asume que la función de producción es Cobb-Douglas, de modo que:

$$Q = \beta_0 L^{\beta_1} K^{\beta_2} \exp(\varepsilon)$$

Para obtener una expresión del modelo que sea lineal en β se plantea:

$$\ln Q = \beta_0 + \beta_1 \ln L + \beta_2 \ln K + \varepsilon$$

Los datos, una vez centrados, permiten obtener las siguientes matrices:

$$X'X = \begin{bmatrix} 0{,}0210648 & 0{,}1063517 \\ 0{,}1063517 & 0{,}5476632 \end{bmatrix}, \ (X'X)^{-1} = \begin{bmatrix} 3030{,}5 & -589{,}47 \\ -589{,}47 & 116{,}49 \end{bmatrix}, \ X'y = \begin{bmatrix} 0{,}03591628 \\ 0{,}18336458 \end{bmatrix}$$

De este modo, la estimación por MCO será (estadísticos *t* entre paréntesis):

$$\ln \hat{Q} = \underset{}{0{,}50} + \underset{(1{,}07)}{0{,}76 \ln L} + \underset{(1{,}36)}{0{,}19 \ln \hat{K}} \qquad R^2 = 0{,}969$$

Un investigador señala que los resultados reflejan existencia de multicolinealidad en los datos.

a) Defina conceptualmente qué es la multicolinealidad aproximada y explique desde el punto de vista económico por qué podría presentarse en este caso.

b) Explique qué elementos del resultado son indicio de multicolinealidad.

c) Muestre empíricamente que la relación existente entre los regresores aporta evidencia clara de multicolinealidad.

d) Dado el resultado anterior, ¿Qué podría esperarse respecto a varianza de los estimadores MCO? Explique utilizando el Factor de Incremento de Varianza.

EJERCICIO 5:

Se dispone de datos de demanda de viajes en ómnibus urbanos para cuarenta ciudades de los Estados Unidos de América del número de viviendas nuevas y sus determinantes, correspondientes al año 1988[2].

Esa información, contenida en el archivo **data4-4.gdt**, se refiere a las siguientes variables:

---

[2] Datos extraídos del libro: "*Introductory Econometrics with Applications*" – Ramanathan, Ramu.

BUSTRAVL demanda de viajes urbanos en colectivos (miles de pasajeros/hora)
FARE valor del boleto (en dólares)
GASPRICE precio del galón de gasolina (en dólares)
INCOME ingreso per cápita de los habitantes de la ciudad (en dólares)
POP población de la ciudad considerada (en miles)
DENSITY densidad de la ciudad (personas/millas^2)
LANDAREA área ocupada por la ciudad (en millas^2)

Se realiza una primera aproximación tendiente a explicar el número de viajes demandados en autobús urbano en función las variables **POP**, **LANDAREA**, **INCOME**, **DENSITY**.

a. Escriba la ecuación del modelo que se estimó (coeficientes genéricos), que se indicará en adelante como **[E1]**. Obtenga la correspondiente salida *Gretl*.
b. ¿Cuál es el significado económico del valor del coeficiente asociado a la variable ***INCOME*** del modelo? Comente. ¿A qué nivel mínimo (de dos cifras decimales) resulta estadísticamente significativa la variable? ¿Por qué?
c. ¿Es el modelo globalmente significativo? ¿La variable **LANDAREA** es ella significativa? Comente, desde el punto de vista de la intuición económica, ambas conclusiones.
d. Defina el factor de agrandamiento de varianza para la variable ***LANDAREA***. (Escriba la ecuación que utilizaría para calcularlo (coeficientes genéricos), justifique las variables que utilizaría, explique la invalidez de qué supuesto se intenta detectar mediante su uso). **(*)**.

A continuación, considere una salida <u>auxiliar</u> en *Gretl*, considerando la variable **LANDAREA**, como función lineal de **POP** y **DENSITY**.

e. Escriba la ecuación del modelo <u>auxiliar</u> que estimó (coeficientes genéricos), que se indicará en adelante como **[E2]**.
f. De acuerdo con la estimación de **[E2]** obtenida, ¿qué tipo de vínculo existe entre las variables **LANDAREA**, **POP** y **DENSITY**?
g. Calcule el factor de agrandamiento de varianza para la variable **LANDAREA**. ¿Cuál hubiera sido el valor estimado de la varianza del estimador asociado a esa variable si no se considerara ese factor? Comente en el contexto de las respuestas dadas en **c** y **d**.