

**PRÁCTICA 2: REGRESIÓN LINEAL MULTIPLE**

1. Se desea estimar un modelo que permita estudiar, para cierta localidad venezolana, la relación existente entre los gastos en alimentación de una familia (Y) (en miles de Bs.), con base a la información que proporcionan las variables regresoras  $X_2$  = Ingresos mensuales (en miles de Bs.) y  $X_3$  = Número de miembros de la familia. Para ello se recoge una muestra aleatoria simple de información perteneciente a 15 familias cuyos resultados son los de la tabla adjunta:

FAMILIA	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Gasto	0,43	0,31	0,32	0,46	1,25	0,44	0,52	0,29	1,29	0,53	0,35	0,78	0,43	0,47	0,38
Ingreso	2,1	1,2	0,9	1,6	6,2	2,3	1,8	1,0	8,9	2,4	1,2	4,7	3,5	2,9	1,4
Tamaño	3	4	5	4	4	3	6	5	3	2	4	3	2	3	4

Obtenga mediante el modelo estimado de regresión múltiple

$$(X'X)^{-1} = \begin{matrix} & 1,362187716 & -0,09296981 & -0,282159759 \\ & -0,09296981 & 0,016625731 & 0,012629161 \\ & -0,282159759 & 0,012629161 & 0,067285613 \end{matrix}$$

(Debe obtener algo como  $Y_{estimado} = -0,055881151 + 0,14405461 X_2 + 0,05577225 X_3$ ); y de esta salida, obtenga e interprete

- Cuál es el error promedio cometido por el modelo? Obtengalo de la salida del programa R.
- Es válido económicamente el modelo?
- Cuál es el error promedio cometido al estimar cada uno de los coeficientes del modelo?
- Son  $X_2$  y  $X_3$  estadísticamente significativas? Considere un nivel de significación del 5%.
- Obtenga  $R^2$  ajustado del modelo e intérpretelos.
- Es el modelo estimado, apto para pronosticar al 5%?
- Obtenga los residuos muestrales, diga cuánto vale la suma de éstos y remítalo a alguna propiedad conocida de los modelos estimados de regresión.

2. Con miras a revisar las tarifas mensuales cobradas al concesionario del cafetín, el coordinador de servicios básicos de cierta facultad monitoreó, durante una muestra aleatoria de 9 meses de los años 2004-2005, el comportamiento de las variables que a continuación se muestran:

- Precio aproximado de los almuerzos vendidos en el cafetín (en miles de Bs.)
- Numero de estudiantes inscritos en la facultad (y las facultades vecinas) del cafetín (en decenas de miles)
- Ingreso mensual del concesionario (en millones de Bs.)
- Número de quioscos informales que funcionan en los alrededores de la facultad (y de las facultades vecinas) (en unidades)

La base de datos empleada es la indicada en la tabla adjunta:

MES	PRECIO ( )	NUMINSCR ( $X_3$ )	INGRESO ( )	NUMQUIOSC ( $X_4$ )
Enero 2004	5	0,1	25	20
Febrero 2004	9	1	59	9
Marzo 2004	10,5	1,5	65	10
Junio 2004	7	0,2	28	18
Julio 2004	9	0,4	55	10
Agosto 2004	6	0,8	33	16
Enero 2005	8,5	0,7	43	15
Abril 2005	7,2	0,6	36	17
Mayo 2005	10	2	68	9

Además, se tiene la inversa de la matriz  $(X'X)$ :

$$(X'X)^{-1} = \begin{bmatrix} 35,43140021 & -2,522426648 & -0,180505631 & -1,084239571 \\ -2,522426648 & 0,209219941 & -0,118196358 & 0,068217689 \\ -0,180505631 & -0,118196358 & 0,778338181 & 0,036100476 \\ -1,084239571 & 0,068217689 & 0,036100476 & 0,036849238 \end{bmatrix}$$

Con base a la información anterior:

- Identifique, justificando apropiadamente, cuales son la variable endógena y explicativa faltante de este modelo
- Estime a través de un programa de computación (R) por el método de los mínimos cuadrados un modelo de regresión múltiple, que le permita al coordinador orientar su decisión. Interprete sus coeficientes.

- c) Al 99% de confianza, cuantifique el efecto parcial de cada variable explicativa del modelo sobre la variable endógena en estudio. Explique e interprete.
- d) Podría decirse que el modelo pasa la validación global al 5% de significación? Y al 1%? Obtenga todos los indicadores posibles de la salida del programa de computación (R).
- e) Obtenga dos estimaciones PUNTUALES para el ingreso mensual esperado del concesionario, UNA para un mes en el que el precio promedio de los almuerzos fue de 10.000 Bs., funcionaban 9 quioscos y habían 2000 alumnos inscritos en la facultad (y facultades vecinas) y OTRA cuando el precio promedio de los almuerzos fue de 9.000 Bs., funcionaban 9 quioscos y habían 1000 alumnos inscritos en la facultad. Compare ambos resultados y diga en donde el error muestral es mayor. Utilice el modulo ANALISIS DE RESIDUALES DE LA SALIDA DE EXCEL.
- f) Antes de estimar el modelo con el que hasta ahora ha venido trabajando, se disponía de un modelo con solo dos variables explicativas:  $X_2$  y  $X_3$ , para el cual el coeficiente de determinación múltiple era de 0,81, ¿considera Ud. que al agregar la variable explicativa  $X_4$  mejoró la bondad de ajuste del modelo múltiple? Justifique apropiadamente.

3. En un hipotético país (dividido territorialmente en 18 estados), que implementó en el año 2000 un programa de vehículos familiares (para uso particular, no público, de máximo cinco puestos, y precio más económico que los restantes del mercado) se publicó, el año pasado (2004) una investigación relativa al comportamiento de las variables que a continuación se muestran:

Y: Número de VEHÍCULOS USADOS (del tipo familiar) vendidos anualmente en el estado (en cientos)

$X_2$ : Precio promedio de los VEHÍCULOS NUEVOS (del tipo familiar) en el estado (en millones de u.m.)

$X_3$ : Precio promedio de los VEHÍCULOS USADOS (del tipo familiar) (en millones de u.m.)

$X_4$ : Precio aproximado de la póliza de seguros contra todo riesgo (en cientos de miles de u.m.)

Disponiéndose de la siguiente base de datos, para una muestra aleatoria, según se indica en la tabla adjunta:

ESTADO	Y	$X_2$	$X_3$	$X_4$
A	10	7	6	8
B	20	8	5	7
C	50	9	5	6
D	60	10	4	9
E	70	7	3	10
F	80	11	4	7
G	90	8	3	9
H	100	9	3	5

Además, se tiene la inversa de la matriz  $(X'X)$ :

$$(X'X)^{-1} = \begin{bmatrix} 19,84137661 & -1,101547705 & -0,916149677 & -0,844119376 \\ -1,101547705 & 0,088552211 & 0,025272644 & 0,030627571 \\ -0,916149677 & 0,025272644 & 0,124469405 & 0,02422778 \\ -0,844119376 & 0,030627571 & 0,02422778 & 0,062953046 \end{bmatrix}$$

Con base a la información anterior:

- a) Estime por el método de los mínimos cuadrados el modelo de regresión múltiple, mediante EXCEL e interpretando sus coeficientes.
- b) Al 5% de significación, resultan significativos los efectos de las variables precio de los vehículos usados y precio de la póliza de seguros sobre la variable endógena del estudio? Obtenga esta respuesta de la respectiva salida de R y explique e interprete estadística y económicamente.
- c) Si se sabe que  $STC=7200$ , calcule e interprete  $r_{múltiple}$  y  $r^2_{múltiple}$  en este modelo estimado.
- d) Podría decirse que el modelo es apto para predecir al 5% de significación? Y al 1%? Haga las pruebas manualmente y compárelas con las respectivas salidas de R.
- e) ¿El modelo estimado sobrevalora o subvalora el número esperado de vehículos usados vendidos anualmente, en un estado donde el precio estimado de las pólizas a todo riesgo es de 900.000 u.m., el precio aproximado de los vehículos usados ronda los tres millones de u.m. y el de los vehículos nuevos alcanzó los 8 millones de u.m.?

4. La nueva operadora de telefonía celular (NOVOCEL)<sup>1</sup> desea conocer para sus tiendas principales de cada estado, la relación existente entre: Y: el número de tarjetas pre-pago de celulares vendidas mensualmente (en decenas de unidades) en el presente año, y las variables:  $X_2$ : Número de habitantes residenciados en el Municipio donde se ubica la tienda (en miles),  $X_3$ : Precio promedio de las tarjetas vendidas en la tienda (en decenas de miles de Bs.) y  $X_4$ : Número de teléfonos celulares vendidos por la empresa en todo el país durante el año 2005 (en miles), disponiéndose de la siguiente información:

Tienda	1	2	3	4	5	6	7	8	9	10	11	12
Y	220	250	310	420	350	510	400	500	500	550	450	320
$X_2$	45	55	62,5	65	63,75	75	64,5	68,5	70	77,5	60	63,75
$X_3$	3,2	2,2	1	0,5	1,5	0,5	1,5	0,5	0,5	0,3	0,5	1,5
$X_4$	2	2	3	4	3	6	4	5	5	6	5	3

<sup>1</sup> Entre las novedades que ofreció esta empresa cuando empezó sus operaciones estaba la venta de tarjetas pre-pago con precios desde 3.000 Bs.

La matriz inversa,  $(X'X)^{-1}$  es:

25,18557279	-0,331398049	-3,039304131	-0,088462719
-0,331398049	0,005562776	0,027231134	-0,014216857
-3,039304131	0,027231134	0,566786278	0,160939353
-0,088462719	-0,014216857	0,160939353	0,204391076

- Estime el modelo de regresión lineal múltiple por el método de los mínimos cuadrados, interpretando sus coeficientes.
- Evalúe si el modelo pasa la validación económica, justificando apropiadamente.
- ¿Es igualmente estable la estimación de cada parámetro en este modelo?. Explique apropiadamente.
- ¿Cuál es el error promedio cometido por el modelo? Interpretélo.
- Obtenga e interprete el error estándar de cada estimador mínimo cuadrado del modelo.
- Es apto el modelo para la predicción al 5%; y al 0,5%?
- Cuando se han vendido 5.000 celulares, hay 70.000 habitantes en el municipio y el precio promedio de las tarjetas alcanza los 5.000 Bs.,estime el número promedio de celulares vendidos.

Salida de una hoja de cálculo: Problema 4 (con indicadores de validación estadística)

Resumen

<i>Estadísticas de la regresión</i>	
Coefficiente de correlación múltiple	0,98484056 R xy
Coefficiente de determinación R <sup>2</sup>	0,96991093 R cuadrado
R <sup>2</sup> ajustado	0,95862752 R cuadrado ajustado
Error típico	22,0582275 regresion S= error de estimación de la
Observaciones	12 n =tamaño de la muestra

ANÁLISIS DE  
VARIANZA

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	Fcalc en la prueba F	VALOR-P
				F	Valor crítico de F
Regresión	3	125474,1434	41824,71448	85,95908028	1,99271E-06
Residuos	8	3892,523219	486,5654023		
Total	11	129366,6667			

	Coefficientes= bi	Error típico =Sbi	Estadístico t	VALORES-P	INT DE CONF PARA LOS BETAS	Inferior 95%	Superior 95%
			t calc en las pruebas t	Probabilidad			
Intercepto	90,4994724	110,6997216	0,817522132	0,4372986		-164,774708	345,773653
Variable X 2	1,45558325	1,64519133	0,884750134	0,402099343		-2,33823721	5,24940371
Variable X 3	-17,005139	16,60658284	-1,023999881	0,335799253		-55,3000123	21,2897346
Variable X 4	58,4468717	9,972443345	5,860837675	0,000378107		35,45036121	81,4433821

BIBLIOGRAFÍA CONSULTADA:

- Gujarati, D. (2004). *Econometría Básica*. Cuarta Edición. Editorial Mc. Graw-Hill
- Newbold, P. (1998). *Estadística para los negocios y la Economía*. Cuarta Edición. Editorial Mc. Graw-Hill
- Pindick, y Rubinfeld, (2001). *Econometría, modelos y pronósticos*. Cuarta Edición. Editorial Mc. Graw-Hill.
- Vilar, J. (2000). *Problemas y prácticas de regresión lineal múltiple*. Disponible en [www. Monografias.doc.com](http://www.Monografias.doc.com).
- Webster, A. (2000) *Estadística aplicada a los negocios y la Economía*. Editorial Mc. Graw-Hill.

Elaborada y reformulada por Joana Matzavracos P. (JULIO 2014)