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

ECONOMETRIA I - CURSO 2015

PRACTICO 3

Modelo de Regresión Lineal Clásico Multivariado – Restricciones Lineales

EJERCICIO 1:

Para una muestra de 10 observaciones se obtuvieron las siguientes sumas:

$$\Sigma Y = 20{,}0 \qquad \Sigma Y^2 = 88{,}2 \qquad \Sigma Y.X_1 = 59{,}0$$
$$\Sigma X_1 = 30{,}0 \qquad \Sigma X_1^{\,2} = 92{,}0 \qquad \Sigma Y.X_2 = 88{,}0$$
$$\Sigma X_2 = 40{,}0 \qquad \Sigma X_2^{\,2} = 163{,}0 \qquad \Sigma X_1.X_2 = 119{,}0$$

Se pide:

1. Estime el modelo por **MCO**. Calcule el $\boldsymbol{R^2}$ y $\boldsymbol{R^2}$ ajustado

2. A un nivel de significación de 5% y asumiendo normalidad ponga a prueba las siguientes hipótesis:

$H_0)\ \beta_1 = 0$ vs. $H_1)\ \beta_1 \neq 0.$

$H_0)\ 3\beta_1 = \beta_2$ vs. $H_1)\ 3\beta_1 \neq \beta_2.$

$H_0)\ \beta 1 = \beta_2 = 0$ vs. $H_1)\ \beta_1 \neq 0$ y/ó $\beta_2 \neq 0.$

EJERCICIO 2:

Dado el siguiente modelo de regresión lineal $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \varepsilon_t$

Se utilizan n observaciones para estimar el modelo. Las variables X_2 y X_3 son no estocásticas; el rango($\boldsymbol{X}$) = 3, el vector formado por las perturbaciones ε de dimensión $n\times 1$ sigue una distribución normal multivariada con media cero y matriz de varianzas y covarianzas $\sigma^2 \boldsymbol{I}$.

Se pide:

1. Estimados los coeficientes del modelo por **MCO** expresar el cociente t (Student) para probar la hipótesis $H_0)\ \beta_2 + \beta_3 = 1$. Desarrollar el cociente, presentar todos los elementos que componen el numerador y denominador y expresar cómo obtener cada uno de ellos.

2. Realizar la misma prueba de hipótesis solicitada en **1.** mediante el estadístico F. Utilizar la expresión general $H_0) R\beta - r = 0$ y deducir la $COV\left(R\hat{\beta} - r\right)$. Desarrollar el cociente F, ¿qué elementos de la matriz $(X'X)^{-1}$ están en el numerador de F?.

Mostrar que el estadístico F solicitado en **2.** es igual a t^2.

EJERCICIO 3: (ejemplo modificado del libro de J. Wooldridge)

Considere el modelo siguiente que explica los sueldos de los jugadores de la liga mayor de béisbol:

$$\ln(salary_i) = \beta_1 + \beta_2 years_i + \beta_3 gamesyr_i + \beta_4 bavg_i + \beta_5 hrunsyr_i + \beta_6 rbisyr_i + \varepsilon_i$$

Las variables incluidas en el modelo son:

salary: sueldo total en 1993

years: años en la liga

gamesyr:	promedio de partidos jugados por año
bagv:	promedio de bateos a lo largo de la carrera de un jugador
hrunsyr:	cuadrangulares por año
rbisyr:	carreras impulsadas por año

Aplicación de Gretl: Considere los datos contenidos en el archivo *mlb1.odt*

a) Estime el modelo por MCO.
b) Analice la significación global del modelo.
c) Contrastar si las variables son significativas en forma individual.
d) Interprete las estimaciones obtenidas para los coeficientes asociados a las variables *years* y *gamesyr*.
e) El desempeño de los jugadores está representado en el modelo por las variables: *bagv*, *hrunsyr* y *rbisyr*. ¿Realmente contribuyen globalmente a explicar el salario?

EJERCICIO 4:

Dado el modelo $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \varepsilon_t$, la teoría económica indica las siguientes hipótesis alternativas sobre la relación entre β_2 y β_3:

i) $2\beta_2 + \beta_3 = 1$
ii) $\beta_2 + \beta_3 = 0$

Se dispone de los datos sobre las sumas de productos para un vector de unos y las variables:

$$X_{2t},\ X_{3t},\ X_{4t} = X_{2t} - 2X_{3t} \ \text{ y } X_{5t} = Y_t - X_{3t}$$

	1	X_{2t}	X_{3t}	X_{4t}	X_{5t}	Y_t
1	20	80	80	-80	380	460
X_{2t}		436	282	-128	1740	2022
X_{3t}			494	-706	1056	1550
X_{4t}				1284	-372	-1078
X_{5t}					9910	10966
Y_t						12516

Se pide:

1. Estimar el modelo por **MCO** el modelo planteado.
2. Probar la hipótesis $2\beta_2 + \beta_3 = 1$ a un nivel del 5% de significación.
3. Reescribir el modelo en función de los parámetros β_2 y β_3 teniendo en cuenta la restricción **(i)** y estimar por **MCO**.
4. Escribir el modelo incorporando la restricción (i). Observar que se obtienen los mismos resultados que en **3.** Estimar la matriz de varianzas y covarianzas de los estimadores con restricciones.
5. Repetir **4.** para la restricción **ii)**.

En Gretl: Retomar las partes **1**. y **2.** y verifique sus cálculos.

EJERCICIO 5: (diciembre de 2012)

Cuando la teoría económica indica que los coeficientes del modelo de regresión lineal general satisfacen determinadas restricciones, resulta más eficiente introducirlas en el modelo econométrico antes de iniciar el proceso de estimación. La técnica de MCO consiste entonces en *minimizar la norma del vector de error, sujeto a las restricciones en el espacio paramétrico* que aporta la teoría económica.

a) Escriba el problema de optimización bajo restricciones y su correspondiente Lagrangiano, supuesto éstas son q y son lineales independientes:
del tipo $R\beta = r$).

b) Plantee las Condiciones de Primer Orden y obtenga el estimador de mínimos cuadrados restringidos, β_{MCR}.

c) Obtenga la expresión de Var(β_{MCR})

d) <u>Aplicación numérica</u>:
Considere el modelo: $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \varepsilon_t$, $t = 1, 2, ...90$.
Se dispone de los datos siguientes:

$$y'y = 80,\quad X'y = \begin{bmatrix} 3 \\ 2 \\ 1 \\ 2 \end{bmatrix} \text{ y } \quad (X'X)^{-1} = \begin{bmatrix} 5 & -3 & 2 & 0 \\ -3 & 6 & -2 & -4 \\ 2 & -2 & 4 & 3 \\ 0 & -4 & 3 & 4 \end{bmatrix}$$

La teoría económica indica que $\beta_3 + 2\beta_2 = 3$ y $\beta_4 = 6$.

d.1- Calcule el β_{MCO}, sin tener en cuenta las restricciones.
d.2- Plantee la ecuación matricial de las restricciones paramétricas.
d.3- Calcule el β_{MCR}.
d.4- Calcule Var(β_{MCR})

<u>EJERCICIO 6</u>: (octubre de 2007)

Con datos extraídos de una "Encuesta de Ingresos y Gastos Familiares de 1990-91" realizada en España (Fuente: Instituto Nacional de Estadística, España) se estimó un modelo de regresión lineal sobre el gasto de consumo de alimentos. El resultado de la estimación por MCO del modelo, con una muestra de 965 hogares, fue:

(1) $LGCA_i = 3{,}464 + 0{,}467\, LGT_i + 0{,}085\, GF_i - 0{,}132\, DF_i + \hat{\varepsilon}_i$ $\qquad R^2 = 0.357$

Las variables son:

$LGCA_i$ = log de gasto de consumo en alimentación del hogar i-ésimo,
LGT_i = log de gasto de consumo total del hogar i-ésimo,
GF_i = cantidad de personas que integran el hogar i-ésimo,
DF_i = situación laboral de la madre del hogar i-ésimo, (igual a 1 si trabaja y 0 en otro caso).
$\hat{\varepsilon}_i$ = residuo_i

La estimación de la matriz de varianzas y covarianzas del vector de coeficientes $\boldsymbol{\beta}_{\mathbf{MCO}}$ fue:

C	*LGT*	*GF*	*DF*
0,054851	-0,005729	4,91E-05	0,001171
	0,000616	-4,07E-05	-0,000161
		8,04E-05	4,18E-05
			0,000764

Se pide:

1) Si agregamos un nuevo regresor al modelo, por ejemplo: *edad promedio de los integrantes del hogar*, y volvemos a regresar la variable *LGCA*, ahora sobre 4 variables explicativas. ¿El R^2 de la nueva regresión podrá ser menor a 0,357? Justifique su respuesta. ¿Qué conclusión deduce de la utilidad de R^2 para juzgar si un modelo - el que contiene 4 variables explicativas – es relativamente mejor para explicar *LGCA* que el que contiene 3? Proponga alguna herramienta útil para comparar modelos que tienen distinto número de regresores.
2) Interprete el significado económico del coeficiente de la variable *LGT*, e indique el valor esperado del porcentaje de incremento del *gasto de consumo en alimentos* si *gasto de consumo total* aumentara un 10%.
3) Estime el error estándar del coeficiente de la variable *DF*, y contraste la hipótesis de que no ejerce influencia sobre la variable *LGCA*. (Exprese hipótesis nula, alternativa, estadístico que utiliza en el contraste y el resultado).
4) Examine la significación del modelo en su conjunto, al 5%. (Exprese hipótesis nula, alternativa, estadístico que utiliza en el contraste y el resultado).
5) Contraste la hipótesis $2\beta_2 + \beta_3 = 0$, al 5%. Detalle el estadístico que utiliza y el resultado del contraste.

EJERCICIO 7:

Se dispone de información de 420 distritos educacionales de California (USA), correspondientes al año lectivo 1998-1999. La información está contenida en el archivo caschool_ej7.gdt. Se desea explicar la matrícula en el distrito (*enrl_tot*), considerando las variables siguientes:

- *teachers*: número de docentes de tiempo completo/número de estudiantes
- *calw_pct*: porcentaje de alumnos incriptos en programas de asistencia pública
- *computer*: promedio de computadoras por salón de clase
- *avinc*: ingreso medio anual en el distrito (en miles de dólares)

1. Escriba la ecuación del modelo considerado.
2. Analice la significación global del modelo.
3. Analice la significación individual de los coeficientes del modelo. Interprete y comente las estimaciones de los coeficientes obtenidas.
4. ¿Es cierto que el efecto sobre la matrícula del número de profesores es opuesto y en módulo 1,25 veces igual al efecto de la variable *avinc*?
5. ¿Es cierto que el efecto sobre la matrícula del número de profesores es opuesto y en módulo 2,5 veces igual al efecto de la variable *calw_pct*?
6. Plantéese las preguntas **4.** y **5.** simultáneamente. Comente si existe diferencia en las conclusiones de ambos planteos.