



**UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA**  
*La Universidad Católica de Loja*

**ÁREA ADMINISTRATIVA**

**TÍTULO DE ECONOMISTA**

**Econometría con Stata: Aplicaciones a la economía ecuatoriana**

**TRABAJO DE TITULACIÓN**

**AUTORAS:**       Iñiguez Fernández, María Paula  
                      Palacio González, María Cecibel

**DIRECTOR:**       Correa Quezada, Ronny Fabián

**Loja- ECUADOR**

**2016**



*Esta versión digital, ha sido acreditada bajo la licencia Creative Commons 4.0, CC BY-NY-SA: Reconocimiento-No comercial-Compartir igual; la cual permite copiar, distribuir y comunicar públicamente la obra, mientras se reconozca la autoría original, no se utilice con fines comerciales y se permiten obras derivadas, siempre que mantenga la misma licencia al ser divulgada. <http://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>*

2016

## **APROBACIÓN DEL DIRECTOR DEL TRABAJO DE TITULACIÓN**

PhD.

Ronny Fabián Correa Quezada.

### **DOCENTE DE LA TITULACIÓN**

De mi consideración:

El presente trabajo de titulación: “Econometría con Stata: Aplicaciones a la economía ecuatoriana” realizado por María Paula Iñiguez Fernández y María Cecibel Palacio González, ha sido orientado y revisado durante su ejecución, por cuanto se aprueba la presentación del mismo.

Loja, agosto de 2016

f) .....

## DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

“Yo María Paula Iñiguez Fernández y María Cecibel Palacio González declaramos ser autoras del presente trabajo de titulación “Econometría con Stata: Aplicaciones a la economía ecuatoriana” de la Titulación de Economía, siendo el PhD. Ronny Fabián Correa Quezada director del presente trabajo; y eximo expresamente a la Universidad Técnica Particular de Loja y a sus representantes legales de posibles reclamos o acciones legales. Además certifico que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo, son de mi exclusiva responsabilidad.

Adicionalmente declaro conocer y aceptar la disposición del Art. 88 del Estatuto Orgánico de la Universidad Técnica Particular de Loja que en su parte pertinente textualmente dice: “Forman parte del patrimonio de la Universidad la propiedad intelectual de investigaciones, trabajos científicos o técnicos y tesis de grado o trabajos de titulación que se realicen con el apoyo financiero, académico o institucional (operativo) de la Universidad”

f. ....

Autora: María Paula Iñiguez Fernández

Cédula: 1105653685

f. ....

Autora: María Cecibel Palacio González

Cédula:1104776438

## DEDICATORIA

Dedico mi trabajo primeramente a Dios por darme la fortaleza necesaria para cumplir mis metas, por brindarme nuevas oportunidades, por ser mi guía en cada paso que doy.

A mis padres Bolívar y Patricia, por apoyarme siempre durante toda mi carrera, por confiar en mí en todo momento, por ser los mejores padres.

A mis hermanos Ligia y Diego, quienes me han brindado todo su cariño y amor, y siempre han estado para mí en todo momento.

*María Paula*

El presente trabajo de titulación está dedicado primeramente a Dios, cuya gracia divina me llevó por senderos de rectitud y verdad.

A mis padres José Palacio y Rosa González, mis hermanos Beatriz y José Alberto, mi demás familiares y amigos que con su incondicional presencia y amor impulsaron la culminación de esta meta.

*María Cecibel*

## **AGRADECIMIENTOS**

Agradezco a mi familia por ser el motor que me impulsa a seguir adelante.

A mi director de tesis PhD. Ronny Correa, quien con sus orientaciones y sabios consejos me ha ayudado a culminar con éxito mi carrera universitaria.

A mi compañera de tesis María Cecibel por sus valiosas aportaciones para este trabajo. ¡Juntas lo logramos!

A la Universidad Técnica Particular de Loja por abrirme sus puertas para poder formarme profesionalmente.

Finalmente, gracias a todas las personas que directa o indirectamente contribuyeron a que este trabajo pudiera llevarse a cabo.

*María Paula*

De manera muy especial agradezco a aquellos seres perfectos que me guiaron con sus enseñanzas y valores morales durante mi vida, mis padres.

A los docentes de la Universidad Técnica Particular de Loja, quienes inculcaron grandes conocimientos durante mi periodo de estudio, en especial al director de tesis PhD. Ronny Correa que fue el gestor principal de este proyecto al igual que los docentes revisores que contribuyeron con valiosos aportes para el mejoramiento del mismo.

Finalmente agradezco a mi compañera de tesis María Paula, por el espléndido aporte realizado en este trabajo de titulación y a mis amigas Anita y Soledad que me acompañaron durante toda mi vida académica y ocupan una importante parte de mi corazón.

*María Cecibel*

## ÍNDICE DE CONTENIDOS

CARÁTULA.....	i
APROBACIÓN DEL DIRECTOR DEL TRABAJO DE FIN DE TITULACIÓN.....	ii
DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS.....	iii
DEDICATORIA .....	iv
AGRADECIMIENTOS.....	v
ÍNDICE DE CONTENIDOS .....	vi
ÍNDICE DE COMANDOS .....	x
RESUMEN .....	1
ABSTRACT .....	2
INTRODUCCIÓN.....	3
CAPÍTULO I: IMPORTANCIA DEL STATA EN LA ECONOMETRÍA.....	5
1.1.    Econometría.....	6
1.1.1.    Definición .....	6
1.1.2.    Origen y evolución.....	7
1.1.3.    Importancia .....	9
1.1.4.    Metodología .....	10
1.2.    Stata .....	12
1.2.1.    Historia.....	12
1.2.2.    Estructura .....	16
1.2.3.    Tipos de archivos .....	18
1.2.4.    Uso actual.....	18
1.3.    Econometría de Gujarati & Porter .....	19
1.3.1    Como relacionamos el libro con el Stata .....	20

1.4.	Dificultades en el uso de Stata en econometría.....	20
CAPÍTULO II: DATOS Y METODOLOGÍA .....		26
2.1	Presentación de datos. ....	27
2.1.1	Descripción de variables económicas.....	27
2.1.2	Descripción de variables sociodemográficas.....	34
2.1.3	Descripción de variables ambientales .....	41
2.1.4	Descripción de variables hipotéticas.....	44
2.1.5	Descripción de variables de encuestas. ....	45
2.2	Metodología .....	47
CAPÍTULO III: APLICACIÓN ECONOMETRICA A CASOS ECUATORIANOS .....		49
3.1	Aspectos básicos.....	50
3.1.1	Abrir y cerrar archivos: dta y do .....	50
3.1.2	Ayuda e instalación de comandos .....	52
3.1.3	Manejo de datos .....	54
3.1.3.1	Introducción de datos.....	54
3.1.3.2	Datos numéricos almacenados como cadena .....	57
3.1.3.3	Eliminar variables .....	59
3.1.3.4	Crear variables .....	60
3.1.3.5	Renombrar y etiquetar variables .....	61
3.1.3.6	Borrar y editar variables .....	63
3.1.4	Tablas .....	64
3.2	Naturaleza del análisis de regresión .....	66
3.2.1	Diagramas de dispersión.....	66
3.2.1.1	Diagrama de dispersión simple .....	66
3.2.1.2	Diagrama de dispersión con ajuste lineal.....	67
3.2.1.3	Diagrama de dispersión con ajuste polinómico.....	68
3.2.1.4	Diagrama de dispersión con línea de tendencia múltiple .....	70
3.2.2	Gráficos de series de tiempo .....	71



3.2.2.1	Evolución de una serie de tiempo .....	71
3.2.2.2	Evolución de dos o más series de tiempo.....	72
3.2.3	Gráficos de barras .....	73
3.2.3.1	Gráfico de barras simple.....	74
3.2.3.2	Gráfico de barras agrupado .....	76
3.2.4	Gráficos de sectores .....	78
3.2.5	Editor de gráficos.....	80
3.3	Análisis de regresión con dos variables: algunas ideas básicas.....	82
3.3.1	Estimación del modelo econométrico .....	83
3.3.2	Función de regresión poblacional .....	85
3.3.3	Función de regresión muestral.....	87
3.3.4	Linealidad .....	88
3.4	Análisis de regresión con dos variables: problema de estimación .....	91
3.4.1	Método de Mínimos Cuadrados Ordinarios (MCO) .....	92
3.4.2	Coeficiente de determinación $r^2$ .....	94
3.4.3	Prueba de significancia de los coeficientes de regresión.....	96
3.4.4	Pruebas de normalidad .....	97
3.5.1	Regresión a través del origen .....	101
3.5.2	Escalas y unidades de medición.....	103
3.5.3	Regresión sobre variables estandarizadas.....	106
3.5.4	Modelo log-lineal .....	109
3.5.5	Modelos semilogarítmicos: log- lin y lin-log .....	110
3.5.6	Modelo recíproco.....	113
3.5.7	Modelo recíproco logarítmico.....	114
3.6	Análisis de regresión múltiple .....	116
3.6.1	Modelo con tres variables .....	116
3.6.2	Otra manera de estimar los coeficientes de regresión parcial .....	118
3.6.3	Regresión múltiple sobre variables estandarizadas .....	120
3.6.4	Efecto simultáneo .....	122

3.6.5	Comparación de dos valores de $R^2$ .....	124
3.6.6	La función de producción Cobb-Douglas: más sobre la forma funcional.....	126
3.6.7	Modelos de regresión polinomial .....	128
3.6.8	Pruebas de hipótesis sobre coeficientes de regresión individual.....	130
3.6.9	Pruebas de significancia general de la regresión muestral.....	133
3.6.10	Prueba de igualdad de dos coeficientes de regresión .....	135
3.6.11	Mínimos cuadrados restringidos: pruebas de restricciones de igualdades lineales.....	137
3.6.12	Prueba de estabilidad estructural: La prueba de Chow .....	141
3.6.13	Prueba de la forma funcional de la regresión: elección entre modelos de regresión lineal y log-lineal.....	145
3.7	Modelos de regresión con variables dicótomas .....	147
3.7.1	Modelos ANOVA .....	148
3.7.2	Modelos ANOVA con dos variables.....	150
3.7.3	Modelos ANCOVA.....	152
3.7.4	La variable dicótoma alternativa a la prueba de Chow .....	153
3.7.5	Efectos de interacción al utilizar variables dicótomas .....	156
3.7.6	Uso de variables dicótomas en el análisis estacional .....	158
3.7.7	Regresión lineal por segmentos.....	161
3.7.8	Variables dicótomas en regresiones semilogarítmicas .....	162
3.8	Multicolinealidad .....	164
3.8.1	Detección de Multicolinealidad .....	164
3.8.2	Corrección de Multicolinealidad .....	170
3.9	Heteroscedasticidad .....	174
3.9.1	Detección de Heteroscedasticidad.....	175
3.9.1.1	Métodos informales .....	175
3.9.1.2	Métodos formales.....	177
3.9.2	Corrección de Heteroscedasticidad .....	182
3.10	Autocorrelación .....	184

3.10.1 Detección de Autocorrelación .....	185
3.10.2 Corrección de Autocorrelación .....	196
3.11. Creación de modelos econométricos.....	202
3.11.1 Pruebas de errores de especificación.....	203
3.11.2 Corrección del subajuste de un modelo .....	207
3.11.3 Test de especificación .....	208
3.11.4 Normalidad en las variables y los residuos.....	211
3.12. Modelos de regresión no lineales.....	214
3.12.1 Estimación de modelos de regresión lineal y no lineal .....	214
3.12.2 Métodos para estimar modelos de regresión no lineales .....	217
3.13 Modelos de regresión de respuesta cualitativa.....	220
3.13.1 Modelo lineal de probabilidad MLP .....	220
3.13.2 Estimación del modelo logit.....	222
3.13.3 Logit para datos agrupados .....	223
3.13.4 Modelo probit .....	225
3.13.5 Pruebas de ajuste para modelos de respuesta cualitativa. ....	227
3.13.6 Modelo tobit .....	229
3.14 Modelos regresión con datos de panel .....	230
3.14.1 Modelo de regresión con MCO .....	230
3.14.2 Modelo de mínimos cuadrados con variable dicótoma (MCVD) de efectos fijos....	232
3.14.3 Modelos de Efectos Fijos (MEF) .....	233
3.14.4 Modelos de Efectos aleatorios (MEFA).....	234
3.14.5 Test de Hausman .....	236
CONCLUSIONES .....	238
RECOMENDACIONES .....	240
BIBLIOGRAFÍA.....	241
ANEXOS .....	244

## ÍNDICE DE COMANDOS

help .....	52
input .....	55
destring .....	58
tostring .....	58
drop .....	59
clear all/clear .....	59
generate/gener/gen .....	60
rename .....	62
label/ label variable .....	62
drop if .....	63
replace .....	63
table .....	64
tab .....	64
scatter .....	66
twoway .....	67
lfit .....	67
qfit .....	68
fpfit .....	68
line .....	71
twoway connected .....	71
graph bar .....	74

graph hbar .....	74
if.....	74
over .....	74
blabel.....	76
graph pie .....	78
title.....	78
by.....	78
regress/reg.....	83
predict .....	89
connect (1).....	89
exp .....	89
predict <i>nombre_residuos</i> , resid.....	92
total .....	92
correlate .....	94
histogram .....	98
histogram <i>nombre_residuos</i> , normal.....	98
pnorm.....	98
summarize/sum .....	98
detail.....	98
sktest.....	98
regress <i>varY varX</i> , noconst.....	102
log .....	109
egen <i>variablestan</i> = std( <i>variable</i> ) .....	120

vce, covariance.....	135
in.....	142
i.....	148
mean .....	158
estat vif.....	165
graph matrix.....	167
rvfplot .....	176
estat imtest, White .....	177
estat hettest .....	180
ivhettest.....	182
robust .....	183
rreg.....	183
rstandard.....	185
rstudent .....	185
ac.....	187
estat dwatson.....	190
estat bgodfrey .....	191
corrgram.....	193
prais .....	298
newey.....	201
ovtest.....	206
linktest .....	209
kdensity/normal.....	211

normal .....	211
sktest.....	213
logit.....	222
glogit.....	224
probit .....	225
mfx .....	226
estat gof .....	227
estat class .....	228
tobit .....	229
xtreg <i>varY varX</i> , fe.....	234
xtreg <i>varY varX</i> , re.....	235
hausman .....	236

## **RESUMEN**

El presente trabajo de fin de titulación desarrolla un conjunto de modelos econométricos aplicados a la realidad ecuatoriana, con el fin de brindar una contribución significativa dentro de la academia bajo el aporte teórico y práctico de Gujarati & Porter (2010). La econometría como herramienta de la economía, surge como un factor determinante en el perfil profesional de una economista, motivo por el cual se abordó dicha temática desarrollando una guía detallada de los procesos econométricos aplicados a la realidad ecuatoriana, ajustable en diversos tipos de datos y permitiendo que el lector pueda replicar estos ejercicios sin ningún problema, además de brindarle las pautas necesarias para generar nuevos modelos en cualquier ámbito de la economía. El surgimiento de este trabajo es respaldado además, por la necesidad de los estudiantes de economía de la Universidad Técnica Particular de Loja quienes expresaron que no hay suficientes recursos metodológicos disponibles para el manejo de Stata. Motivo por el cual, se espera que el presente trabajo tenga un efecto no solo en el ámbito académico local sino nacional y con posibles repercusiones dentro de la región.

**PALABRAS CLAVES:** Econometría, Stata, Ecuador.



## **ABSTRACT**

The following research develops a set of econometric models applied to Ecuadorian reality, in order to provide a significant contribution within the academy under the theoretical and practical contribution of Gujarati & Porter (2010). Econometrics as a tool of the economics, appears as a determining factor in the professional profile of an economist, this is why the issue was addressed by developing a detailed guide to the econometric processes applied to the Ecuadorian reality, adjustable in various types of data and allowing the reader to replicate these exercises without any problem, and also providing the guidelines necessary to generate new models in any field of economics. The arising of this work is further supported by the need from students of economics at the Technical University of Loja who expressed the insufficient in methodological resources available for handling Stata, therefore this work can extend its effect not only in the academic scope, but also in a national one with possible impact within the region.

**KEYWORDS:** Econometrics, Stata, Ecuador

## INTRODUCCIÓN

El presente trabajo realiza aplicaciones de la econometría a la realidad ecuatoriana a través de diversos ejercicios econométricos utilizando el software estadístico Stata 13. Estos ejercicios abarcan los 16 primeros capítulos del libro “Econometría” de Gujarati & Porter (2010), ya que generalmente éste es el texto base que utilizan los estudiantes de la Titulación de Economía.

Por tanto, este trabajo comprende tres capítulos. El primero de ellos consiste en una descripción básica de la econometría como herramienta económica, en la que se analiza el uso del software estadístico Stata, la aplicación del libro de “Econometría” de Gujarati & Porter (2010) como bibliografía básica de referencia, para luego culminar con un estudio exploratorio a los estudiantes de econometría de la “Universidad Técnica Particular de Loja” (UTPL) con el fin de averiguar las dificultades por el uso de Stata en la econometría. El capítulo II describe las variables utilizadas en el desarrollo de los ejercicios econométricos y explica la metodología empleada para la resolución de un ejercicio estándar. Finalmente, el capítulo III desarrolla 108 ejercicios econométricos resueltos con Stata abarcando las temáticas de la bibliografía aplicados a la realidad ecuatoriana.

La importancia de este trabajo radica en el aporte e impacto que podría generar en los estudiantes de econometría, ya que serviría como un recurso metodológico de aprendizaje en la aplicación de la econometría. Los estudiantes tendrían acceso a una guía econométrica con ejercicios resueltos de manera clara y explicativa. Es así que se da respuesta a la necesidad de los alumnos de la materia de econometría al existir una guía instructiva usando Stata. Con ello se espera que los estudiantes desarrollen habilidades en el manejo del este software estadístico y puedan aplicar con facilidad la econometría a través de la creación de modelos econométricos.

El objetivo de elaborar una guía con diversos modelos aplicados a la realidad ecuatoriana se ha cumplido casi en su totalidad, ya que solamente 4 de los 108 ejercicios elaborados utilizan datos hipotéticos por la limitante de que estos no se ajustan a distintos aspectos teóricos de la econometría que son necesarios conocer. A pesar de que las bases de datos de las fuentes oficiales de información disponibles para Ecuador son poco desagregadas, además de que no se encuentran datos de algunas variables interesantes para hacer modelaciones, se ha logrado utilizar variables económicas, ambientales y sociodemográficas del Ecuador en 104 ejercicios desarrollados en el capítulo III del presente trabajo.

La metodología utilizada para la resolución de los ejercicios es aquella elegida por los estudiantes de econometría en el estudio exploratorio, como consecuencia de ello el desarrollo de un ejercicio estándar consta de 4 partes: la primera contiene una breve explicación del tema a abordar en el ejercicio, la segunda consta del listado de comandos a utilizar con su respectiva definición y sintaxis añadiendo además la rutina de comandos utilizada, la tercera parte comprende el desarrollo del ejercicio paso a paso y la cuarta parte brinda una interpretación de los resultados obtenidos.

## **CAPÍTULO I**

### **IMPORTANCIA DEL STATA EN LA ECONOMETRÍA**

## **1.1. Econometría**

### **1.1.1. Definición.**

La econometría es un método que ha tomado relevancia dentro del campo económico para la resolución y análisis de los diversos y variados problemas que acontecen en la sociedad. Definir la econometría desde la economía cuantitativa, estadística económica y matemática económica surge como un gran reto, puesto que la definición de una ciencia debe ser progresiva y provisional (Tintner, 1953). Desde los años treinta hasta la actualidad han surgido diversas aportaciones a la definición de la econometría, tales como la de Åkerman (1938) quien argumenta que la econometría está basada en una síntesis intuitiva de observación y construcción y la de Sloan (1949) el cual menciona que la econometría es “Un instrumento de medición cuyo fin es probar y desarrollar a la teoría económica” a lo que Schumpeter (1946) agrega “No es la teoría cuantitativa lo que se pretende comprobar, toda teoría económica es cuantitativa por naturaleza, sino a la teoría numérica”.

Un sin número de definiciones se han presentado desde los años 30 hasta la actualidad, dentro de la presente investigación se muestran los aportes más relevantes de grandes figuras económicas y destacados en el campo estadístico tales como:

- Frisch (1933a): “la econometría implica la conjunción de teoría económica cuantitativa y observación estadística.”
- Samuelson, Koopmans y Stone (1954): “La econometría puede ser definida como el análisis cuantitativo de fenómenos económicos reales basados en el desarrollo simultáneo de la teoría y la observación, relacionados mediante métodos apropiados de inferencia.”
- Valavanis (1959): “El objetivo de la econometría es expresar las teorías económicas bajo una forma matemática a fin de verificarlas por métodos estadísticos y medir el impacto de una variable sobre otra.”
- Klein (1962): “El principal objetivo de la econometría es dar contenido empírico al razonamiento a priori de la economía.”
- Goldberger (1964): “la econometría es la ciencia social en la cual la teoría económica, las matemáticas y la inferencia estadística son herramientas aplicadas al análisis de los fenómenos económicos.”
- Malinvaud (1966): «La econometría consiste en encontrar una serie de supuestos lo suficientemente específicos y realistas, para que permitan aprovechar de la mejor manera los datos disponibles.”

- Christ (1966): “Métodos cualitativos que explican el comportamiento de variables ya observadas, o predicen la conducta de variables aún no observadas.” Intriligator (1978): “Rama de la economía que se ocupa de la estimación empírica de relaciones económicas”
- Chow (1983): “Arte y ciencia de usar métodos para la medida de relaciones económicas”
- Griliches e Intriligator (1984): “La econometría es la aplicación de las matemáticas y los métodos estadísticos al análisis de los datos económicos.”
- Hill (1988): “La econometría es el uso de la teoría económica, economía matemática e inferencia estadística como fundamentos analíticos para aseverar o refutar conocimientos dentro de la teoría económica, de modo que esta información pueda servir de base para la toma de decisiones y la elección.”
- Maddala (1996): “La econometría es la aplicación de métodos matemáticos y estadísticos para el análisis de datos económicos con el fin de brindar contenido empírico a las teorías económicas y someterlas a verificación.”

Resumiendo los distintos puntos de vista y agrupando las diferentes aportaciones en una sola definición se podría decir que la econometría es una ciencia que integra elementos cualitativos y cuantitativos mediante herramientas matemáticas y estadísticas que permiten analizar datos frente a la teoría económica, de esta manera se tendría una base para la toma de decisiones. Esta definición, tal como se mencionó anteriormente, deberá ser progresiva en el tiempo y se complementará con los avances en las distintas áreas que conforman la econometría.

### **1.1.2. Origen y evolución.**

Etimológicamente, el término *econometría* significa medición económica y fue mencionado por primera vez por el economista polaco Pawel Ciompa en 1910 bajo el término “Oekonometrie”. Ragnar Frisch era considerado el primer economista en mencionar este término en una publicación científica de 1926, el cual en una nota publicada 10 años más tarde explica que Pawel menciona tempranamente a la econometría desde un aspecto demasiado descriptivo mientras que su aporte se ajusta a una definición teórica, estadística y matemática mucho más cercana a la definición actual del término Frisch (1936).

Si bien los primeros vestigios de lo que es econometría se remontan a 1900, aportes estadísticos y económicos previos a esta fecha han sido fundamentales para desarrollar a la ciencia econométrica en sí. Los primeros aportes teórico-cuantitativos surgen en el siglo XVII por los políticos aritméticos mediante el análisis de las economías nacionales en relación a las balanzas comerciales, ingresos, nivel de precios y demás. Sus primeros enfoques fueron deterministas buscando encontrar leyes del comportamiento económico similares a las de

otras ciencias exactas como la física o la biología, dentro de los principales exponentes de aquel siglo destacan: Gregory King con la estimación de la riqueza mediante el enfoque del ingreso, Charles Davenant con la relación entre el precio del trigo y cantidades cosechadas y finalmente William Petty con el primer cálculo de renta nacional frente al gasto e ingreso nacional bruto y pionero en el uso de gráficos y datos en el análisis económico.

En los siglos posteriores las aportaciones económicas de Quesnay (1758), Cournot (1838) y Juglar (1862) sobre sistemas económicos, oferta-demanda y regularidades temporales de ciclos económicos respectivamente ocasionaron una notable evolución del pensamiento económico y fueron base de las primeras aportaciones econométricas que seguirían siglos más tarde. El ámbito estadístico tuvo grandes contribuciones en el siglo XVII y XVIII con Bayes (1763) y su teorema de probabilidad y análisis estadístico, de igual forma Gauss (1809,1821) supone la normalidad de los errores y restringe los estimadores a funciones lineales y genera así método de mínimos cuadrados y el modelo lineal de Gauss, finalmente Gosset (1908) quien aporta a la estadística la distribución *t de student*, el error probable de una media y los residuos estudentizados.

Paulatinamente las aportaciones económicas y estadísticas se complementaron dando cabida a las aportaciones de Yule (1895) con la correlación simple de la pobreza y Hooker (1901) con la relación entre la tasa de matrimonio y el nivel de prosperidad, además se efectúa la primera regresión múltiple sobre la demanda del café en función del precio del café y del azúcar en 1907 por el italiano Benini, este logro es también atribuido a Moore (1914) con una regresión múltiple sobre la demanda, siguiendo esta misma línea le siguieron Schultz, Hoolbrook y Douglas. Ragnar Frisch es uno de los principales exponentes en este campo y logra consolidar la "*Sociedad Econométrica*" en 1930 la cual estipula:

*"La Sociedad Econométrica es una sociedad internacional para el avance de la teoría económica en su relación con la estadística y las matemáticas.*

*Ragnar Frisch da origen a la Sociedad Econométrica junto con Charles Roos e Irving Fisher en diciembre de 1930 en Cleveland, donde la Asociación Económica Americana, la Asociación Estadística American y la Asociación Matemática Americana celebraban su reunión anual. Joseph Schumpeter presidió el encuentro que fundó la Sociedad Econométrica y eligió a Irving Fisher como su primer presidente. {...} Alfred Cowles ofreció financiación para la revista Econometrica publicada en 1933, con Ragnar Frisch como su editor."*<sup>1</sup>

---

<sup>1</sup> The Econometric Society. Retrieved and translated on December 15, 2015, from <https://www.econometricsociety.org/society/about>

La creación de esta sociedad incentivó a distintos académicos a incursionar dentro de la econometría, rama de conocimiento relativamente nueva en aquella época que se convirtió en una herramienta importante dentro de la búsqueda por respuestas a diversos fenómenos económicos. A partir de este punto la econometría tuvo un impresionante desarrollo y ha contribuido en campos como la economía cuantitativa, teoría económica y diversos estudios empíricos.

Al existir una publicación exclusiva para aportaciones a la econometría, como lo es *Econometrica*, diversos economistas de la época brindaron teorías y métodos básicos como es el caso de Tinbergen (1939) quien fue el primero en desarrollar un modelo de ecuaciones simultáneas para una economía completa, Haavelmo (1944) por su parte brinda un enfoque probabilístico de contrastación de hipótesis e inferencia estadística. Contrastando el estudio estadístico de las teorías económicas Koopmans (1947) menciona que si un modelo estructural no es el adecuado las estimaciones carecerán de sentido al igual que distintos académicos que resaltan los problemas en las estimaciones, las variables y las bases de datos de aquel entonces. Después de la segunda guerra mundial y las crisis ocasionadas por el alza de precios en el petróleo, se evidenció la ineficiencia de ciertos modelos y concepciones económicas, además se cuestionó a la econometría como herramienta en la toma de decisiones dentro de la política económica (Lucas, 1976; Malinvaud, 1981). Frente a estas críticas Portillo (2006) destaca importantes contribuciones de Granger y Newbold (1974) en el estudio de ecuaciones simples de series temporales; en esta misma época se dieron hallazgos relevantes sobre análisis espectral, modelos Box-Jenkins, metodología de vectores autorregresivos y la cointegración.

### **1.1.3. Importancia.**

Desde sus orígenes la econometría ha tomado un importante rol dentro la economía como una herramienta de análisis, en la actualidad es aplicable a distintas áreas de esta ciencia ya que permite la cuantificación de parámetros para explicar distintos hechos en el entorno y para predecirlos. La economía moderna se fundamenta en las diversas aplicaciones de la econometría ya sea en la toma de decisiones de política o corroborando hipótesis y teorías mediante el uso y manejo de datos, esta herramienta ha permitido corroborar o desmentir un sin número de supuestos en función a los datos brindados, ya que si bien un país europeo puede producir bienes bajo la forma de Cobb-Douglas esto no necesariamente se aplicaría a un país en el Norte de África. Las implicaciones económicas que ofrece la econometría van más allá de predicciones de modelos, análisis de políticas públicas o de fluctuaciones económicas, esta herramienta brinda una base de análisis con implicaciones a diversas áreas del conocimiento integrando perspectivas sociales, ambientales comerciales y demás para el



uso de los econométricos en beneficio de la sociedad, asimismo fundamenta la investigación y desarrollo de nuevas teorías y supuestos ante economías tan fluctuantes y evolutivas como las que se presentan en la actualidad.

Hong (2006) señala el reconocimiento e importancia que ha tenido la econometría al haber una mayor demanda de docentes que la impartan, llegando a convertirse en una materia obligatoria para todos los estudiantes de pregrado en economía y un requisito en algunos exámenes de admisión para postgrados. Hendry (1980) menciona que la econometría busca abstraer fenómenos de interés para luego expresarlas de forma matemática como una teoría, los modelos resultantes son analizados bajo el contexto de los datos obtenidos y la teoría económica establecida, esto conlleva a importantes resultados que muestran realidades concordantes con la teoría en algunos casos y en otros una disyuntiva de la misma. De igual forma Hoover (2005) señala cuatro utilidades dentro de la econometría como son: el poner a prueba las teorías preestablecidas, como instrumento de medición de variables desconocidas o no observadas, para la predicción de los valores de una variable y finalmente para caracterizar una relación o fenómeno dentro de la economía.

De manera general se resume la importancia de la econometría como una herramienta económica para todo aquel que busque cuantificar los procesos, fenómenos o realidades que acontecen en un entorno específico bajo una amplia gama de metodologías y procesos que se ajustan a las anomalías y características de los datos que se brindan. Esta ciencia dista de ser perfecta y presenta vacíos en algunos aspectos metodológicos, sin embargo ofrece un sin número de beneficios y aportaciones a la economía moderna.

#### **1.1.4. Metodología.**

La metodología según Boland (1989) es aquel proceso de decisión necesario para llevar las ideas a evidencias, dentro del campo econométrico Hendry (1995) menciona que existen distintos métodos para obtener estimadores como lo son: el método Bayesiano, método no paramétrico, método de momentos y de verosimilitud los cuales brindan distintos enfoques para solución de una problemática, sin embargo Gujarati & Porter (2010) se guía por la metodología clásica de la econometría bajo ocho sencillos lineamientos:

##### **1. Planteamiento de la teoría o hipótesis.**

Se analizan las diversas aportaciones y teorías frente al modelo econométrico a desarrollar, considerando la naturaleza de los datos y el medio en el cual se desarrolla la investigación. De esta manera se tendrá una base comparativa para aseverar o negar con los resultados obtenidos.

## **2. Especificación del modelo matemático de la teoría.**

Una vez planteada la teoría se procede a especificar el modelo matemático, es decir se establece la relación entre los parámetros a utilizar y que variables cuantificarán los efectos que se buscan analizar con el modelo.

## **3. Especificación del modelo econométrico o estadístico de la teoría.**

Se establece el tipo de modelo econométrico a usar, este puede ser por vectores autorregresivos (VAR), un modelo lineal simple, un sistema de ecuaciones o una regresión múltiple de las variables como se muestra a continuación:

$$Y = \beta_1 + \beta_2 X_1 + \beta_3 X_2 + u$$

## **4. Obtención de datos.**

Una vez especificadas las variables dependientes e independientes se procede a la recolección de datos ya sea a partir de fuentes estadísticas oficiales o mediante el levantamiento de información propia, la confiabilidad de los datos es crucial para obtener estimadores que reflejen el contexto verdadero de lo que se planea investigar.

## **5. Estimación de los parámetros del modelo econométrico.**

La estimación numérica de los parámetros se realiza bajo la técnica estadística conocida como análisis de regresión para obtener valores estimados a partir de los datos recolectados.

## **6. Pruebas de hipótesis.**

Mediante la inferencia estadística se comprueba si las relaciones obtenidas y cada uno de los estimadores presentados se ajustan a la teoría económica, las pruebas de hipótesis permiten comprobar ciertas características de las variables y aseverar si estas cumple o no ciertos cánones establecidos dentro de la estadística y la teoría económica.

## **7. Pronóstico o predicción.**

Si el modelo resulta correcto al no refutar ninguna prueba de hipótesis y se mantiene acorde a la teoría es posible predecir los valores futuros de las variables de análisis, esto permite disminuir la incertidumbre de ciertos aspectos en la economía y direccionar medidas de política frente a los posibles resultados establecidos por el modelo econométrico.

## **8. Utilización del modelo para fines de control o de políticas.**

Como se mencionó anteriormente, los gobiernos pueden usar estas estimaciones para saber los efectos que desencadenan variaciones en los impuestos frente al consumo de un bien específico, por ejemplo, o simplemente para comprobar los efectos de medidas de política en sectores específicos de la economía.

## 1.2. Stata

“Stata es un paquete de tratamiento de datos que permite analizarlos y describirlos mediante el uso de métodos estadísticos y gráficos” (Mestre, 1994, p.163). En otras palabras, Stata es un software estadístico completo que brinda al usuario lo necesario para realizar análisis de datos, tablas, y gráficas (MULTION, 2015). A más de ser un instrumento de apoyo para la estadística, tiene una fuerte especialización en problemas asociados con la econometría, ayudando a modelar el comportamiento de la economía en general. Este software estadístico se diferencia de la mayoría en un aspecto clave: la forma en que los usuarios contribuyen a un lenguaje extensible. Los usuarios pueden aportar con sus propios algoritmos que al añadirlos al software ayudan a responder necesidades de investigadores. Una comunidad que contribuye a Stata directa o indirectamente impulsa a un circuito de retroalimentación poderosa, haciendo que Stata crezca de manera continua (Cox, 2005). Cabe mencionar que, Stata está disponible para sistemas Windows, Macintosh, Unix, y algunas variaciones de éstos.

### 1.2.1. Historia.<sup>2</sup>

Según Cox (2005) Stata se inició en California a mediados de la década de 1980. En 1984 Willian Gould era dueño de un pequeño negocio de computadoras llamado Centro de Recursos para la Computación (CRC) con sede en Santa Mónica. Gould se dio cuenta de que varias personas estaban escribiendo programas estadísticos directamente destinados a los ordenadores personales, tales como Sysat o MicroTSP. Por ello, creyó que con ayuda podría escribir un programa que no se compare con ninguno en el mercado. Consecuencia de ello, nació Stata que en ese tiempo se llamaba “Dial”. Este nombre fue cambiado a Stata poco antes de su lanzamiento en 1985.

Willian Gould y Sean Beckett demoraron un año en escribir la primera versión de Stata. Ello fue escrito en el lenguaje de programación C. Stata 1.0 era un pequeño programa que no cubría con los aspectos estadísticos principales, más de lo que sus competidores lo hacían, pues solamente contaba con 44 comandos para gestión de datos. Esta versión era ejecutable en el sistema operativo DOS. Además esta versión tenía errores, por lo cual surge Stata 1.1 para corregir estos errores, pero la versión 1.2 trajo un nuevo sistema de menús y mejor ayuda en línea. Posteriormente, Stata 1.3 trajo gráficos reales a través de Stata/Gráficos.

Hacer Stata extensible, de manera que los usuarios pueden añadir sus propios comandos, y así el software empiece a crecer a partir de un paquete en un idioma, fue uno de los pasos importantes de su historia. En 1986 Stata 1.4 poseía nueva documentación, y el primer

---

<sup>2</sup> Cox, N. J. (2005). A brief history of Stata on its 20th anniversary. *Stata Journal*, (1), 2–18

contenido oficial de programación. Luego, en 1987 Stata 1.5 presentó adiciones muy notables, que fueron ANOVA, logit y probit. En 1988 se presenta el primer Stata ejecutable para Sun/Unix. Posteriormente para otras variantes de Unix. En 1992, sale la primera versión para Macintosh.

El primer libro de Stata fue “Estadística con Stata” por Lawrence Hamilton (1990). Entre muchos libros posteriores y empresas de enseñanza uno que debe ser mencionado es StataQuest con libros escritos por Ted Anagnoson y Richard De León. En 1990 se presentó Stata 2.0 que era una nueva versión con los gráficos en el Stata ejecutable con variables de cadena, superficies de análisis de supervivencia (Cox y Kaplan-Meier) y regresión por pasos. Stata 2.05 dio un nuevo formato. Y Stata 2.1 hizo otras adiciones principales, tales como variables de bytes, análisis factorial, y ado-files, distribuido previamente en un kit a integrarse en Stata.

Copias de Stata se vendieron en varios países fuera de Estado Unidos. El primer distribuidor fue Peter Hedsrom de Metrica Consulting en Suecia en 1990, y Ana de Timberlake de Timberlake Consulting en Gran Bretaña fue el segundo en 1991. Stata se convirtió en el producto de línea principal.

Otra iniciativa en 1991 fue la introducción del Boletín Técnico Stata (BTS), como una publicación emitida seis veces al año. Actualmente continúa como Stata Journal. Este sirvió como un medio para la publicación de programas extra-oficiales entre lanzamientos, reemplazando los kits y discos de apoyo, pero principalmente sirvió para que los usuarios hagan públicos sus propios programas y así formalizar la existencia de una auténtica comunidad de usuarios.

En 1992 Stata 3.0 amplía el corpus de comandos para modelos de regresión para delimitadas y múltiples variables limitadas, categóricas, dependientes, robusta regresión cuantil, y el apoyo a las distribuciones no normales. Este conjunto de comandos fue el motor para el modelado basado en máxima verosimilitud. Un año más tarde la empresa se convirtió en StataCorp LP. En 1994 se creó un servidor de listas Statalist iniciado por David Wormuth, en donde debaten temas o aspectos relacionados con Stata.

En 1995 Stata 4.0 introdujo comandos para datos de panel, series de tiempo, y análisis de conglomerados. Además vio un primer puerto a Microsoft Windows. Más tarde se añadió un comando ssc a Stata basado en programa por Baum y Cox, dando así a los usuarios una forma de instalar programas directamente sin la intermediación de un navegador.

El sitio web: [www.stata.com](http://www.stata.com) fue lanzado en 1996, para cubrir con las quejas y necesidades de los usuarios. Desde entonces se ha convertido en el centro de la empresa. Meses más

tarde Stata 5.0 introdujo nuevos comandos al programa. Pero no es hasta inicios de 1999 con el Stata 6.0 que se añaden comandos de gran utilidad para datos de series de tiempo como son arima, arch, entre otros. Y en el año 2000 Stata 7.0 añadió aún más comandos.

En el 2001 el Boletín Técnico Stata (BTS) fue lanzado como Stata Journal. En este mismo año se vio la primera versión de 64bits para Solaris, Linux, Windows, y OS X. Otro ajuste a los tamaños más grandes era la liberación de Stata/SE, con soporte para conjuntos de datos grandes.

En el 2003 Stata 8 fue el lanzamiento más grande en la historia de Stata, debido a la gran masa de códigos y documentación. Éste contó con un sistema completamente nuevo de gráficos con una sintaxis más complicada, y la introducción de diálogos, haciendo a los diálogos compatibles con el lenguaje de comandos. El mismo año, Stata desde la versión 8.1 ya permitía una actualización gratuita. Meses después Stata 8.2 presentó cambios sustanciales en gráficos, y adiciones a las series tiempo como los llamados VEC s que son modelos de vectores de corrección de error.

A partir del 2005 Stata ha evolucionado a pasos agigantados, con características especiales como mejor procesamiento de datos de panel, realiza operaciones con números reales o complejos, incorpora métodos multivariantes, modelos VAR, resultados binarios y conteo continuo, cadenas largas, mejoras en datos de panel, post-estratificación, diseños y modelos multinivel, funciones impulso respuesta, más comandos, más documentación, entre otros avances. Actualmente, la última versión disponible es Stata 14.1 lanzada en octubre del 2015. Se han lanzado 32 versiones hasta el momento, de las cuales solamente 9 versiones tienen una actualización descargable gratuita.

**Tabla 1.1.** Evolución de Stata

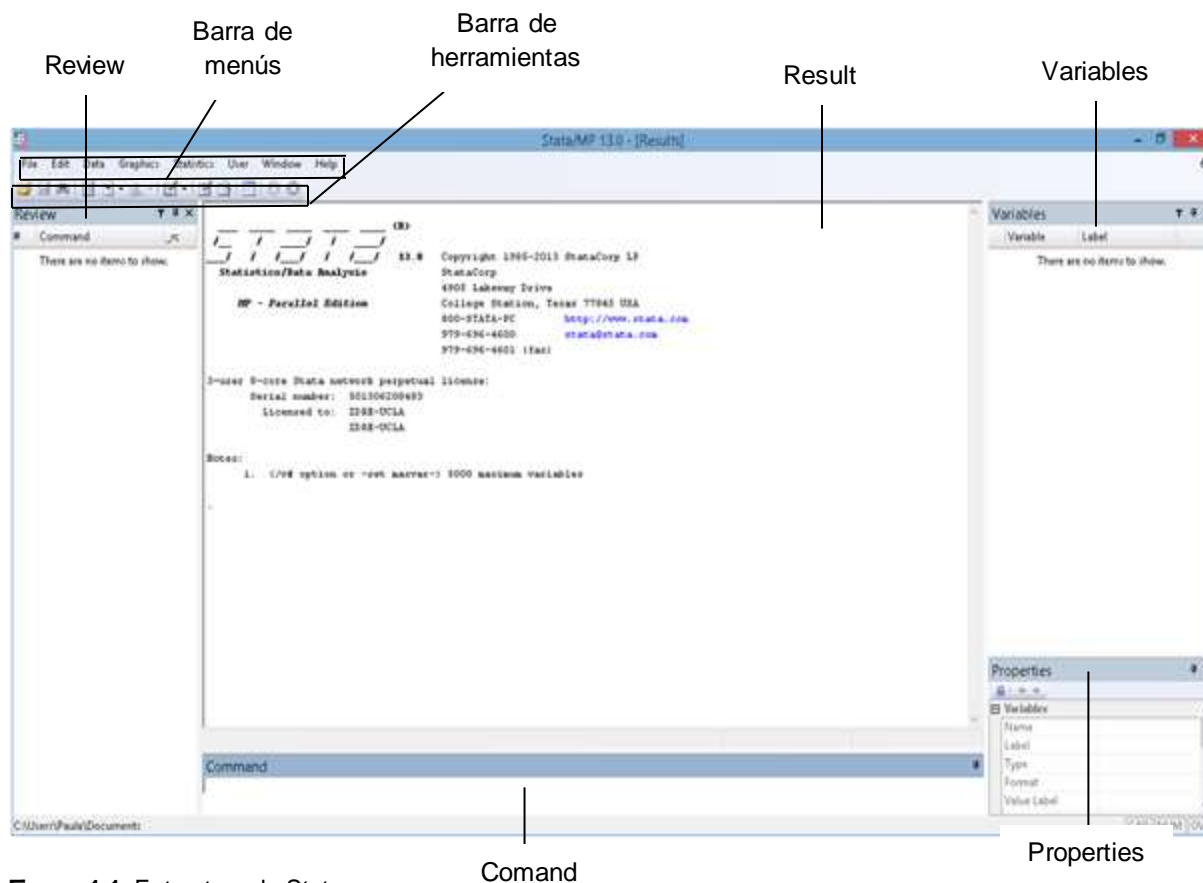
<b>Versión</b>	<b>Fecha</b>	<b>Actualización descargable gratuita</b>
14.1	octubre 2015	NO
14.0	abril 2015	NO
13.1	octubre 2013	SI
13.0	junio 2013	NO
12.1	enero 2012	SI
12.0	julio 2011	NO
11.2	marzo 2011	SI
11.1	junio 2010	SI
11.0	julio 2009	NO
10.1	agosto 2008	SI
10.0	junio 2007	NO
9.2	abril 2006	SI
9.1	septiembre 2005	SI
9.0	abril 2005	NO
8.2	octubre 2003	SI
8.1	julio 2003	SI
8.0	enero 2003	NO
7.0	diciembre 2000	NO
6.0	enero 1999	NO
5.0	septiembre 1996	NO
4.0	enero 1995	NO
3.1	agosto 1993	NO
3.0	marzo 1992	NO
2.1	agosto 1990	NO
2.05	abril 1989	NO
2.0	junio 1988	NO
1.5	febrero 1987	NO
1.4	mayo 1986	NO
1.3	agosto 1985	NO
1.2	mayo 1985	NO
1.1	febrero 1985	NO
1.0	enero 1985	NO

**Fuente:** Gould, W., & Cox, N. (2015). Stata | FAQ: History of Stata. Recuperado de <http://www.stata.com/support/faqs/resources/history-of-stata/>

**Elaboración:** Las autoras

### 1.2.2. Estructura.

La ventana principal de Stata 13, que aparece al abrir el programa, está compuesta por cinco ventanillas, además de la barra de menús y la barra de herramientas.



**Figura 1.1.** Estructura de Stata

**Fuente:** Programa Stata 13

**Elaboración:** Las autoras

**Comand.** También llamada línea de comando. Esta ventana permite escribir las órdenes que el usuario quiere que sean ejecutadas por el programa. (Muñoz, s.f.).

**Review.** En esta ventana se archivan las órdenes ejecutadas por el usuario durante una sesión de trabajo en Stata. Con ello, facilita repetir las operaciones ya realizadas, y recuperar comandos para introducir modificaciones (Muñoz, s.f.).

**Variables.** Muestra el listado de variables que contiene la base de datos, así como también sus etiquetas. Esta ventana facilita la selección de las variables mediante doble clic, sin necesidad de escribirlas (Muñoz, s.f.; Rojo, 2008).













**Properties.** Muestra información acerca de las variables como nombre, etiqueta, tipo, formato, tamaño, entre otras propiedades.

**Result.** En esta ventana aparecen los resultados de los comandos ejecutados. Es decir, el programa a través de esta ventana responde a las órdenes de usuario (Muñoz, s.f.).

**Barra de menús.** Se encuentra en la parte superior de la pantalla. A través de sus opciones se puede abrir una variedad de cuadros de diálogo que permiten realizar parte de los procedimientos de Stata (Muñoz, s.f.). Esta barra consta con las opciones: file, edit, data, graphics, statistics, user, windows, y help.

**Barra de herramientas.** Cuenta con 12 íconos con distintas aplicaciones. La descripción de los mismos se encuentra en la tabla 2.

**Tabla 1.2.** Descripción de barra de herramientas

Ícono	Descripción/ Función
	Abrir archivos de datos en formato Stata (.dta).
	Guardar archivo actual de datos.
	Imprimir los resultados, gráficos y órdenes.
	Empezar, cerrar, suspender, y/o reanudar un archivo en formato log (.log; .smcl)
	Abrir el visor de ayuda.
	Traer una ventana de gráfico al frente de todas las demás ventanas, es decir a primer plano.
	Crear un nuevo archivo do (.do). En estos archivos se puede crear, modificar, abrir, ejecutar y guardar rutinas de comandos (sintaxis) de Stata.
	Modificar la base de datos.
	Visualizar la base de datos.
	Brindar información de la base de datos como: variable, etiqueta, tipo, formato, y notas.
	Continuar cuando se ha hecho una pausa en el medio de una larga salida.
	Detener la tarea actual en Stata.

**Fuente:** Rojo, J. (2008). Curso de introducción al paquete Stata. Centro de Ciencias Humano Y Sociales. Recuperado de <http://investigadores.cide.edu/aparicio/data/IntroduccionStata9.pdf>

**Elaboración:** Las autoras



### 1.2.3. Tipos de archivos.

Los tipos de archivos básicos en Stata son:

**Datos (.dta).** Según Muñoz (s.f.) “Los datos utilizados en Stata se guardan en formato .dta, que es el formato propio de Stata” (p.5). Sin embargo el programa si está en la capacidad de importar datos de diversos formatos, tales como Excel, SPSS, ASCII, CSV, etc. Los archivos .dta son la base de datos de Stata (Fuentes & Palma, 2003).

**Sintaxis (.do).** Los archivos de sintaxis o rutinas de comandos se guardan en formato .do, es por ello que se los conoce como do-files. Estos archivos brindan principalmente la facilidad de conservar las rutinas de nuestro trabajo, las mismas que pueden ser editadas de acuerdo a las necesidades del usuario en cualquier momento.

**Output (.log).** Si se quiere conservar los resultados se lo puede hacer en un ‘log file’. Podemos elegir los formatos .log y .smcl, aunque el más recomendable es .log, que no tiene formato y garantiza compatibilidad con otros editores de textos.

**Extensiones (.ado).** Son uno de los puntos fuertes de Stata. Los usuarios, con relativa facilidad, pueden escribir funciones y comandos nuevos y distribuirlos gratuitamente. Según Muñoz (s.f.) “la potencia del programa se multiplica, creciendo de manera continua, ya que allí donde la distribución oficial no llega, sí lo hacen los usuarios” (p.6). Por lo tanto, estos archivos contienen códigos que están incluidos en el paquete de Stata.

**Help (.hlp).** Son archivos de ayuda, que brindan información relevante para guiar al usuario en una serie de inconvenientes ocasionados por su desconocimiento.

**Gráficos (.gph).** Los gráficos hechos en Stata se los puede guardar en archivos .gph, brindando la capacidad de conservar gráficos hechos en determinada sesión de trabajo en Stata (Jiménez-Martín, 2001).

**Diccionario (.dct).** Son archivos de instrucciones para leer datos ASCII (útil para simulaciones).

### 1.2.4. Uso actual.

“Stata es distribuido en más de 150 países y es usado por profesionales en muchos campos de investigación” (MULTION, 2015, p.1) como economía, ciencia política, ciencia ambiental, biometría, sociometría, etc (Stata, 2009). Su frecuente uso en la investigación se debe a su amplia gama de capacidades estadísticas. Stata mediante técnicas estadísticas es capaz de gestionar y manejar datos, tablas, y gráficos. Con ello los investigadores consiguen evidencia clara para el análisis e interpretación de resultados de sus respectivas investigaciones.

También es utilizado en el ámbito académico. A menudo programas de estudio incluyen el aprendizaje de softwares que complementen la formación académica de los estudiantes. Stata es uno de los programas que se cree pertinente que, tanto docentes como estudiantes, lo manejen por su rápido procesamiento de datos, exactitud, fácil manejo, y gran uso investigativo.

A través de Stata se puede dar tratamiento a los datos como en la transformación de datos y construcción de índices, manejar distintos tipos de variables, estimar modelos econométricos, realizar gráficas e inclusive mapas en calidad de publicación. Una característica importante de mencionar de este programa es que facilita el análisis de los datos de encuestas, tanto de fuentes de datos oficiales como de las elaboradas por los mismos investigadores. Este es un aspecto muy útil para la investigación cuantitativa de cualquier campo investigativo.

### **1.3. Econometría de Gujarati & Porter**

Damodar N. Gujarati, profesor emérito de economía en la Academia Militar de Estados Unidos y Dawn C. Porter docente de la Universidad del Sur de California son autores del libro “Econometría” en su quinta edición, cuya página oficial describe al texto de la siguiente manera:

*“Econometría de Gujarati y Porter ofrece una introducción elemental pero completa de la econometría al carecer de álgebra matricial, cálculo o estadísticas más allá del nivel elemental. Con la incorporación de más de 100 nuevas bases de datos, así como investigación actualizada y ejemplos, la quinta edición responde a importantes desarrollos en la teoría y práctica de la econometría. Econometría es ampliamente utilizada por los estudiantes de todos los campos debido a que los temas y aplicaciones dentro del texto se aplican a una amplia gama de estudios.”*

Este libro aborda cuatro temas principales dentro de su contenido, como son los modelos de regresión uniecuacionales, flexibilización de los supuestos del modelo clásico, temas de econometría y finalmente los modelos de ecuaciones simultáneas y econometría de series de tiempo. Dentro de estos apartados se aborda de manera concisa y breve las principales aportaciones econométricas desde los orígenes de esta ciencia hasta las metodologías de resolución más actuales y relevantes. La quinta edición de “Econometría” es el resultado de años de investigación y la renovación de teoría y aplicaciones que comenzó en el año de 1978 en donde Gujarati fue el único autor, en 1995 inicia una colaboración con Dawn Porter y se presenta la tercera edición de este libro, la última y quinta edición se publica en el 2005 y es traducida al español, francés, portugués, coreano, chino y turco en los años posteriores.

Sobre los autores se puede destacar que Gujarati es un economista de nacionalidad indio-americano con estudios realizados en la universidad de Bombay y Chicago, posee una

extensa trayectoria en el campo de la docencia y la econometría al enseñar por 28 años en la Universidad de New York y luego en la Academia Militar de Estados Unidos dentro del Departamento de Ciencias Sociales. Se le atribuyen reconocidas publicaciones dentro de las revistas más importantes del medio económico tales como el “Review of Economic and Statistics” “The Economic Journal”, “Journal of Financial and Quantitative Analysis”: en adición a diversos libros como: “Pensiones y la Crisis Fiscal en Nueva York”, “Gobierno y Negocios” y finalmente “Econometría” la cual fue traducida a varios idiomas en sus cinco ediciones y es tomada como bibliografía principal para la presente tesis. Por su parte Dawn C. Porter es una estadista cuyas áreas de conocimiento se enfocan al análisis categórico, modelos multivariados y aplicaciones dentro de la psicología, además es docente colaboradora en instituciones como la Universidad de California, Universidad de Georgetown y la Universidad de Nueva York, es coautora del libro “Fundamentos de Estadística Empresarial”, posee más de diez publicaciones científicas y es consultora en empresa como Ginnie Mae, Inc; Toys R Us Corporation; IBM; Cosmaire, Inc y la Universidad de Nueva York. En 2011 recibió el Golden Apple Award y el premio a la enseñanza de Evan C. Thompson además de varias nominaciones para estos mismos galardones en años previos.

### **1.3.1 Como relacionamos el libro con el Stata.**

Al existir un libro de econometría que contenga teoría detallada y aplicación de la misma mediante numerosos ejercicios, como lo es el libro de Damodar Gujarati & Dawn Porter, este recurso es además uno de los libros en español sobre enseñanza econométrica más utilizados por los estudiantes según Osuna (2015), volviéndose referencia casi obligatoria para todo aquel que inicie en este campo de conocimiento.

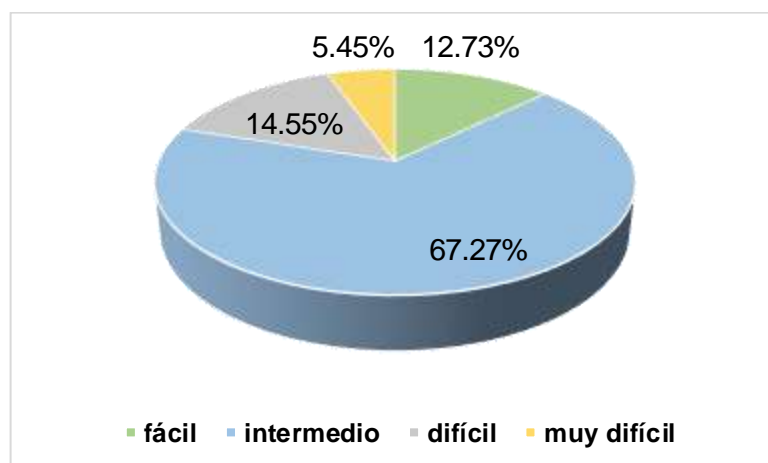
La presente tesis se busca replicar aquellos conocimientos mediante la aplicación de ejercicios enfocados a la realidad ecuatoriana utilizando el software estadístico Stata, para cumplir con este cometido se estructuraron gran parte de los temas presentados en el libro de Gujarati de manera que se apliquen todos los contenidos que este brinda con cada ejercicio presentado. Al mantener la estructura anteriormente descrita los estudiantes podrán usar la presente como una guía de resolución de ejercicios de manera que asimilen teoría y aplicación de ejercicios para una realidad latinoamericana, dado que gran parte de los recursos bibliográficos que se brindan en las aulas de clase poseen aplicaciones que no siempre concuerdan con el entorno local

### **1.4. Dificultades en el uso de Stata en econometría**

Con el fin de averiguar las dificultades que se presentan en el uso del software estadístico Stata en econometría, se ha realizado un estudio exploratorio a los estudiantes de

econometría de la titulación de economía de la “Universidad Técnica Particular de Loja” en el período académico octubre 2015 - febrero 2016. Dicho estudio se lo realizó con información primaria, mediante la aplicación de encuestas, a un total de 55 estudiantes, representando a la totalidad de estudiantes matriculados en econometría (componente de séptimo ciclo de la carrera). De los cuales 37 estudiantes son de género femenino y 18 de género masculino.

La encuesta aplicada (Anexo 1) está direccionada a descubrir el grado de dificultad que tuvieron los estudiantes en el uso de Stata, las razones por las cuales les resultó difícil el uso del mismo, la necesidad de aplicar datos de Ecuador en la aplicación de sus conocimientos de econometría, saber si creen importante que se elabore una guía pedagógica de econometría utilizando el software Stata que utilice datos del Ecuador, y principalmente determinar la metodología de resolución de ejercicios de econometría utilizando Stata más adecuada, clara, y entendible para ellos.



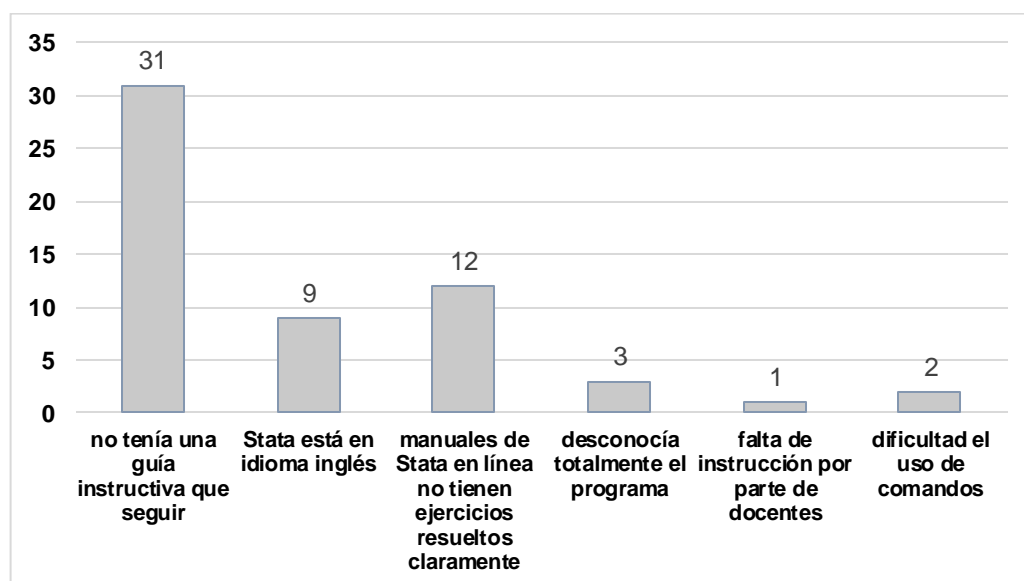
**Gráfico 1.1.** Grado de dificultad del uso de Stata

**Fuente:** Estudio exploratorio

**Elaboración:** Las autoras

Como se muestra en el gráfico 1.1, al preguntarles a los estudiantes el grado de dificultad que tuvieron al empezar a utilizar Stata en econometría, el 67.27% respondió que tuvo un grado de dificultad intermedio, el 14.44% lo consideró difícil, el 12.73% lo consideró fácil, y el 5.45% mencionó que les resultó muy difícil el uso del mismo. Como el manejo del software estadístico Stata por lo general se lo considera un programa sencillo y muy interactivo, por lo tanto, el empezar a utilizarlo por primera vez no ocasiona un alto grado de dificultad. Sin embargo, si no hay una correcta metodología de enseñanza puede llegar a convertirse en una dificultad. Esto también depende de las habilidades que tenga el estudiante de aprender un paquete informático, y el nivel intuitivo que tiene en el uso del mismo. Esto es importante, ya que la sintaxis de Stata es en gran parte intuitiva.

Las razones por la cuales los estudiantes tienen un cierto grado de dificultad en el manejo de Stata se presentan en el gráfico 1.2.



**Gráfico 1.2.** Razones de dificultad del uso de Stata

**Fuente:** Estudio exploratorio

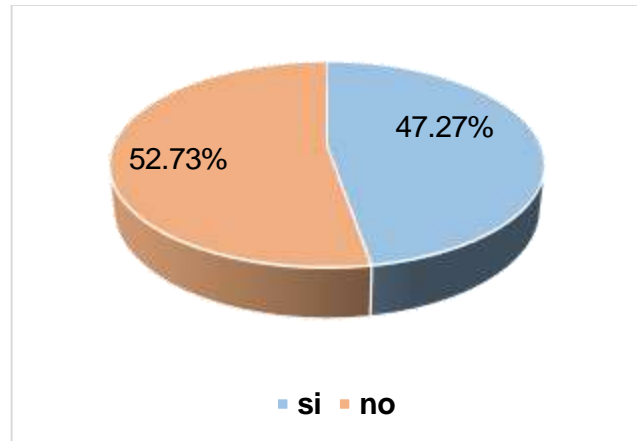
**Elaboración:** Las autoras

Los estudiantes señalaron con mayor frecuencia (31 veces) la razón de que no tenían una guía instructiva que seguir. La segunda razón más mencionada (12 veces) es que los manuales de Stata en línea no tienen ejercicios resueltos claramente. Y la tercera (9 veces) es que Stata está en idioma inglés. Siendo estas las tres principales razones que impiden a los estudiantes tener una mayor facilidad en el manejo del programa.

Con una guía que brinde instrucciones claras y entendibles, resultaría más fácil aprender un paquete informático como lo es Stata. Esto se debe a que, al iniciar con el manejo de un nuevo programa, los estudiantes se encuentran desorientados, y la ayuda que brindan los maestros no es suficiente para ellos, ya que los docentes ocupan la mayor parte del tiempo a enseñar la materia de econometría teóricamente y cuando van a práctica manejan del software rápidamente. Por consiguiente, es trabajo de cada estudiante, adquirir las competencias necesarias para lograr un manejo eficiente de Stata. Habrá quienes entiendan con mayor facilidad el uso de Stata, pero para otros les será más difícil. Lo ideal sería que todos los estudiantes aprendan a manejar Stata en econometría. Por lo que una guía instructiva ayudaría mucho en ello.

Los estudiantes podrían acudir a manuales que brinda Stata en línea. Sin embargo, la mayoría de los manuales se encuentran en idioma inglés, lo cual también es un impedimento para el entendimiento del estudiante. Si bien es cierto hay manuales en línea en idioma español, los

estudiantes consideran que no contienen ejercicios resueltos claramente. Ellos en muchos de los casos carecen de conocimientos básicos en el manejo de Stata, lo que imposibilita el entendimiento de ejercicios más complejos.

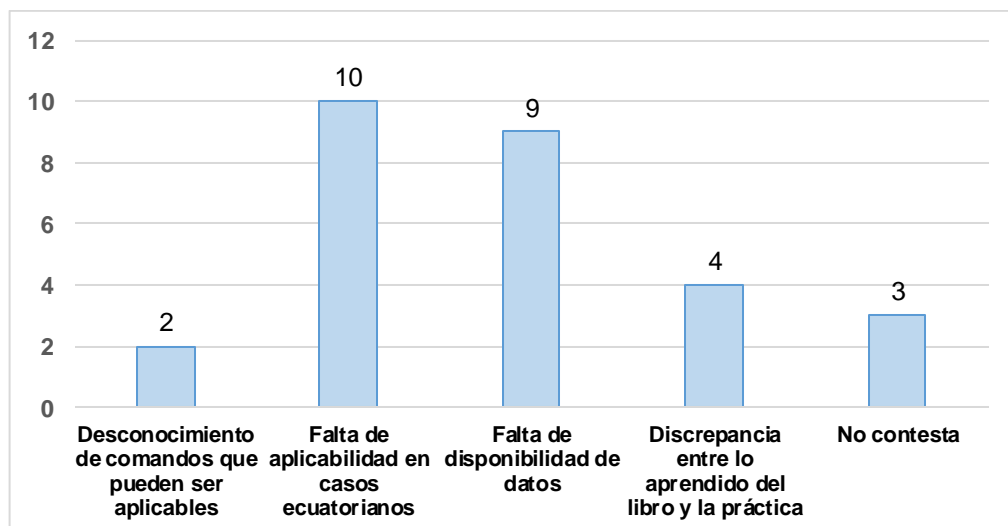


**Gráfico 1.3.** Dificultad en la creación de modelos con datos de Ecuador

**Fuente:** Estudio exploratorio

**Elaboración:** Las autoras

A través del gráfico 3 se puede observar que a un poco más de las mitad de los estudiantes 52.73%, no presentan dificultad en la creación de modelos econométricos con datos de Ecuador, y un poco menos de la mitad de los estudiantes, 47.27%, sí presentan dificultad en la creación de dichos modelos. Pero, ¿por qué casi la mitad tiene problemas en ello?. Para dar respuesta a esto se analiza el gráfico 1.4.



**Gráfico 1.4.** Razones de dificultad en la creación de modelos con datos de Ecuador

**Fuente:** Estudio exploratorio

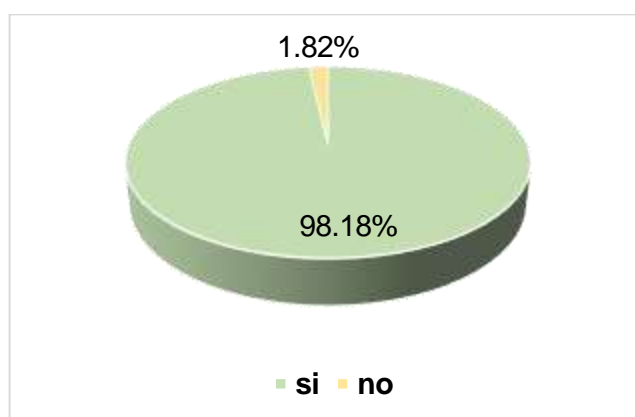
**Elaboración:** Las autoras

La razón predominante es la falta de aplicabilidad en casos ecuatorianos. Es decir, la mayoría de modelos econométricos que se realizan en las aulas, se elaboran con casos

internacionales, haciendo poco uso de datos de Ecuador. Cabe mencionar que el libro base que utilizan los estudiantes de econometría es el de Gujarati & Porter (2010), el cual contiene en su mayoría casos de la economía estadounidense.

La segunda razón es la falta de disponibilidad de datos de Ecuador. Muchas de las veces no hay datos suficientes para realizar los modelos econométricos que se quiere. Sin embargo, no siempre es así, ya que a menudo los estudiantes desconocen algunas fuentes de datos oficiales en las cuales podrían encontrar la información necesaria. O simplemente no saben cómo y en dónde encontrar la información requerida.

Como tercera razón se tiene la discrepancia entre lo aprendido del libro y la práctica. Los estudiantes al crear modelos con datos de Ecuador frecuentemente se encuentran con resultados no esperados, por la razón de que todas las economías son distintas, y por ende, modelos que se ajusten al caso de Estados Unidos, no siempre se ajustan al caso de Ecuador. Por lo tanto, esto produce dudas y confusiones con lo aprendido. Y como última y cuarta razón se tiene el desconocimiento de comandos que puedes ser aplicables. Aquí volvemos a los mismo mencionado anteriormente, acerca de la dificultad que presentan algunos estudiantes en el uso de Stata en la construcción de modelos econométricos.

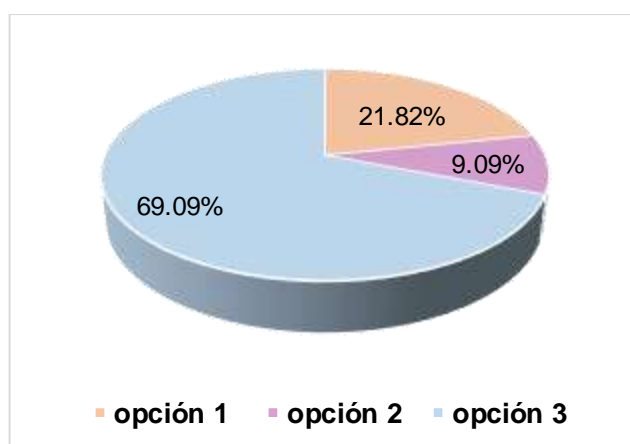


**Gráfico 1.5.** Necesidad de una guía pedagógica de econometría usando Stata  
**Fuente:** Estudio exploratorio  
**Elaboración:** Las autoras

Luego de conocer las dificultades de los estudiantes de econometría tanto como en el uso de Stata como en la creación de modelos usando datos de Ecuador. Se cree pertinente saber si ellos consideran necesaria la creación de una guía de econometría utilizando el software Stata que utilice datos del Ecuador. El gráfico 5 indica que el 98,18% de los estudiantes sí consideran necesaria la creación de dicha guía, mientras que solamente el 1,82% no la considera necesaria.

Indudablemente, de acuerdo con los resultados del presente estudio exploratorio, la creación de una guía que contenga ejercicios resueltos claramente usando Stata y datos del Ecuador

es una necesidad para los estudiantes. Por lo tanto, es importante determinar qué metodología es la más adecuada para la resolución de ejercicios de econometría, en términos explicativos y de proceso. Para ello se presentaron tres diferentes opciones (anexo 1, preg.8). La primera, explica el procedimiento a seguir mediante el uso de las opciones de la barra de menús, en la cual presenta imágenes del proceso paso a paso, y también hace uso de los comandos para obtener lo deseado. La segunda, explica el procedimiento en una versión más corta que la primera, es decir sin necesidad de introducir muchas imágenes. Y la tercera, empieza con una parte introductoria de lo que se va a desarrollar, seguida del listado de comandos a utilizar, explicando el uso de cada uno de ellos, luego consta del desarrollo del ejercicio, y por último una breve interpretación de los resultados.



**Gráfico 1.6.** Metodologías de resolución de ejercicios econométricos  
**Fuente:** Estudio exploratorio  
**Elaboración:** Las autoras

El gráfico 6 indica que la opción metodológica con mayor acogida para la resolución de ejercicios econométricos de la guía pedagógica es la opción 3, con el 69.09% de acogida. Seguida de la opción 1 con el 21.82%, y por último la opción 2 con el 9.09%. Por consiguiente, la guía pedagógica debe contener ejercicios realizados con la opción metodológica número 3, la cual resulta la más adecuada para que el lector entienda y aprenda correctamente.

Los resultados de este estudio exploratorio validan la pertinencia del presente trabajo, el mismo que aplica la econometría a la realidad ecuatoriana, mediante la resolución de una variedad de ejercicios, haciendo uso del software estadístico Stata. La descripción de los datos y modelos empleados se muestran en el capítulo II sobre metodologías, en donde se detallan la naturaleza y características de los datos utilizados para el desarrollo de los ejercicios econométricos presentes en el capítulo III.



## **CAPÍTULO II**

### **DATOS Y METODOLOGÍA**

## 2.1 Presentación de datos

Los datos utilizados en los ejercicios econométricos se obtuvieron de distintas fuentes, tales como encuestas y fuentes oficiales de información como: Banco Mundial, Banco Central del Ecuador, Instituto Nacional de Estadísticas y Censos – INEC entre otros. Dichos datos fueron agrupados en las áreas de economía, socio demografía y ambiente en adición a algunas variables de origen hipotético y de encuestas que son descritas en las tablas subsecuentes.

### 2.1.1 Descripción de variables económicas.

Dentro de esta categoría se encuentran 58 variables correspondientes a bases de datos del Banco Central del Ecuador (BCE), en periodos de tiempo desde 1993 hasta el 2015 y segregado a nivel regional y nacional, Instituto Nacional de Estadísticas y Censos (INEC) con diversas encuestas realizadas en los años de 2010 y 2013, Banco Mundial con las tablas de Indicadores Desarrollo Mundial (WDI) para Ecuador y el mundo en diversos periodos que van desde el año 1960 hasta el 2014.

**Tabla 2.1.** Variables económicas utilizadas

Fuente	Año	Variable	Denominación	Naturaleza
Banco Central – Cuentas Regionales	2010	Valor agregado bruto cantonal	VAB	Cuantitativa
	1993-2012	Tasa de crecimiento del VAB	tc	Cuantitativa
		Logaritmo de la productividad (VAB/PEA)	lprod	Cuantitativa
Banco Central del Ecuador- Tablas de Oferta y Utilización Final	2002-2015	Producto Interno Bruto	PIB	Cuantitativa
		Gasto de consumo final total	GCF	Cuantitativa
		Formación bruta de capital fijo	FBKF	Cuantitativa
		Exportación de bienes y servicios	EXP	Cuantitativa
BCE – Cuentas Trimestrales	2000 I-2015 IV	Gasto de consumo final de hogares	CONH	Cuantitativa
INEC- CENEC	2010	Empresas exportadoras, publicas y privadas	EXP	Cuantitativa
		Producción cantonal	produccion	Cuantitativa
		Dicótoma de polos cantonales	DP	Cualitativa
		Ingresos	ING	Cuantitativa
		Número de empresas	EMP	Cuantitativa
		Personal remunerado	L	Cuantitativa
		Valor de activos fijos	K	Cuantitativa
INEC-Encuesta de Comercio	2013	Producción total	PT	Cuantitativa
INEC-Encuesta de Edificaciones	2013	número de unidades de edificación	UE	Cuantitativa
		total de pisos	TP	Cuantitativa
		valor total del terreno	VTT	Cuantitativa
		valor total de la edificación	VTE	Cuantitativa

**Fuente:** Banco Central del Ecuador, Instituto Nacional de Estadísticas y Censos, Banco Mundial.

**Elaboración:** Las autoras.

**Tabla 2.1.** Variables económicas utilizadas (*Continuación*)

Fuente	Año	Variable	Denominación	Naturaleza
INEC- Encuesta hoteles y servicios	2013	Producción total	PT	Cuantitativa
		Participación en las utilidades	U	Cuantitativa
		Publicidad	P	Cuantitativa
INEC- Estadísticas de recaudación	2010	Impuestos cantonales	IMP	Cuantitativa
Banco Mundial- WDI Ecuador	1960-2011	Pib per cápita	pibp	Cuantitativa
	1960-2014	Población Total	PT	Cuantitativa
		Año	year	Cuantitativa
		Gasto de consumo total	consumo	Cuantitativa
			cons_tot	Cuantitativa
		Gasto de consumo de hogares	cons_hog	Cuantitativa
		Producto interno bruto	PIB	Cuantitativa
		Exportaciones de bienes y servicios	EXP	Cuantitativa
		Pib per cápita	PIBPC	Cuantitativa
	1965-2014	Producto interno bruto	PIB	Cuantitativa
		Formación bruta de capital fijo	FBKF	Cuantitativa
	1966-2013	Crecimiento del pib de Ecuador	PIB_por_ecu	Cuantitativa
		Crecimiento del pib de Estados Unidos	PIB_por_eu	Cuantitativa
		Crecimiento de la FBKF de Ecuador	FBKF_por_ecu	Cuantitativa
		Crecimiento de la FBKF de Estados Unidos	FBKF_por_eu	Cuantitativa
	1970-2014	Exportaciones de bienes y servicios	EXP	Cuantitativa
		Producto interno bruto	PIB	Cuantitativa
		Comercio de mercaderías	CM	Cuantitativa
	1972-2011	Petróleo en miles de barriles	Petroleo	Cuantitativa
		Producto interno bruto	PIB	Cuantitativa
	1976-2013	Industrialización, valor agregado	IND	Cuantitativa
		Gasto en educación	GE	Cuantitativa
		Inversión extranjera directa	IED	Cuantitativa
		Índice de precios al consumidor	IPC	Cuantitativa
		Rentas del petróleo (% del PIB)	RPET	Cuantitativa
		Pib per cápita	PIBPC	Cuantitativa
	1980-2014	Producto interno bruto	pib	Cuantitativa
			PIB	Cuantitativa
	1984-2013	Pib per cápita	pibp	Cuantitativa
	1988-2013	Ingreso nacional neto ajustado	IN	Cuantitativa
		Trabajadores asalariados, total	TA	Cuantitativa
		Gasto de consumo final de los hogares per cápita	GC	Cuantitativa
Banco Mundial – WDI Mundo	1960-2013	Cantidad de dinero	M2	Cuantitativa
		Tiempo	anio	Cuantitativa

**Fuente:** Banco Central del Ecuador, Instituto Nacional de Estadísticas y Censos, Banco Mundial.

**Elaboración:** Las autoras.

La descripción de cada variable económica utilizada dentro de los ejercicios econométricos es la siguiente:

- Valor Agregado Bruto

Es una variable de corte cantonal y de naturaleza cuantitativa, es tomada de las cuentas nacionales del año 2010 del BCE y su valor está expresado en dólares americanos.

- Tasa de crecimiento del VAB

Es una variable cuantitativa de corte temporal y espacial para los años de 1993 al 2012 aplicada a 22 provincias del Ecuador. Es obtenida a partir de las cuentas regionales de Banco Central del Ecuador.

- Logaritmo de la productividad.

Esta variable cuantitativa es el resultado de la división del Valor Agregado Bruto sobre la Población económicamente activa expresada en logaritmos, al igual que la variable anterior es de corte temporal para los años de 1993 al 2012 dentro de 22 provincias del Ecuador.<sup>3</sup>

- Producto Interno Bruto

Se utiliza seis valores diferentes del PIB para el desarrollo de los ejercicios econométricos, el primero de estos es obtenido de las tablas de oferta y utilización del BCE para el periodo 2002-2015 los cuales se expresan en dólares americanos, en segunda instancia se toman los datos del Banco Mundial expresados a precios constantes del años 2005 durante los periodos: 1960-2014 (dólares); 1965-2014 (miles de millones de dólares); 1970-2014 (dólares) ; 1972-2011 (dólares) y 1980-2014 (miles de millones de dólares). Todos estos son de corte temporal y de naturaleza cuantitativa.

- Gasto de consumo final total

Esta variable abarca el consumo de los hogares y del gobierno para el periodo 2002-2015 según datos del BCE y del periodo 1960-2014 de acuerdo a datos del WDI, ambas variables son de naturaleza cuantitativa y están expresadas en dólares americanos.

- Formación bruta de capital fijo

Se consideran dos fuentes de información distintas para esta variable, el BCE con las tablas de oferta y utilización en el periodo 2002-2015 que presentan valores en dólares americanos y las obtenidas en el WDI durante el periodo 1965-2014 que se encuentran en miles de dólares a precios constantes del 2005.

---

<sup>3</sup> A partir del año 2007 el Ecuador consta de 24 provincias, por tanto para esta variable se omitió a Santa Elena y Santo Domingo de los Tsachilas frente al periodo temporal analizado.

- Exportación de bienes y servicios

Esta variable es de naturaleza cuantitativa y de corte temporal, por lo que para la resolución de diversos ejercicios se utilizaron tres periodos de tiempo: 2002 al 2015 utilizando los datos del BCE expresados en dólares, series del WDI en los periodos 1960 al 2014 expresados en dólares a precios constantes del 2005 y 1970 2014 cuyos valores están medidos como porcentaje del PIB.

- Gasto de consumo final de hogares y Gasto de consumo final de los hogares per cápita

En base a las cuentas trimestrales del BCE e indicadores del WDI se utilizan tres versiones de estas variables cuantitativas y de corte temporal. La primera de ellas está expresada en millones de dólares durante el primer trimestre del 2000 hasta el cuarto trimestre del 2015, en la segunda se utilizan datos anuales del WDI en el periodo 1960-2014 expresado en dólares a precios constantes del 2005 y la tercera se encuentra esta misma variable dividida entre los habitantes del país y expresada a precios constantes del 2005.

- Empresas exportadoras, publicas y privadas

Utilizando el Censo Nacional Económico (CENEC) del año 2010 proporcionado por el INEC, se extrajeron datos para los 221 cantones del Ecuador que poseen empresas exportadoras, siendo estas públicas y privadas.

- Producción cantonal

Esta variable es de carácter cuantitativo y con corte transversal, es tomada del CENEC del año 2010 y se expresa en dólares americanos para cada uno de los cantones del Ecuador.

- Dicótoma de Polos Cantonales

Es una variable cualitativa generada a partir de los tres cantones más poblados en el Ecuador durante el año 2010, se asignó el valor de 1 para los cantones Quito, Guayaquil y Cuenca y el valor de 0 para el resto de cantones.

- Ingresos

A partir del Censo Nacional Económico del año 2010 se utiliza esta variable cantonal expresada en millones de dólares, es de corte transversal y de naturaleza cuantitativa.

- Número de empresas

Al igual que la variable anterior, en base al CENEC 2010 se obtiene el número de empresas que registra cada cantón en el año de estudio y se obtiene esta variable de tipo transversal y cuantitativa.

- Personal remunerado

Dentro de esta variable cuantitativa se encuentra el número de personas que perciben una remuneración durante el año 2010 en cada uno de los cantones que presenta el Censo Nacional Económico.

- Valor de activos fijos

Esta variable de corte transversal y cuantitativa tomada del CENEC, toma el valor de los activos registrados al 31 de diciembre del 2010 para cada cantón del Ecuador.

- Producción total

En base a la Encuesta de Comercio realizada por el INEC, se extrae los valores de producción total para el año 2013<sup>4</sup>, generando así una variable cuantitativa y de corte transversal expresada en dólares americanos. Además, esta variable también es utilizada para cuantificar la producción total de las microempresas hoteleras que invirtieron en publicidad en el Ecuador durante el año 2013, según la Encuesta de Hoteles y Servicios realizada por el INEC.

- Número de unidades de edificación

Esta variable de carácter cuantitativo y transversal fue tomada de la Encuesta de Edificaciones realizada por el INEC, cuantifica a las edificaciones en la ciudad de Loja durante el año 2013.

- Total de pisos

Esta variable cuantifica el número de pisos por edificación en la ciudad de Loja, es de corte transversal ya que toma los datos de la Encuesta de Edificaciones del año 2013.

- Valor total del terreno

En base a la Encuesta de Edificaciones del año 2013, se toma esta variable de corte transversal expresada en dólares americanos para medir el valor del terreno en la que se encuentran las edificaciones dentro de la ciudad de Loja.

---

<sup>4</sup> Con el fin de limitar la muestra, en algunas variables se utilizaron datos a nivel provincial o cantonal.

- Valor total de la edificación

Esta variable cuantifica el valor monetario total que tiene una edificación dentro de la ciudad de Loja, es de corte transversal ya que toma los datos de la Encuesta de Edificaciones del año 2013.

- Participación en las utilidades

Dentro de los datos de esta variable cuantitativa y de corte transversal expresada en dólares americanos, se considera de manera exclusiva a las microempresas hoteleras que invirtieron en publicidad en el Ecuador durante el año 2013, según la Encuesta de Hoteles y Servicios realizada por el INEC.

- Publicidad

Al igual que la variable descrita anteriormente, los datos de publicidad pertenecen a las microempresas hoteleras del Ecuador durante el año 2013 y se mantiene el carácter cuantitativo y transversal.

- Impuestos cantonales

Esta variable expresada en dólares americanos es obtenida de las estadísticas de recaudación del Instituto Nacional de Estadísticas y Censos del año 2010, su naturaleza es cuantitativa y su corte es transversal.

- Pib per cápita

Considerando la tabla del WDI de Ecuador se obtuvieron tres series de tiempo para los años 1960 al 2011; 1960-2014 y 1976-2013 a precios constantes del 2005 de estas variables, las cuales se expresaron en dólares americanos. Esta variable es el resultado del cociente entre el Producto Interno Bruto y todos los habitantes de las país dentro de una determinado periodo de tiempo.

- Población Total

Esta variable de carácter cuantitativo y temporal cuantifica el número de habitantes del país durante el periodo 1960-2011, de acuerdo a datos del WDI de Ecuador.

- Año

Esta variable cuantifica el número de años que se encuentran dentro de un determinado periodo de tiempo, dentro de la cual se utilizan datos de los periodos de 1960 al 2011 y de 1960 al 2013.

- Crecimiento del pib de Ecuador y Crecimiento del pib de Estados Unidos

Estas variables proporcionadas por la tabla de Indicadores de Desarrollo Mundial, representan la tasa de cambio que tiene el PIB del Ecuador y el de Estados Unidos durante un periodo determinado de tiempo. Por tanto, la naturaleza de estos indicadores es temporal y se la analiza durante el periodo 1966-2013.

- Crecimiento de la FBKF de Ecuador y Crecimiento de la FBKF de Estados Unidos.

Los valores de estas variables son proporcionadas por la tabla de WDI de Ecuador, representan la tasa de cambio que tiene la formación bruta de capital fijo del Ecuador y el de Estados Unidos durante un periodo determinado de tiempo. Por tanto, la naturaleza de estas variables es temporal y se la analiza durante el periodo 1966-2013.

- Comercio de mercaderías

Esta variable está expresada como porcentaje del PIB y su naturaleza es cuantitativa y de corte temporal, ya que es analizada durante el periodo 1970-2014.

- Petróleo en miles de barriles

Cuantifica el número de barriles producidos dentro del país durante el periodo 1972-2011 conforme a los datos presentados por el Banco mundial en la tabla del WDI de Ecuador, por tanto esta variable es cuantitativa y de dimensión temporal.

- Industrialización, valor agregado

Esta variable cuantitativa y de corte temporal, mide el crecimiento anual del valor agregado en la industria manufacturera en moneda local a precios constantes, además es utilizada durante el periodo 1976-2013 en base a los datos del WDI de Ecuador.

- Ahorro ajustado: gasto en educación

Esta variable tomada del WDI de Ecuador se refiere a los gastos operativos corrientes en educación expresados como porcentaje del Ingreso Nacional Bruto, por tanto es de naturaleza cuantitativa y de corte temporal para el periodo 1976-2013.

- Inversión extranjera directa

Esta variable fue obtenida de la tabla de indicadores de desarrollo del Banco mundial y consiste en el valor neto total de inversiones que entran al Ecuador durante el periodo 1976-2013, está expresada como porcentaje del PIB y por tanto esta variable es de dimensión temporal y naturaleza cualitativa.



- Índice de precios al consumidor

Esta variable refleja la variación porcentual anual en el costo para el consumidor medio de adquirir una canasta de bienes y servicios en un año para el caso ecuatoriano, los datos de naturaleza cuantitativa y de corte transversal fueron tomados del Banco Mundial para el periodo 1976-2013.

- Rentas del petróleo

Esta variable tomada del WDI de Ecuador cuantifica el monto económico que se obtiene por la diferencia entre valor de producción y sus costos para el periodo 1976-2013. Su naturaleza es cuantitativa y está expresada como porcentaje del PIB.

- Ingreso nacional neto ajustado

Esta variable obtenida del WDI de Ecuador es igual al ingreso nacional menos el consumo de capital físico y el agotamiento de los recursos naturales para el periodo 1988-2013. Su naturaleza es cuantitativa y de dimensión temporal y está expresada a precios constantes del 2005 en dólares americanos.

- Trabajadores asalariados

Esta variable está expresada como porcentaje del total de contratados en el Ecuador para el periodo 1988-2013, según datos del Banco Mundial, además posee un carácter temporal y cuantitativo.

- Cantidad de dinero

Se considera el periodo 1988-2013 para esta variable obtenida del WDI de Ecuador, es de naturaleza cuantitativa con de corte temporal y se encuentra expresada a precios corrientes de dólares americanos.

### **2.1.2 Descripción de variables sociodemográficas.**

Dentro de esta sección de datos se encuentran 41 variables de distintas fuentes de información como los son: el Banco Central del Ecuador, Dirección Nacional de Migración. Instituto Nacional de Estadísticas y Censos y el Banco Mundial a través de la tabla de Indicadores de Desarrollo Mundial. Dentro de estas se encuentran datos de corte transversal para los años 2010, 2013 y 2014; además de series de tiempo que van desde el año de 1980 hasta el 2014

**Tabla 2.2.** Variables sociodemográficas utilizadas

<b>Fuente</b>	<b>Año</b>	<b>Variable</b>	<b>Denominación</b>	<b>Naturaleza</b>
BCE – Sistema de información macroeconómica	1980 - 2013	Número de desempleados	DES	Cuantitativa
	1980 - 2014	Tasa de desempleo	DES	Cuantitativa
Dirección Nacional de Migración	1980-2013	Migración neta	MIG	Cuantitativa
INEC-Encuesta de Comercio	2013	Total mujeres empleadas	TME	Cuantitativa
		Total hombres empleados	THE	Cuantitativa
		total remuneraciones	TR	Cuantitativa
INEC- Encuesta hoteles y servicios	2013	Total empleados	TE	Cuantitativa
		Total de horas trabajadas en el mes de noviembre	HT	Cuantitativa
		Sueldos y salarios ejecutivos y gerenciales hombres	SEG	Cuantitativa
		Sueldos y salarios empleados mujeres	SEM	Cuantitativa
		Sueldos y salarios empleados hombres	SEH	Cuantitativa
INEC- ENEMDU	2010	Recuento de personas que estudian desde pre-escolar a postgrado	EST	Cuantitativa
		Dicótoma de Escolaridad	DE	Cualitativa
		Empleo cantonal	empleo	Cuantitativa
	2014	Salario mensual	w	Cuantitativa
			W	Cuantitativa
		Escolaridad	esc	Cuantitativa
			ESC	Cuantitativa
		Experiencia	exp	Cuantitativa
		Sexo	sexo	Cualitativa
			SEXO	Cualitativa
		Edad	edad	Cuantitativa
		Condición de empleo	cond_empleo	Cualitativa
		Motivo de desempleo	motivo_desempleo	Cualitativa
		Nivel de instrucción	NIVINST	Cualitativa
		Género	GEN	Cualitativa
		Salarios	SALARIO	Cuantitativa
		Instrucción primaria	PRIM	Cualitativa
		Instrucción secundaria	SEC	Cualitativa
		Instrucción superior	SUP	Cualitativa
		Instrucción post-grado	POST	Cualitativa
		Estado civil	ESTCIVIL	Cualitativa
Banco Mundial – WDI Ecuador	1960-2014	Población total	pob	Cuantitativa
		Población urbana	pob_urb	Cuantitativa
		Población rural	pob_rural	Cuantitativa
		Mortalidad infantil	MI	Cuantitativa
		Tasa de alfabetización de mujeres	TAM	Cuantitativa
	1960-2013	Esperanza de vida	esp_vida	Cuantitativa
		Tasa de mortalidad	tasa_mortalidad	Cuantitativa
	1984-2013	Tasa de mortalidad	mort	Cuantitativa
	1988-2013	Inscripción escolar, nivel secundario	IES	Cuantitativa

**Fuente:** Banco Central del Ecuador, Instituto Nacional de Estadísticas y Censos, Dirección de Migración.

**Elaboración:** Las autoras.

La descripción de cada variable económica utilizada dentro de los ejercicios econométricos es la siguiente:

- Número de desempleados

Esta variable cuantifica el número de personas que se encuentran desempleadas dentro del país, los datos se obtuvieron del sistema de información macroeconómica del Banco Central para el periodo 1980-2013. Su naturaleza es cuantitativa y su corte temporal, además el número de desempleados se expresó en miles de personas.

- Tasa de desempleo

Expresa el porcentaje de personas desempleadas con respecto a la población económicamente activa, su naturaleza es cuantitativa y de dimensión temporal ya que se analiza el periodo 1980-2014 con los datos proporcionados por el sistema de información macroeconómica del Banco Central del Ecuador.

- Migración neta

Esta variable cuantitativa y de carácter temporal es resultado de la diferencia entre emigrantes e inmigrantes del Ecuador para el periodo 1980-2013, esta está expresada en miles de personas para cada año de análisis.

- Total de hombres y mujeres empleados

En base a la Encuesta de Comercio 2013 realizada por el Instituto Nacional de Estadísticas y Censos, se extraen estas dos variables de corte cuantitativo y transversal cuya cuantificación es por número de personas dentro de estas categorías.

- Total remuneraciones

Esta variable cuantitativa y de corte transversal cuantifica el total de remuneraciones que perciben los empleados, se encuentra expresada en dólares americanos durante el año 2013 y la fuente de estos datos es la Encuesta de Comercio del mismo año.

- Total empleados

La variable está cuantificada por número de personas y es de carácter cuantitativo y transversal sesgada a las microempresas hoteleras que invirtieron en publicidad, los datos son obtenidos en el INEC a partir de la Encuesta hoteles y servicios realizada en el año 2013.

- Total de horas trabajadas en el mes de noviembre

A partir de la encuesta de hoteles y servicios, se extrae el número de horas trabajadas en las microempresas hoteleras que invirtieron en publicidad para el año 2013. Esta variable es por consiguiente de naturaleza cuantitativa y de dimensión espacial.

- Sueldos y salarios de ejecutivos y gerenciales hombres

Esta variable cuantifica los sueldos y salarios de los ejecutivos y gerentes que trabajan en las microempresas hoteleras que invirtieron en publicidad para el año 2013 en el Ecuador, esta expresada en dólares americanos y es de carácter cuantitativo y espacial.

- Sueldos y salarios empleados mujeres y hombres

Estas dos variables cuantifican los sueldos y salarios de los empleados hombres y mujeres que trabajan en las microempresas hoteleras que invirtieron en publicidad para el año 2013 en el Ecuador, esta expresada en dólares americanos y es de carácter cuantitativo y espacial.

- Recuento de personas que estudian desde pre-escolar a postgrado

Esta variable es tomada de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) realizada por el INEC en el 2010 para todos los cantones del Ecuador, esta expresada en número de personas y su naturaleza es cuantitativa y su corte es transversal.

- Dicótoma de Escolaridad

Esta variable es obtenida del Instituto Nacional de Estadísticas y Censos a través de la ENEMDU 2010, sus valores son de carácter cualitativo y espaciales, donde el valor de 1 corresponde a las personas con más de 7 años de escolaridad y 0 para las personas con menos de 7 años de escolaridad.

- Empleo cantonal

Esta variable está expresada en número de personas empleadas por cantón, fue obtenida a partir de la ENEMDU 2010 y su naturaleza es cuantitativa y de corte transversal.

- Salario mensual

A partir de la ENEMDU realizada en el año 2014 se obtienen el salario mensual de los habitantes de la ciudad de Quito expresados en dólares americanos, por tanto esta variable es de corte transversal y de naturaleza cuantitativa.

- Escolaridad

Continuando con los datos de la ENEMDU 2014, se presenta esta variable como el número de años de educación completada de una muestra aleatoria de datos, los cuales son de corte temporal y de naturaleza cuantitativa.

- Experiencia

Con esta variable se expresa los años laborados de una muestra al azar de la Encuesta Nacional de Empleo, Desempleo y Subempleo en el año 2014, esta es por tanto de naturaleza cuantitativa y de corte espacial.

- Sexo

Esta variable de naturaleza cualitativa expresa con valores de 1 a las observaciones de hombres y con el valor de 2 a las observaciones de las mujeres, fue obtenida de la ENEMDU 2014 y es de corte transversal aplicada a la ciudad de Quito.

- Edad

En base a una muestra aleatoria de datos provenientes de la Encuesta Nacional de Empleo, Desempleo y Subempleo 2014, se cuantifica la variable en número de años para cada observación, siendo así de carácter cuantitativo y espacial.

- Condición de empleo

Esta variable cualitativa obtenida a partir de la ENEMDU 2014 es de carácter transversal y enuncia la condición de empleo bajo las categorías de empleado y desempleado, además la muestra utilizada para este conjunto de datos fue de manera aleatoria.

- Motivo de desempleo

Esta variable cualitativa obtenida a partir de la ENEMDU 2014 en una muestra aleatoria de datos, representa el motivo de desempleo en tres categorías: falta de oportunidades en la actualidad, motivos personales o familiares y finalmente el resultado de sus errores del pasado.

- Nivel de instrucción

A partir de la ENEMDU 2014 se genera esta variable de carácter cualitativo y de corte transversal aplicada a la ciudad de Quito en los habitantes empleados en el rango de edad de 15 a 65 años, las categorías que presentan son: 1 para nivel de primaria, 2 para el nivel secundario, 3 para secundaria y 4 para el post-grado.

- Género

Esta variable dicotómica es obtenida de la ENEMDU 2014 de la cual se tomaron únicamente los datos de la parroquia Cumbe de la ciudad de Cuenca con salarios menores a \$600.00 dólares, las categorías que esta presenta es de 1 para las observaciones de hombres y el valor de 0 para las observaciones de las mujeres.

- Salarios

Esta variable cuantitativa y de corte transversal es obtenida de la ENEMDU 2014 de la cual se tomaron únicamente los datos de la parroquia Cumbe de la ciudad de Cuenca con salarios menores a \$600.00 dólares, esta se encuentra expresada en cientos de dólares americanos para el periodo 2014.

- Instrucción primaria, secundaria, superior y de post-grado

A partir de la ENEMDU 2014 se generan estas cuatro variables de carácter cualitativo y de corte transversal, que son aplicadas a la ciudad de Quito en los habitantes empleados en el rango de edad de 15 a 65 años. Los criterios para la variable de Instrucción Secundaria es: 1 para los que estudian en secundaria y 0 para todos los demás; Instrucción Superior con valor de 1 para los que estudien en nivel superior y 0 para todos los demás; Instrucción de Post-Grado con categoría de 1 para los que se encuentren en un post-grado y cero para todos los demás.

- Estado civil

Esta variable de naturaleza cualitativa expresa con valores de 1 en las observaciones de personas casadas y 0 para el resto de observaciones de la ciudad de Quito, fue obtenida de la ENEMDU 2014 y es de corte transversal.

- Población total

Esta variable es de serie de tiempo para el período 1960-2014, extraída del Banco Mundial de la fuente de datos del WDI para Ecuador. Está medida en número de personas, por lo tanto es de naturaleza cuantitativa.

- Población urbana

Es una variable de serie de tiempo para el período 1960-2014. Sus datos han sido tomados del Banco Mundial específicamente del WDI para Ecuador. Expresa el porcentaje de la población urbana con respecto a la población total, siendo así una variable cuantitativa.

- Población rural

Al igual que la población total y la población urbana, esta variable es de serie de tiempo para el período 1960-2014, extraída del WDI para Ecuador del Banco Mundial. Expresa el porcentaje de la población rural con respecto a la población total, por lo tanto es una variable cuantitativa.

#### Mortalidad infantil

Esta variable es de serie temporal para el período 1960-2014, medida por el número de muertes por cada 1000 niños menores a 5 años. Se la ha tomado de la base de datos del WDI para Ecuador del Banco Mundial, y es de naturaleza cuantitativa.

- Tasa de alfabetización de mujeres

Es una variable de serie de tiempo para el período 1960-2014, encontrada en el Banco Mundial en la base del WDI para Ecuador. Esta variable mide el porcentaje de mujeres alfabetizadas de 15 o más años, por lo cual es de naturaleza cuantitativa.

- Esperanza de vida

La variable esperanza de vida al nacer consta de datos de serie temporal del período 1960 - 2013, tomados del Banco Mundial específicamente de WDI para Ecuador. Esta expresada en número de años y es de naturaleza cuantitativa.

- Tasa de mortalidad

A esta variable se la puede encontrar en el WDI para Ecuador en las bases del Banco Mundial como tasa de mortalidad en un año, expresada en el número de muertes por cada 1000 personas, siendo de esta manera una variable cuantitativa. Además se ha utilizado a esta variable en dos períodos diferentes: 1960-2013; y 1984-2013.

- Inscripción escolar, nivel secundario

Es una variable de serie temporal para el período 1988-2013, extraída del Banco Mundial de la base de datos del WDI para Ecuador. Esta variable corresponde al número total de estudiantes matriculados en educación secundaria, independientemente de su edad, expresado como porcentaje de la población total en edad oficial de cursar la secundaria, por lo tanto es de naturaleza cuantitativa.

### 2.1.3 Descripción de variables ambientales.

Para el desarrollo de ciertos ejercicios se han utilizado 16 variables ambientales, las mismas que fueron tomadas de la base de datos del Ministerio de Agricultura, Ganadería, Acuacultura y Pesca (MAGAP) en períodos de tiempo desde el año 2000 hasta el 2012 y del Banco Mundial en períodos de tiempo desde el año 1960 hasta el 2014.

**Tabla 2.3.** Variables ambientales utilizadas.

Fuente	Año	Variable	Denominación	Naturaleza
MAGAP -Costo de Producción productos	2000-2012	Superficie sembrada arroz	SS	Cuantitativa
		Superficie cosechada arroz	SC	Cuantitativa
		Superficie sembrada banano	SS	Cuantitativa
		Producción de arroz	PA	Cuantitativa
		Producción de banano	Pbanano	Cuantitativa
			PAbanano	Cuantitativa
			PTbanano	Cuantitativa
		Rendimiento del banano	R	Cuantitativa
		Rendimiento del arroz	R	Cuantitativa
		Rendimiento del cacao	cacao	Cuantitativa
		Rendimiento del café	cafe	Cuantitativa
		Rendimiento del maracuyá	maracuya	Cuantitativa
		Rendimiento de la papa	papa	Cuantitativa
		D1= 1 si la producción de arroz es de la provincia de Napo y 0 si es otra.	D1	Cualitativa
		D2= 1 si la producción de arroz es de la provincia de Orellana y 0 si es otra.	D2	Cualitativa
		D3= 1 si la producción de arroz es de la provincia de Sucumbios y 0 si es otra.	D3	Cualitativa
		D4= 1 si la producción de arroz es de la provincia de Zamora y 0 si es otra.	D4	Cualitativa
	2000, 2005, 2010	Rendimiento del arroz	arroz	Cuantitativa
		Rendimiento del banano	banano	Cuantitativa
		Rendimiento del cacao	cacao	Cuantitativa
		Rendimiento del café	cafe	Cuantitativa
		Rendimiento del maracuyá	maracuya	Cuantitativa
		Rendimiento de la papa	papa	Cuantitativa
Banco Mundial-WDI Ecuador	1960-2011	Emisiones de CO <sub>2</sub>	CO <sub>2</sub>	Cuantitativa
	1970-2014	Emisiones de CO <sub>2</sub>	CO <sub>2</sub>	Cuantitativa

**Fuente:** Ministerio de Agricultura, Ganadería, Acuacultura y Pesca y Banco Mundial.

**Elaboración:** Las autoras.



- Superficie sembrada arroz

Es una variable de serie de tiempo para el período 2000-2012. Sus datos son segregados a nivel provincial tomados de la base de Costo de Producción encontrada en la página del MAGAP. La superficie sembrada del arroz está medida en hectáreas, por lo cual es de naturaleza cuantitativa.

- Superficie cosecha arroz

Es una variable de serie de tiempo para el período 2000-2012. Sus datos son segregados a nivel provincial tomados de la base de Costo de Producción encontrada en la página del MAGAP. La superficie de cosecha del arroz está medida en hectáreas, siendo de naturaleza cuantitativa.

- Superficie sembrada banano

Es una variable de serie de tiempo para el período 2000-2012. Sus datos son segregados a nivel provincial tomados de la base de Costo de Producción encontrada en la página del MAGAP. La superficie sembrada del banano está medida en hectáreas, siendo así una variable cuantitativa.

- Producción de arroz

La producción de arroz en cáscara, seco y limpio es un variable temporal a nivel provincial, que contiene datos del período 2000-2012 tomados de la base del Costo de Producción de la página del MAGAP. Esta variable está medida en toneladas métricas, por lo tanto es de naturaleza cuantitativa.

- Producción de banano

La producción en fruta fresca del banano es una variable temporal a nivel provincial, la cual fue extraída de la base de datos del Costo de Producción del MAGAP para el período 2000-2012. Está expresada en toneladas métricas, y es de naturaleza cuantitativa. Es importante mencionar que a esta variable se la ha utilizado varias veces en distintos modelos econométricos, por lo cual tiene 3 denominaciones distintas como se muestra en la Tabla 2.3.

- Rendimiento del banano, arroz, cacao, café, maracuyá y papa

Son variables cuantitativas, temporales a nivel provincial. Han sido tomadas de la base de datos Costo de Producción del MAGAP. Están medidas en toneladas por hectárea (Tm/Ha).

Estas variables son utilizadas en dos ocasiones cada una, la primera para el período 2000-2012 y la segunda para los años 2000, 2015, y 2010. Los datos de estos 3 años son utilizados simplemente para la realización de gráficas comparativas.

- D1

Es una variable cualitativa (dicótoma) de serie temporal a nivel provincial, que toma el valor de 1 si la producción de arroz es de la provincia de Napo y 0 en todos los demás casos. Estos valores se basan en la fuente de datos del Costo de Producción del MAGAP para el período 2000-2012.

- D2

Es una variable cualitativa (dicótoma) de serie de tiempo a nivel provincial, que toma el valor de 1 si la producción de arroz es de la provincia de Orellana y 0 en todos los demás casos. Estos valores se basan en la fuente de datos del Costo de Producción del MAGAP para el período 2000-2012.

- D3

Es una variable cualitativa (dicótoma) de serie temporal a nivel provincial, que toma el valor de 1 si la producción de arroz es de la provincia de Sucumbios y 0 en todos los demás casos. Estos valores se basan en la fuente de datos del Costo de Producción del MAGAP para el período 2000-2012.

- D4

Al igual que D1, D2 y D3 es una variable cualitativa (dicótoma) de serie de tiempo para el período 2000-2012, segregada a nivel provincial. Basada en la fuente de datos del Costo de Producción disponible en la página del MAGAP. Pero esta variable toma el valor de 1 si la producción de arroz es de la provincia de Zamora y 0 en todos los demás casos.

- Emisiones de CO<sub>2</sub>

Es una variable de serie temporal, de naturaleza cuantitativa, que ha sido tomada de la base de datos del Banco Mundial específicamente del WDI para Ecuador. Está medida en toneladas métricas per cápita. Y además se ha utilizado a esta variable en diferentes períodos: 1960-2011 y 1970-2014.

### 2.1.4 Descripción de variables hipotéticas.

Con el fin de abarcar algunas cuestiones teóricas de la econometría que no se han podido establecer con datos reales del Ecuador se han utilizado variables con datos hipotéticos. Son únicamente 6 variables hipotéticas utilizadas en ciertos ejercicios, las cuales se muestran en la Tabla 2.4.

**Tabla 2.4.** Variables hipotéticas utilizadas.

Fuente	Año	Variable	Denominación	Naturaleza
Variables hipotéticas	-	Consumo familiar	consumof	Cuantitativa
		Ingreso familiar	ingresof	Cuantitativa
		Costo total	CT	Cuantitativa
		Producción	PROD	Cuantitativa
		Comisión por ventas	COMISION	Cuantitativa
		Ventas	VENTAS	Cuantitativa

**Elaboración:** Las autoras.

- Consumo familiar e ingreso familiar

Al igual que Gujarati y Porter (2010) se han utilizado estas dos variables como hipotéticas para resolver ejercicios que expliquen de manera sencilla algunas ideas básicas del análisis de regresión. Ambas variables están medidas en dólares americanos.

- Costo total

Esta variable está medida en dólares, y es utilizada para ajustar el comportamiento cúbico del costo total frente a los niveles de producción. Los modelos de costos (grado 3) no siempre se ajustan al comportamiento de sus costos de empresas del Ecuador.

- Producción

Esta variable está expresada en cantidades producidas. Se la usa como variable explicativa en el modelo de costos totales como se mencionó anteriormente.

- Comisión por ventas y ventas

Estas variables están medidas en dólares. Se las ha utilizado con el fin de estimar una regresión lineal por segmentos bien ajustada, que muestre claramente que después de alcanzar un determinado nivel de ventas (umbral) en el segundo segmento la comisión por ventas crece a una tasa mayor que en el primer segmento lineal.

### 2.1.5 Descripción de variables de encuestas.

No toda la información tomada para la realización de los ejercicios ha sido extraída de fuentes oficiales de información, ya que también se hizo uso de información producto de encuestas realizadas a estudiantes de la “Universidad Técnica Particular de Loja” (UTPL), y a personas residentes en el cantón El Pangui, Provincia de Zamora Chinchipe. En consecuencia se han utilizado un total de 9 variables de estas fuentes.

**Tabla 2.5.** Variables de encuestas varias.

Fuente	Año	Variable	Denominación	Naturaleza
Encuesta aplicada a estudiantes de UTPL	Abril-Agosto 2013	Edad del estudiante	edad	Cuantitativa
		Nota del examen de admisión.	NEA	Cuantitativa
		Número de componentes matriculados	NCM	Cuantitativa
		Número de créditos que aprobados	NCA	Cuantitativa
		Promedio obtenido	Prom	Cuantitativa
		Ciclo actual	ciclo	Cuantitativa
Encuesta de seguridad alimentaria realizada en el cantón El Pangui	2015	Edad	EDAD	Cuantitativa
		Personas incluidas dentro de la edad	N	Cuantitativa
		Personas enfermas el último año	ENF	Cuantitativa

**Fuente:** Encuesta a estudiantes (2013), Encuesta de seguridad alimentaria (2015).

**Elaboración:** Las autoras.

- Edad del estudiante

Es una variable de corte transversal para el período académico abril-agosto 2013. Está expresada en el número de años de edad de los estudiantes, por tanto es de naturaleza cuantitativa. Esta variable ha sido tomada de una encuesta realizada a estudiantes de la UTPL, modalidad presencial.

- Nota del examen de admisión

Esta variable muestra información de la nota sobre 100 puntos obtenida por cada estudiante en el examen de admisión a la UTPL. Es una variable cuantitativa de corte transversal para el período académico abril-agosto 2013, que se la ha obtenido de una encuesta realizada a estudiantes de la UTPL, modalidad presencial.

- Número de componentes matriculados

Es una variable cuantitativa y de corte transversal para el período académico abril-agosto 2012. Refleja el número de componentes (materias) en los cuales cada estudiante se ha

matriculado en el período académico antes mencionado. Estos datos son extraídos de una encuesta realizada a estudiantes de la UTPL, modalidad presencial.

- Número de créditos aprobados

Esta variable está expresada por el número de créditos aprobados por cada estudiante en el período académico abril – agosto 2013, por lo tanto es de naturaleza cuantitativa. Además es una variable de corte transversal al extraer datos de un solo periodo académico. Esta variable fue tomada de una encuesta realizada a estudiantes de la UTPL, modalidad presencial.

- Promedio obtenido

Expresa la nota promedio obtenida sobre 40 puntos por cada estudiante en el período académico abril-agosto 2013. Esta variable es cuantitativa y de corte transversal, con datos obtenidos de una encuesta realizada a estudiantes de la UTPL, modalidad presencial.

- Ciclo actual

Indica el nivel de ciclo que está cursando el estudiante, estos valores pueden ser del 1 al 10, siendo 1 el primer ciclo y 10 el décimo ciclo. Esta variable es de naturaleza cuantitativa y de corte transversal, extraída de una encuesta realizada a estudiantes de la UTPL, modalidad presencial.

- Edad

Es una variable cuantitativa y de corte transversal para el año 2015, cuya información se la tomó de una encuesta de seguridad alimentaria realizada a los habitantes del cantón El Pangui. Esta variable está medida en el número de años de edad de cada persona.

- Personas incluidas dentro de la edad

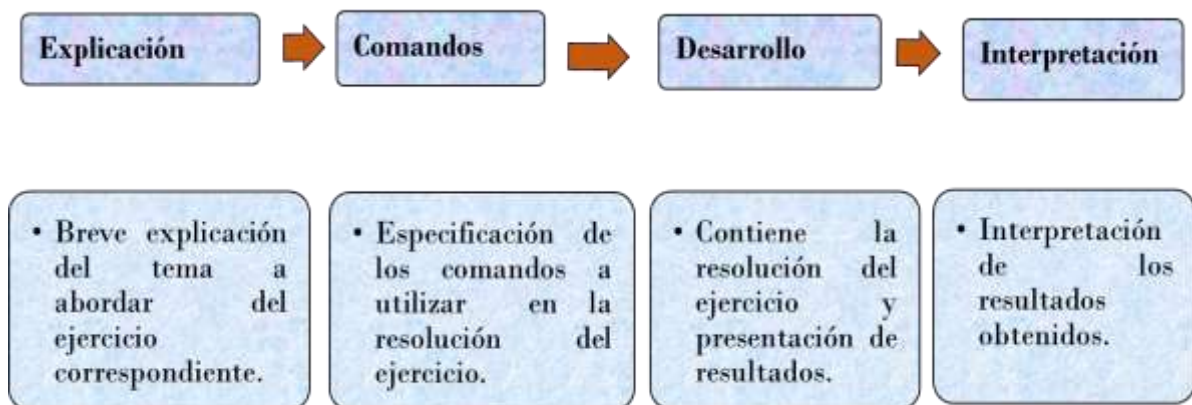
Es una variable cuantitativa expresada por el número de personas según la edad. Esta variable es de corte transversal para el año 2015. Y ha sido tomada de una encuesta de seguridad alimentaria realizada a los habitantes del cantón El Pangui.

- Personas enfermas el último año

Esta variable muestra el número de personas enfermas según la edad, por tal razón esta variable es de naturaleza cuantitativa. Además al contener datos solamente del 2015 es una variable de corte transversal, cuya información se ha obtenido de una encuesta de seguridad alimentaria realizada a los habitantes del cantón El Pangui.

## 2.2 Metodología

Para resolver los 108 ejercicios econométricos se ha utilizado la metodología 3 propuesta en el estudio exploratorio a través de una encuesta realizada a los estudiantes de econometría (Anexo 1). Esta metodología referente a la resolución de un ejercicio estándar ha sido la más entendible frente a las otras dos opciones propuestas como se muestra en el gráfico 1.6. En este sentido, la estructura de un ejercicio estándar consta de 4 partes: explicación, comandos a utilizar, desarrollo e interpretación.



**Figura 2.1.** Estructura de un ejercicio estándar

**Elaboración:** Las autoras

**Explicación.** Se empieza citando brevemente el tema a abordar del ejercicio, con la finalidad de que el lector tenga un claro entendimiento de lo que se va a desarrollar. En otras palabras, se brinda una explicación básica de los aspectos teóricos de la econometría a abordar en el ejercicio.

**Comandos a utilizar.** Se especifican los comandos del programa estadístico Stata 13 a utilizar en la resolución del ejercicio. Al mismo tiempo, se explica el uso de cada comando, se incluye la sintaxis o estructura de cada comando, y se agrega la rutina de comandos utilizada. Cabe mencionar que si los comandos ya fueron utilizados en ejercicios anteriores, estos serán omitidos ya que no será necesario explicarlos nuevamente, con el fin de evitar explicaciones repetitivas.

**Desarrollo.** Es la parte central del ejercicio, ya que contiene la resolución y presentación de resultados, ya sea mediante tablas, gráficas, ecuaciones, pruebas de hipótesis, etc. Aquí se aplican los comandos anteriormente mencionados explicando detalladamente el procedimiento para llegar el resultado deseado.

**Interpretación.** Se interpretan los resultados obtenidos en la etapa de desarrollo del ejercicio. No todos los ejercicios llevan interpretación ya que en algunos no es necesario, como en el caso de los ejercicios del apartado de aspectos básicos donde solamente se explican cuestiones básicas del programa Stata lo cual no se puede interpretar.

### **CAPÍTULO III**

#### **APLICACIÓN ECONOMÉTRICA A CASOS ECUATORIANOS**



El presente capítulo comprende múltiples aplicaciones de la econometría usando datos del Ecuador y algunos datos hipotéticos, los segundos son presentados con el único fin de sustentar aplicaciones teóricas. Para introducir al lector al ambiente de trabajo en Stata, se empieza explicando algunos aspectos básicos del programa para luego continuar resolviendo ejercicios que abarcan los primeros 16 capítulos del libro de Econometría de Gujarati & Porter (2010), desde el capítulo 1. Naturaleza del análisis de regresión hasta el capítulo 16. Modelos de regresión con datos de panel. Cabe indicar que se han omitido algunos temas del libro que son netamente teóricos. La equivalencia de los temas de este capítulo con los temas del libro de Econometría de Gujarati & Porter (2010) se puede ver en el Anexo 2.

### 3.1 Aspectos básicos

Para conocer el manejo básico de Stata se ha considerado necesario explicar algunos procedimientos que implican abrir, cerrar y guardar archivos dta y do, y sobretodo cuestiones relacionadas con gestión de datos. Para lo cual, se han resuelto nueve ejercicios que utilizan datos de variables económicas de empleo y salarios contenidos en las Tablas 3.1.1, Tabla 3.1.2 y Tabla 3.1.3 (Anexo 5).

#### 3.1.1 Abrir y cerrar archivos: dta y do.

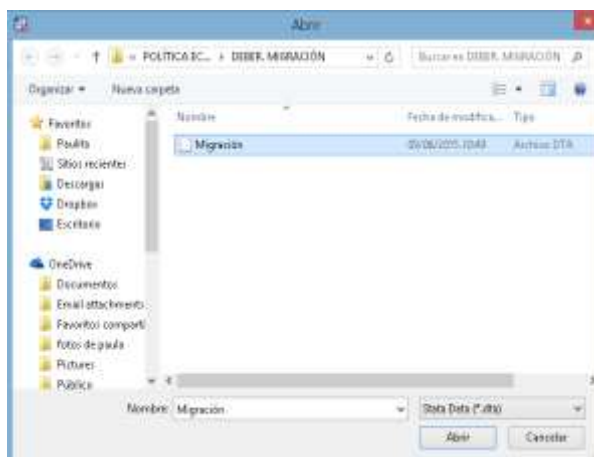
##### Ejercicio 1:

En los archivos formato *.dta* se almacena la base de datos del trabajo realizado, mientras que en los archivos formato *.do* se almacenan las rutinas de comandos utilizados en una sesión de trabajo en Stata.

Desarrollo:

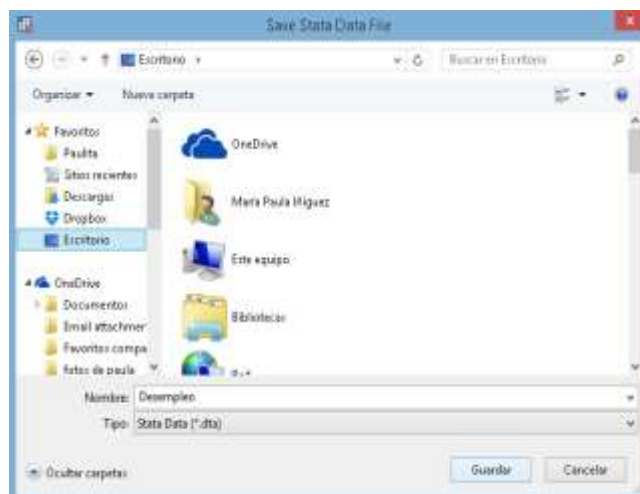
- **Abrir y guardar archivo dta.**

Para abrir un archivo dta, se da clic en *File/ Open*. Luego se abre la siguiente ventanilla:




Se busca el archivo dta que se desee abrir, se lo selecciona, y se da clic en *abrir*.

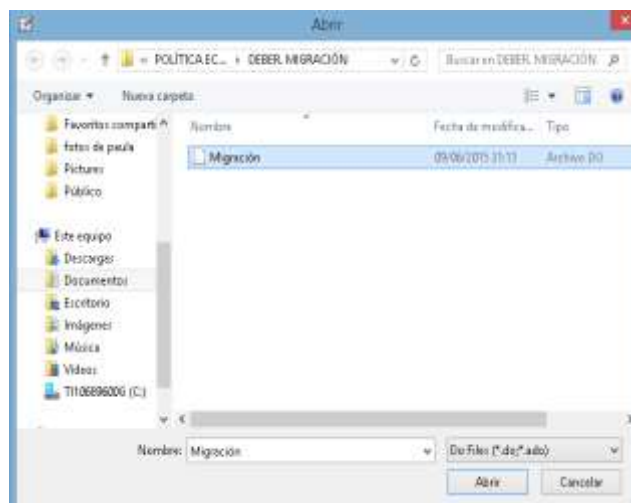
Para guardar un archivo dta, se da clic en *File/ Save as*. Luego se abre la siguiente ventanilla:



El usuario se ubicará en donde desee guardar su archivo dta (en este caso se procederá a guardar en el escritorio). Luego se escribe el nombre del archivo, y por último se da clic en *guardar*.

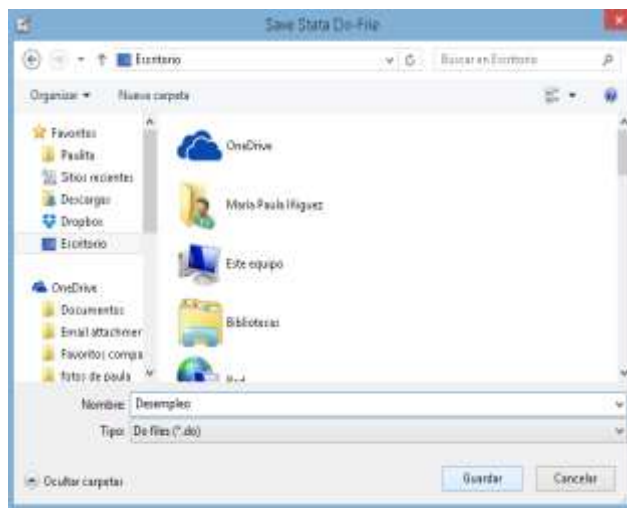
- **Abrir y guardar archivo do.**

Para abrir un archivo do, se da clic en el ícono,  *New Do-File Editor*, y se abre la ventanilla del archivo do. Luego se da clic en *File / Open*, y se abre la siguiente ventanilla:



Se busca el archivo do que se desea abrir, se lo selecciona, y se da clic en *abrir*.

Para guardar un archivo do, en la ventanilla del archivo do a guardar, se da clic en *File/ Save as*, y se abre la siguiente ventanilla:



El usuario se ubicará en donde desee guardar su archivo do (en este caso se guardará en el escritorio). Luego se escribe el nombre del archivo, y por último se da clic en *guardar*.

### 3.1.2 Ayuda e instalación de comandos.

#### **Ejercicio 2:**

Existen comandos y rutinas que no se encuentran en el programa, pero se los puede añadir de acuerdo a las necesidades del usuario.

Comandos a utilizar:

**help:** Permite obtener ayuda de Stata sobre un comando en particular. En el caso de que este comando no esté en la memoria del programa permite añadir comandos nuevos.

Su estructura es:  
**help** comando

Rutina utilizada:  
**help** whitetst

Desarrollo<sup>5</sup>:

El comando *whitetst*, que sirve para realizar el test de heteroscedasticidad de White (se verá más adelante en el apartado 3.9) no se encuentra instalado en Stata 13. Para añadirlo primeramente se ejecuta el siguiente comando: **help** whitetst. Luego aparecerá la siguiente ventana:

<sup>5</sup> Se puede utilizar este mismo procedimiento con otros comandos que nos estén instalados.

```

Example . . . . . Stata web books: Regression with Stata
. . . . . Chen, Ender, Mitchell & Wells (UCLA Academic Technology Services)
6/07 web book Regression with Stata by (in alphabetical
order) Xiao Chen, Philip B. Ender, Michael Mitchell
& Christine Wells
http://www.ats.ucla.edu/stat/stata/webbooks/reg/

Example . . . . . Textbook Examples: Econometric Analysis, Fourth Edition
. . . . . UCLA Academic Technology Services
10/05 examples from the book Econometric Analysis, Fourth
Edition by William Greene
http://www.ats.ucla.edu/stat/examples/greene/

Web resources from Stata and other users

(contacting http://www.stata.com)

2 packages found (Stata Journal and STB listed first)

sgl37 from http://www.stata.com/stb/stb55
STB-55 sgl37. Test for heteroskedasticity in reg. error distribution /
STB insert by Christopher F. Baum, Boston College / Nicholas J. Cox,
University of Durham / Vincent Wiggins, Stata Corporation / Support:
baum@bc.edu / n.j.cox@durham.ac.uk / vwiggins@stata.com / After

whitetst from http://www.bc.edu/RePEc/bocode/w
'WHITETST': module to perform White's test for heteroskedasticity /
whitetst computes White's test for heteroskedasticity following / regress
or cnsreg. This test is a special case of the / Breusch-Pagan test (q.v.
bpagan). The White test does not require / specification of a list of

```

Se da clic en el link [sgl37 from http://www.stata.com/stb/stb55](http://www.stata.com/stb/stb55) y aparecerá una ventana para instalar el comando.

```

package sgl37 from http://www.stata.com/stb/stb55

TITLE
STB-55 sgl37. Test for heteroskedasticity in reg. error distribution

DESCRIPTION/AUTHOR(S)
STB insert by Christopher F. Baum, Boston College
Nicholas J. Cox, University of Durham
Vincent Wiggins, Stata Corporation
Support: baum@bc.edu
n.j.cox@durham.ac.uk
vwiggins@stata.com
After installation, see help bpagan, help whitetst

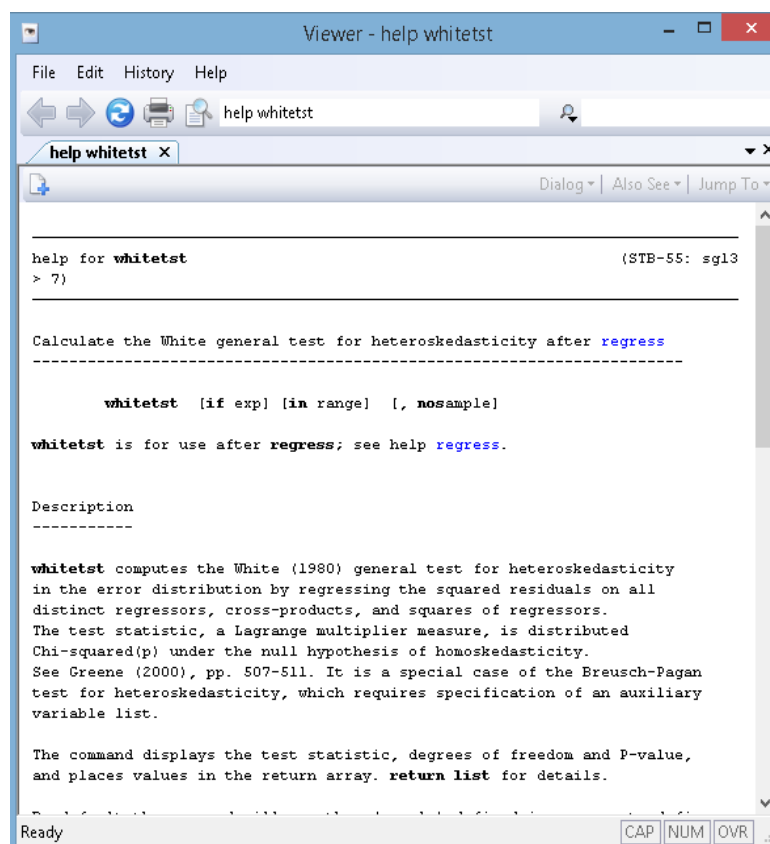
INSTALLATION FILES (click here to install)
sgl37/bpagan.ado
sgl37/bpagan.hlp
sgl37/whitetst.ado
sgl37/whitetst.hlp

(click here to return to the previous screen)
net install sgl37.pkg

```

Se da clic en [click here to install](#), y ya se instalará el nuevo comando o rutina.

Si se vuelve a ejecutar el comando: **help** whitetst, ahora aparecerá esta ventana:



Esta ventana muestra información acerca del comando **whitetst**. Por lo tanto, si se quiere conocer acerca de determinado comando que ya se encuentra instalado en Stata, el comando **help** brindará la información del mismo.

### 3.1.3 Manejo de datos.

Para poder trabajar con Stata es necesario conocer cómo se van a utilizar los datos, empezando por la introducción de éstos hasta la transformación de los mismos. Frecuentemente los datos necesitan ser modificados, eliminados, renombrados, etiquetados, creados, etc., de acuerdo a las necesidades del usuario.

#### 3.1.3.1 Introducción de datos.

#### Ejercicio 3:

Una vez iniciada la sesión en Stata, es decir una vez abierto el programa se procede a introducir los datos para empezar el trabajo. Existen algunas formas de introducir los datos al programa, tales como:

- Utilizando el comando **imput**.
- Por medio de la opción *Import*.

Comandos a utilizar:

**input:** Permite introducir los datos directamente en la memoria de datos.

Su estructura es:

**input** var1 var2... var\_n

.....**end**

Rutina utilizada:

**input** w esc exp


.....**end**

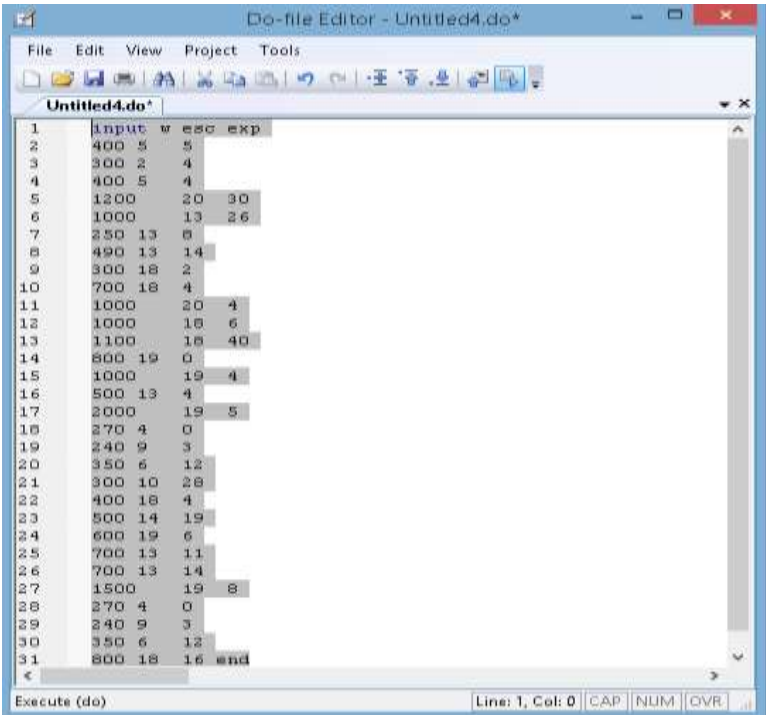
Desarrollo:

Se van a introducir los datos de las variables: salario, escolaridad y experiencia (Tabla 3.1.1) al programa Stata.


- **Utilizando el comando *input*.**

Se hace uso de *Do-file-editor*, ventana en la cual se puede correr todos los comandos para el trabajo. La ventaja de utilizar esta ventana es que se puede guardar toda una rutina de comandos como un archivo *do*, el cual puede ser utilizado nuevamente cuando se lo requiera.

Se da clic en ícono  ubicado en la barra de herramientas, y se abre la ventana *Do-file-editor*. En esta ventana escribe **input**, se pegan o se escriben las variables con sus distintos valores, y al final se escribe la palabra **end** para indicar que se ha terminado de introducir los datos. Esto se lo hace de la siguiente manera:




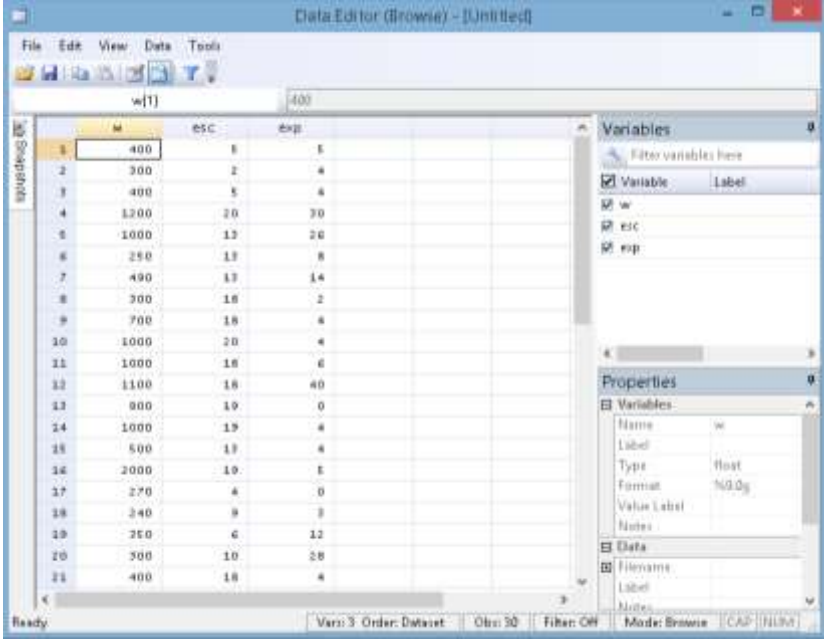
```
1 input w esc exp
2 400 5 5
3 300 2 4
4 400 5 4
5 1200 20 30
6 1000 13 26
7 250 13 0
8 490 13 14
9 300 18 2
10 700 18 4
11 1000 20 4
12 1000 18 6
13 1100 18 40
14 800 19 0
15 1000 19 4
16 500 13 4
17 2000 19 5
18 270 4 0
19 240 9 3
20 350 6 12
21 300 10 28
22 400 18 4
23 500 14 19
24 600 19 6
25 700 13 11
26 700 13 14
27 1500 19 8
28 270 4 0
29 240 9 3
30 350 6 12
31 800 18 16 end
```

Se seleccionan todos los datos introducidos, y se da clic en , ícono que ejecuta los comandos seleccionados.

Para verificar que los datos se hayan ingresado correctamente en el programa, se observa en la parte derecha de la ventana principal, en *Variables*. Así:

Variables		Y	q	X
Variable	Label			
w				
esc				
exp				

Efectivamente ya se encuentran las variables ingresadas, sin embargo no se puede apreciar los datos de las mismas, por lo tanto se recurre a la opción *Data Editor (Browse)*, haciendo clic en , y se abre la siguiente ventana:



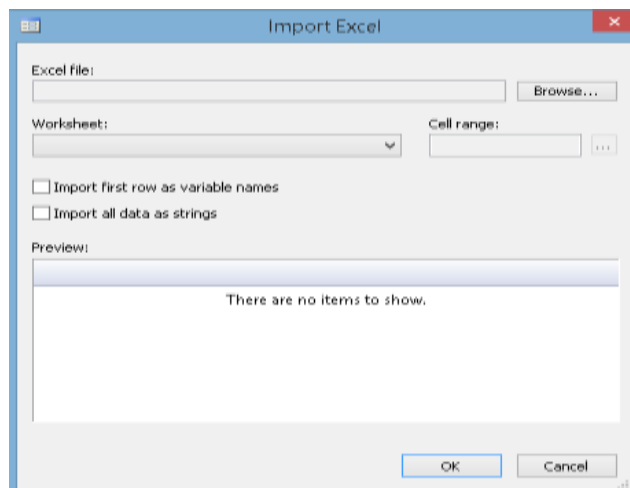
	w	esc	exp
1	400	8	5
2	300	2	4
3	400	5	4
4	1200	20	30
5	1000	12	20
6	250	12	8
7	490	12	14
8	300	18	2
9	700	18	4
10	1000	20	4
11	1000	18	6
12	1100	18	40
13	800	10	0
14	1000	19	4
15	500	12	4
16	2000	10	5
17	270	4	0
18	240	9	2
19	250	6	12
20	300	10	28
21	400	18	4

De esta manera se corrobora que los datos se han ingresado correctamente al programa.

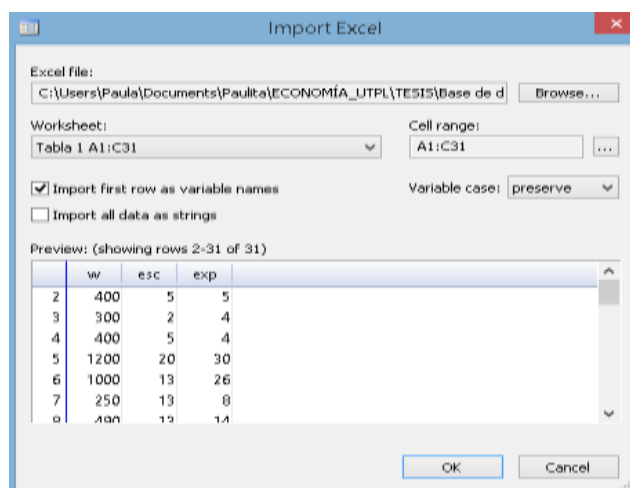
- **Por medio de la opción *Import*<sup>6</sup>.**

Como ya se mencionó anteriormente también se pueden introducir datos haciendo uso de la opción *import*. Para ejemplificarlo se van a utilizar datos que se encuentran en formato xls (archivo de excel), y se sigue el siguiente procedimiento: *FILE/ IMPORT/ EXCEL SPREADSHEET*. Y a continuación se abre la siguiente ventana:

<sup>6</sup> Para las versiones anteriores a Stata 13 es necesario que el archivo de excel a introducir haya sido previamente guardado en modo compatibilidad (formato .xls).



En esta ventana se hace clic en *Browse* y se selecciona el documento xls para importarlo.



En *Worksheet* se muestra el nombre de la hoja de cálculo del documento xls y a su derecha se halla una pestaña que permite seleccionar la hoja de cálculo que contenga los datos que se quieren introducir. Además, se da clic en *import first row as variable names* para que la primera fila del archivo sea considerada como el nombre de las variables. Por último, se da clic en *OK* y los datos ya estarán introducidos en el programa.

### 3.1.3.2 Datos numéricos almacenados como cadena.

#### Ejercicio 4:

Después de importar los datos en Stata, se podría presentar el caso de que los datos numéricos están en cadena. Esto se lo puede apreciar al momento de visualizar las variables introducidas, pues se observa que algunas de estas variables están de color rojo más no de negro como las variables numéricas.



Comandos a utilizar<sup>7</sup>:

**destring:** Convierte variables en cadena a numéricas. En otras palabras, permite desencadenar variables.

Su estructura es:

**destring** variable, **gen**(nueva\_variable)

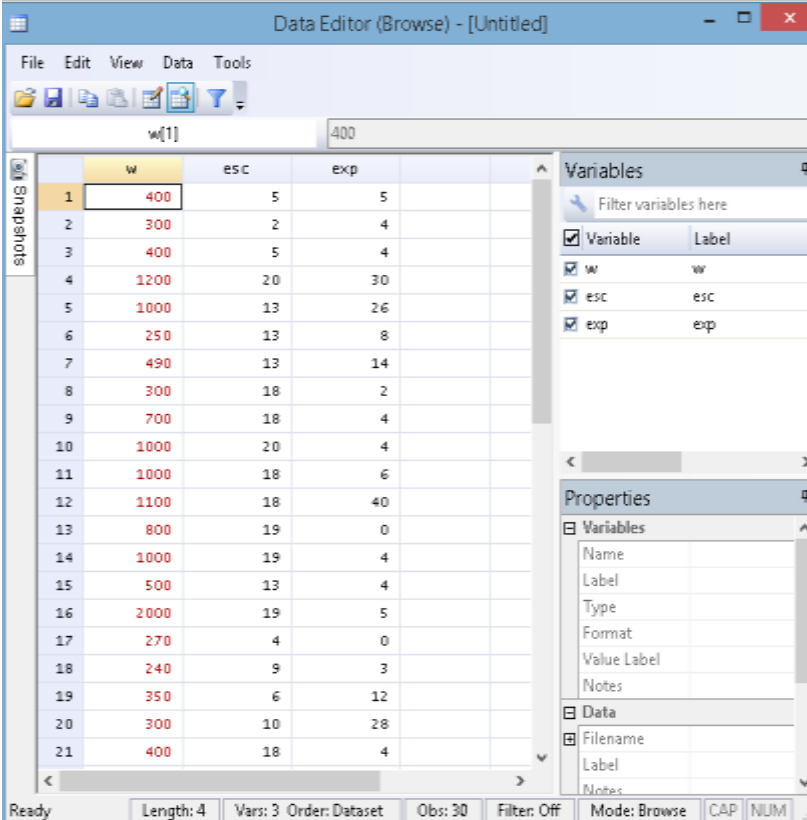
Rutina utilizada:

**destring** w, **gen**(salario)

Desarrollo:

Se introducen los datos del salario, nivel de escolaridad y experiencia en Stata (Tabla 3.1.1).

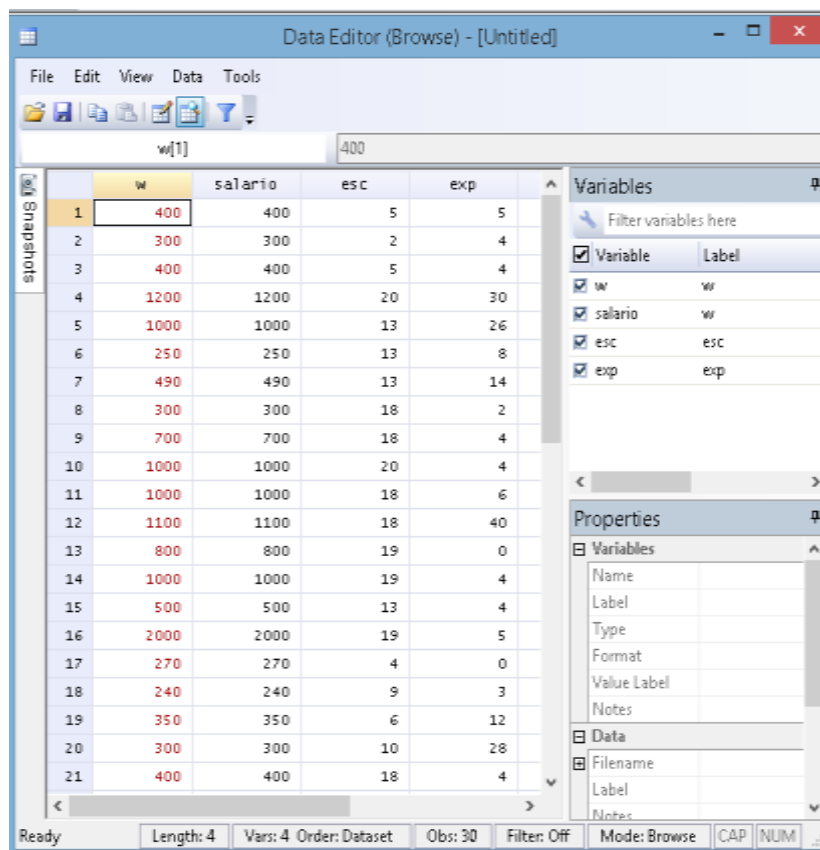
Pero al observar los datos haciendo uso del ícono *data editor*,  el salario “w” aparece así:



	w	esc	exp
1	400	5	5
2	300	2	4
3	400	5	4
4	1200	20	30
5	1000	13	26
6	250	13	8
7	490	13	14
8	300	18	2
9	700	18	4
10	1000	20	4
11	1000	18	6
12	1100	18	40
13	800	19	0
14	1000	19	4
15	500	13	4
16	2000	19	5
17	270	4	0
18	240	9	3
19	350	6	12
20	300	10	28
21	400	18	4

Se aprecia que la variable “w” está de color rojo, es decir encadenada. Para corregir este tipo de error se ejecuta el comando: **destring** w, **gen**(salario). Con ello ya estará convertida la variable en cadena en numérica con su nuevo nombre “salario”, y para comprobarlo se pueden observar nuevamente las variables.

<sup>7</sup> Se puede utilizar el comando **tostring** para convertir variables numéricas a textuales (en cadena).



Se puede evidenciar que “salario” es la variable numérica creada a partir de la variable encadenada “w”.

### 3.1.3.3 Eliminar variables.

#### Ejercicio 5:

A menudo es necesario eliminar una o varias variables que no se van a utilizar o no son necesarios para el trabajo en Stata, y en otros casos se quieren eliminar todas las variables que se introducen al programa. Para ello, se utilizan algunos comandos y acciones que ayudarán a cumplir con dicha tarea.

Comandos a utilizar:

- **drop**: Permite eliminar la variable que se desee.  
Su estructura es:  
**drop** variable
- **clear all**: Permite eliminar toda la lista de variables que están introducidas en Stata.  
También se puede utilizar solamente **clear**.

Rutina utilizada:

**drop w**  
**clear all**

Desarrollo:

- **Eliminar una variable.**

Continuando con el ejercicio 5 se tienen dos variables con los mismos datos (salario & w). Se va a proceder a eliminar la variable “w” que es aquella que ya no interesa. Para ello, se ejecuta el siguiente comando: **drop w**. Con ello la variable “w” se ha eliminado, por lo tanto, ya no aparece en la lista de variables.

Variables		Y	q	x
Variable	Label			
salario	salario			
esc	esc			
exp	exp			

Otra manera de eliminar alguna variable es haciéndolo directamente desde la lista de variables. Se da clic derecho sobre el nombre de la variable, en este caso sobre “w”, se selecciona la opción *Drop Selected Variables*, y damos clic en *Sí*.

- **Eliminar todas las variables.**

Continuando con el mismo ejemplo, ahora se quiere eliminar todas las variables introducidas en Stata. Para ello, se ejecuta el siguiente comando: **clear all**. De esta manera, se podrá observar que ya no existen variables en la base de datos.

Variables		Y	q	x
Variable	Label			
There are no items to show.				

### 3.1.3.4 Crear variables.

#### **Ejercicio 6:**

En ocasiones es necesario crear nuevas variables para utilizarlas en nuestro trabajo. Estas nuevas variables se pueden generar a partir de aquellas ya existentes en la base de datos de Stata según sea el caso.

Comandos a utilizar:

**generate:** Permite crear una nueva serie o variable. También se puede escribir en forma abreviada como **gen** o **gener**.

Ejemplos:

**generate** nueva\_variable = var1+ var2

**generate** nueva\_variable = var1\*var2

**generate** nueva\_variable = var1^2

Rutina utilizada:

**generate** exp\_cuadrado=exp^2

Desarrollo:

Como el modelo básico de salarios de Mincer (1974) establece que el salario está en función del nivel de escolaridad, experiencia y experiencia al cuadrado. Por consiguiente, tomando la información de la *Tabla 3.1.1* se va a crear la variable experiencia al cuadrado que es aquella que falta para completar las variables incluidas dicho modelo de salarios.

Para crear la variable experiencia al cuadrado ejecutamos el comando:

**generate** exp\_cuadrado=exp^2

En donde, “exp\_cuadrado” es el nombre de la nueva variable a creada. Y se escribió  $exp^2$  porque se quiere que la nueva variable sea igual a la experiencia elevada al cuadrado. Finalmente se puede observar que ya se ha creado la nueva variable “exp\_cuadrado”.

Variables		Y	+	x
Variable	Label			
ww	ww			
esc	esc			
exp	exp			
exp_cuadrado				

### 3.1.3.5 Renombrar y etiquetar variables.

#### **Ejercicio 7:**

Se suelen tener variables codificadas en su nombre, por lo que es aconsejable renombrarlas para facilitar el trabajo y utilizar correctamente la información. Por otro lado, se puede etiquetar las variables introducidas al programa con la palabra o palabras que resulten más pertinentes. La utilización de etiquetas (*label*) servirá para tener una información más clara de las variables al momento de trabajar en Stata. Estas tareas se pueden realizar de dos maneras:

- Desde *Properties* con la opción *Lock/Unlock*.
- Utilizando los comandos **rename** y **label** según sea el caso.

Comandos a utilizar:

- **rename:** Permite cambiar el nombre de una variable. Es decir, renombrarla.  
Su estructura es:  
**rename** variable nueva\_variable
- **label:** Permite etiquetar variables. Las etiquetas pueden ser máximo de 80 caracteres.  
Su estructura es:  
**label variable** variable “etiqueta”

Rutina utilizada:


**rename** w S

**label variable** w “Salario”

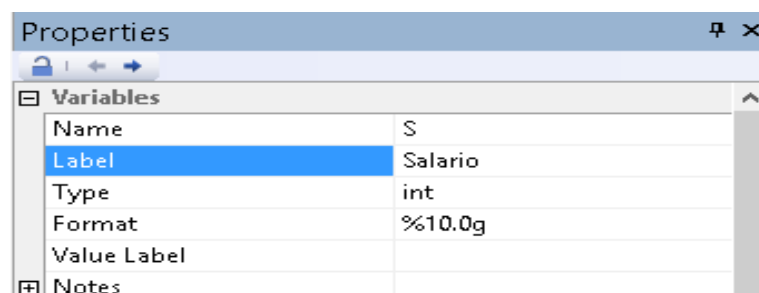
Desarrollo:

Tomando nuevamente la información de la *Tabla 3.1.1* se pretende editar las etiquetas de las variables introducidas, y luego renombrar dichas variables. Para ello, se lo hará de las dos maneras ya mencionadas.

- **Desde *Properties* con la opción *Lock/Unlock*.**

Para desbloquear la edición de las propiedades de una variable, se selecciona la variable a editar dando un clic sobre ella, y posteriormente en *Properties* (ubicada en la parte derecha inferior de la pantalla principal del programa) se da clic en el ícono *Lock/Unlock* . En consecuencia se han desbloqueado las propiedades de la variable “S” para su edición.

Ahora bien, como se quiere renombrar y etiquetar la variable se hace uso de las propiedades: *name* y *label*, y se prosigue a editar la información así:



Con ello ya se ha cambiado el nombre de la variable, ya que ha pasado de “w” a “S”, y su etiqueta también se ha editado con la palabra “Salario”<sup>8</sup>.

---

<sup>8</sup> Para colocar la etiqueta es necesario que el texto que se introduzca como etiqueta vaya entre comillas.

- Utilizando los comandos *rename* y *label*.

Se ejecutan los siguientes comandos:

**rename** w S

**label variable** w “Salario”

### 3.1.3.6 Borrar y editar variables.

#### Ejercicio 8:

En trabajos de investigación y análisis estadísticos en general es necesario eliminar y editar datos para tener una base de datos pertinente para el propósito a conseguir. Por lo que se suelen desechar un grupo de datos, o simplemente se transforman datos de acuerdo a la necesidad del usuario.

Comandos a utilizar:

- **drop if:** Es una variación del comando **drop**, el cual permite eliminar un rango de observaciones determinadas.  
Su estructura es:  
**drop if** variable>=número  
**drop if** variable<=número  
**drop if** variable==número
- **replace:** Reemplaza el contenido de un variable.  
Su estructura es:  
**replace** variable=número **if** variable>=número  
**replace** variable=número **if** variable<=número  
**replace** variable=número **if** variable==número  
**replace** variable=número **if** variable==.

Rutina utilizada:

**replace** sexo=0 **if** sexo==2  
**drop if** edad<15  
**drop if** edad>29

Desarrollo:

Haciendo uso de la *Tabla 3.1.2*, se quiere obtener información del sexo solamente de jóvenes, es decir de aquellos con edad comprendida entre 15 a 29 años.

Primeramente, observando los datos de la variable sexo, aparece 1 para masculino y 2 para femenino. Sin embargo, se prefiere tener 0 para femenino, por lo cual es necesario reemplazar este valor ejecutando el siguiente comando: **replace** sexo=0 **if** sexo==2.

Por otro lado, para obtener solamente los valores de los jóvenes (15 a 29 años), se procede a eliminar las edades que no se requieran. Para ello, se ejecutan los siguientes comandos:

**drop if** edad<15

**drop if** edad>29

### 3.1.4 Tablas.

#### Ejercicio 9:

Las tablas son de mucha utilidad para el análisis de datos, especialmente cuando se quiere cruzar variables. Las tablas de contingencia ayudan a analizar la asociación de dos o más variables que generalmente son de naturaleza cualitativa (nominal u ordinal). En Stata es posible cruzar dos variables haciendo uso de comandos para el efecto. A este tipo de tablas de contingencia también se las denomina comúnmente como tablas de doble entrada.

Comandos a utilizar:

- **table**: Permite clasificar los datos de una variable, brindando información de la proporción (en términos absolutos) de cada categoría de la variable. También permite generar tablas de doble entrada para analizar la distribución conjunta de dos variables.

Su estructura es:

**table** variable (tabla de una variable)  
**table** var1 var2 (tabla de doble entrada)

- **tab**: También conocido como *tabulate*. Clasifica los datos de una variable, brindando información de la proporción (en términos absolutos y relativos) de cada categoría de la variable. También permite generar tablas de doble entrada para analizar la distribución conjunta de dos variables. Estas tablas incluyen el “total” en la columna derecha y fila inferior.

Su estructura es:

**tab** variable (tabla de una variable)  
**tab** var1 var2 (tabla de doble entrada)

Rutina utilizada:

**table** cond\_empleo

**tab** cond\_empleo

**tab** cond\_empleo sexo

Desarrollo:

Utilizando los datos de la Tabla 3.1.3, primeramente se quiere conocer cuál es la condición de empleo de los habitantes de la ciudad de Loja. Para ello, se realizará una tabla que clasifique los datos de la variable “cond\_empleo”.

Ejecutando el comando: **table** cond\_empleo, se obtiene la siguiente tabla:

cond_empleo	Freq.
desempleados	53
empleados	1,587

Y ejecutando el comando: **tab** cond\_empleo, se obtiene la presente tabla:

cond_empleo	Freq.	Percent	Cum.
desempleados	53	3.23	3.23
empleados	1,587	96.77	100.00
Total	1,640	100.00	

La tabla ejecutada con el comando **tab** nos brinda mayor información que la tabla ejecutada con el comando **table**.

Ahora, se va a averiguar la condición de empleo en la ciudad de Loja según el sexo. Para ello, se realiza una tabla de contingencia que cruce la variable “cond\_empleo” y “sexo”, ejecutando el siguiente comando: **tab** cond\_empleo sexo

Y se obtiene la siguiente tabla de contingencia:

cond_empleo	Freq.	Percent	Cum.
desempleados	53	3.23	3.23
empleados	1,587	96.77	100.00
Total	1,640	100.00	

Por último, se quiere averiguar cuáles fueron las razones más influyentes por las cuales el 3.23% hombres y mujeres lojanas están desempleados. Para conocer dicha información se ejecuta el siguiente comando: **tab** motivo\_desempleo sexo

Y se obtiene la siguiente tabla de contingencia:

motivo_desempleo	sexo		Total
	hombre	mujer	
falta de oportunita..	19	22	41
motivos personales ..	4	6	10
resultado de sus er..	1	1	2
Total	24	29	53

Interpretación:

El 3.23% de la población en edad de trabajar de la ciudad de Loja es desempleada. Estas personas desempleadas lo están por 3 motivos: el motivo más predominante tanto en hombres como mujeres es por la falta de oportunidades, seguido por motivos personales y por último por el resultado de errores del pasado, razón que es menos predominantes tanto en hombres como mujeres. Además, se evidencia que hay más mujeres desempleadas que hombres desempleados, pues 29 de cada 53 personas desempleadas son mujeres y 24 de cada 53 son hombres. Analizando el motivo de desempleo, tenemos que 22 de cada 29 mujeres



desempleadas y 19 de cada 24 hombres desempleados están en dicha condición por falta de oportunidades. 6 de cada 29 mujeres desempleadas y 4 de cada 24 hombres desempleados están en esta condición por motivos personales, y 1 de cada 29 mujeres y 1 de cada 24 hombres son personas desempleadas resultado de sus errores del pasado.

### 3.2 Naturaleza del análisis de regresión

“El análisis de regresión trata del estudio de la dependencia de una variable dependiente  $Y$  respecto de una o más variables explicativas  $X$ ” (Gujarati & Porter, 2010, p.15). Por lo tanto, la naturaleza del análisis de regresión radica en el grado de relación que existe entre dos variables.

Se puede conocer la relación entre dos variables (antes de realizar una predicción) mediante la elaboración de gráficas, principalmente por medio de diagramas de dispersión. Es por ello, que se desarrollan diez ejercicios en los cuales se grafican diagramas de dispersión, gráficos de series de tiempo, gráficos de barras y gráficos de sectores o de pastel. Para realizar dichos gráficos se utilizan nueve tablas (Tabla 3.2.1–Tabla 3.2.9) con información económica, social y ambiental, extraída del Banco Mundial (2015), MAGAP (2015), e INEC (2010).

#### 3.2.1 Diagramas de dispersión.

Los diagramas de dispersión son gráficos que representan la relación y asociación existente entre dos variables.

##### 3.2.1.1 Diagrama de dispersión simple.

#### Ejercicio 1:

También se lo denomina nube de puntos. Este tipo de diagrama de dispersión nos permite identificar la tendencia y relación de los datos entre dos variables. Mientras más dispersos estén los datos, menor será la relación entre las dos variables.

Comandos a utilizar:

**scatter:** Permite graficar diagramas de dispersión.

Su estructura es:

**scatter** varY varX

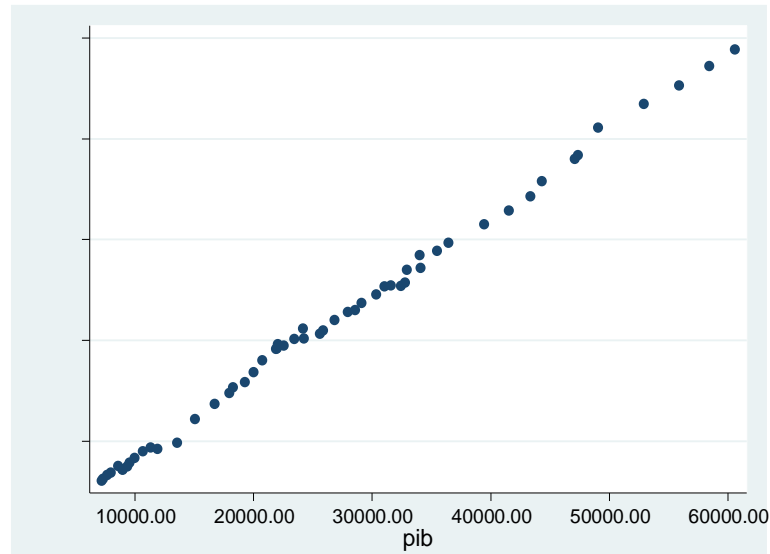
Rutina utilizada:

**scatter** consumo pib

Desarrollo:

Se va a graficar un diagrama de dispersión con las variables gasto en consumo y pib (Tabla 3.2.1). Sabiendo que,  $\text{consumo} = f(\text{pib})$ , se ejecuta el siguiente comando: **scatter** consumo pib

Y se obtiene el diagrama de dispersión:



Interpretación:

Existe una relación positiva casi lineal entre el gasto de consumo y el pib de Ecuador. Pues a mayor pib, mayor gasto en consumo. Por lo tanto, se podría esperar que el coeficiente de correlación entre estas variables sea cercano a 1.

### 3.2.1.2 Diagrama de dispersión con ajuste lineal.

#### Ejercicio 2:

Este tipo de diagrama de dispersión agrega una línea de tendencia lineal, la cual permite apreciar que tan dispersos están los datos respecto a su media.

Comandos a utilizar:

- **twoway:** Permite unir en una misma línea de comandos dos o más comandos para realizar gráficos.  
Su estructura es:  
**twoway** (*comando1* var1 var2)...(*comando\_n* var1 var 2)
- **scatter:** Permite graficar diagramas de dispersión.  
Su estructura es:  
**scatter** varY varX
- **lfit:** Permite agregar una línea de tendencia de ajuste lineal.  
Su estructura es:  
**lfit** varY varX

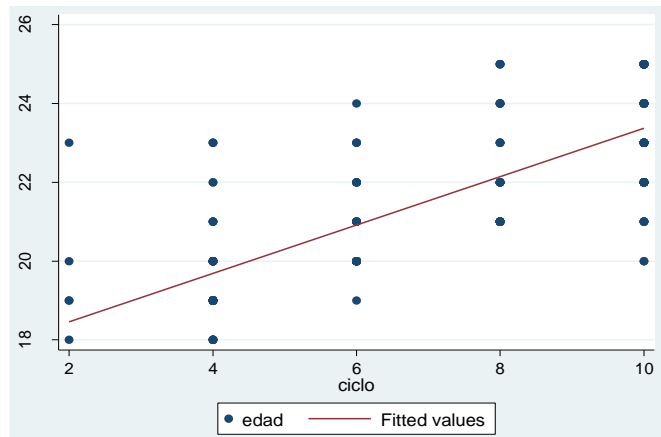
Rutina utilizada:  
**twoway** (**scatter** edad ciclo) (**lfit** edad ciclo)

Desarrollo:

A partir de la Tabla. 3.2.2 que contiene información de estudiantes de la Universidad Técnica Particular de Loja respecto a la edad y ciclo, se va a graficar un diagrama de dispersión con ajuste lineal entre las dos variables. Por consiguiente, se ejecuta el siguiente comando:

**twoway (scatter edad ciclo) (lfit edad ciclo)**

Y se obtiene el siguiente diagrama de dispersión con ajuste lineal:



Interpretación:

Las variable ciclo y edad se relación positivamente. Si el estudiante se encuentra en un ciclo mayor tiende a tener mayor edad. Sin embargo, notamos que los datos si se dispersan considerablemente de su valor medio.

### 3.2.1.3 Diagrama de dispersión con ajuste polinómico.

#### Ejercicio 3:

Estos diagramas de dispersión se caracterizan por presentar una tendencia curvilínea. Por lo que el comportamiento de las variables puede ser polinómico, ya sea cuadrático (orden 2) o de un orden mayor.

Comandos a utilizar:

- **qfit:** Permite agregar una línea de tendencia cuadrática.  
Su estructura es:  
**qfit varY varX**
- **fpfit:** Permite agregar una línea de tendencia polinómica  
Su estructura es:  
**fpfit varY varX**

Rutina utilizada:

**twoway (scatter esp\_vida tasa\_mortalidad ) (qfit esp\_vida tasa\_mortalidad )**  
**twoway (scatter esp\_vida tasa\_mortalidad ) (fpfit esp\_vida tasa\_mortalidad )**

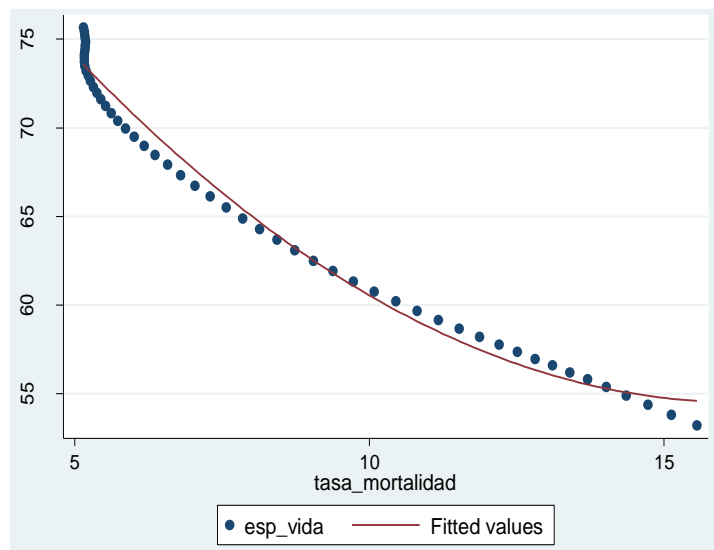
Desarrollo:

Utilizando los datos la Tabla 3.2.3 que contiene información de esperanza de vida al nacer y tasa de mortalidad en un año por cada 1000 personas, se va a graficar un diagrama de dispersión con ajuste cuadrático y otro con ajuste polinómico.

Para el diagrama de dispersión con ajuste cuadrático se ejecuta el siguiente comando:

**twoway (scatter esp\_vida tasa\_mortalidad) (qfit esp\_vida tasa\_mortalidad)**

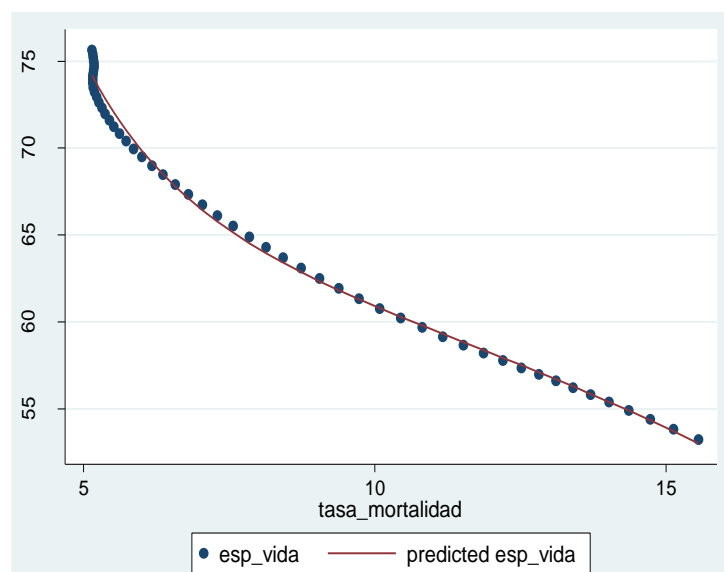
Se obtiene el siguiente diagrama de dispersión:



Para el diagrama de dispersión con ajuste polinómico ejecutamos el siguiente comando:

**twoway (scatter esp\_vida tasa\_mortalidad) (fpfit esp\_vida tasa\_mortalidad)**

Se obtiene el siguiente diagrama de dispersión:



Interpretación:

Existe una relación inversa (negativa) entre la esperanza de vida y la tasa de mortalidad. A mayor tasa de mortalidad infantil menor será la esperanza de vida. Por otro lado, los datos de estas dos variables se ajustan de mejor manera a un ajuste polinómico.

#### 3.2.1.4 Diagrama de dispersión con línea de tendencia múltiple.

##### Ejercicio 4:

Los diagramas de dispersión también se usan como herramienta de análisis y comparación de diferentes conjuntos de datos. Unifica dos o más diagramas de dispersión en un solo gráfico.

Comandos a utilizar:

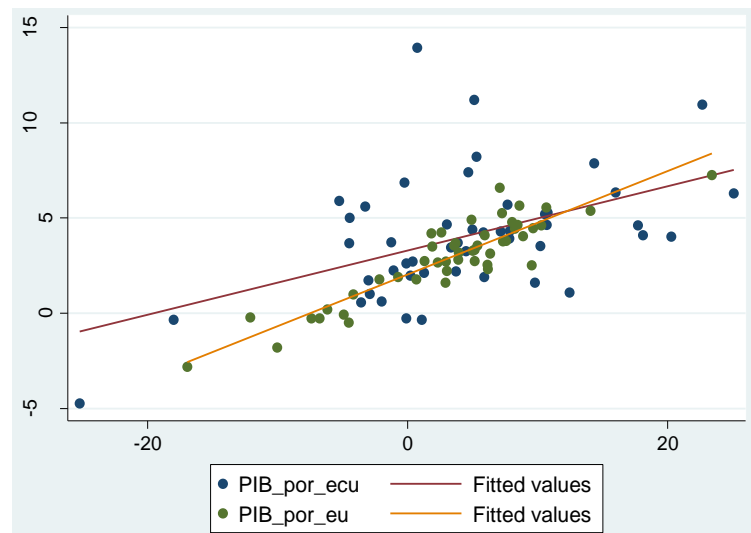
Rutina utilizada:

```
twoway (scatter PIB_por_ecu FBKF_por_ecu) (lfit PIB_por_ecu FBKF_por_ecu) (scatter PIB_por_eu FBKF_por_eu) (lfit PIB_por_eu FBKF_por_eu)
```

Desarrollo:

Se quiere realizar un gráfico de dispersión múltiple que compare la relación entre el PIB y la formación bruta de capital fijo (fbkf) de Ecuador y Estados Unidos. Para ello se utilizarán los datos de la Tabla 3.2.4. Se procede a ejecutar el siguiente comando: **twoway (scatter PIB\_por\_ecu FBKF\_por\_ecu) (lfit PIB\_por\_ecu FBKF\_por\_ecu) (scatter PIB\_por\_eu FBKF\_por\_eu) (lfit PIB\_por\_eu FBKF\_por\_eu)**

Y se obtiene el siguiente diagrama de dispersión múltiple:



Los datos de color azul representan a los de Ecuador y los datos de color verde a los de Estados Unidos.

Interpretación:

Comparando los datos de ambos países en análisis, se puede afirmar que los datos de Estados Unidos tienen una relación positiva mayor entre el PIB y FBKF que los datos de Ecuador. Esto se lo evidencia mediante la inclinación de las líneas de tendencia, puesto que la de Estados Unidos se encuentra con mayor pendiente.

### 3.2.2 Gráficos de series de tiempo.

En Stata es posible graficar una o más series de tiempo en un mismo gráfico.

#### 3.2.2.1 Evolución de una serie de tiempo.

#### Ejercicio 5:

Es importante conocer la evolución que ha tenido una variable con el paso del tiempo. Esta evolución podría mostrar una tendencia clara en unos casos, pero no muy clara en otros. Por lo general, las series de tiempo presentan muchas fluctuaciones, e inclusive algunas variables producen cambios demasiado altos, conocidos como “picos” en determinado año, semestre, mes, día, etc.

Comandos a utilizar:

- **line:** Permite trazar una línea evolutiva de una variable de serie de tiempo.  
Su estructura es:  
**line** varY varX
- **twoway connected:** Permite conectar mediante líneas los puntos (datos) de una variable de serie de tiempo.  
Su estructura es:  
**twoway connected** varY varX

Rutina utilizada:

**line** M2 anio

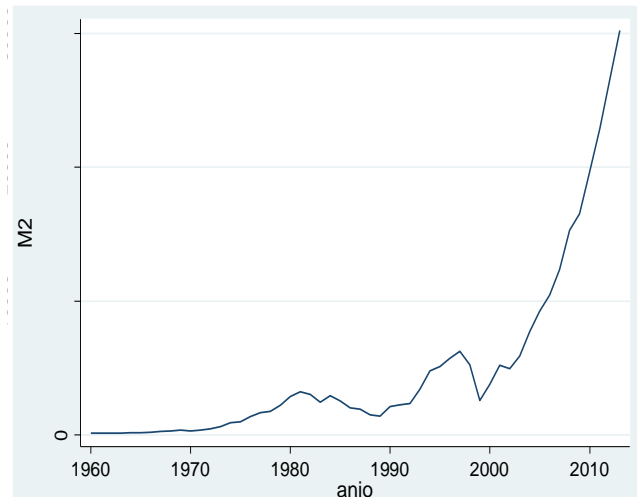
**twoway connected** M2 anio

Desarrollo:

Se va a conocer el comportamiento de la variable cantidad de dinero (M2) de Ecuador en el período 1960-2013 (Tabla 3.2.5). Se realiza un gráfico de línea, y otro de línea y puntos.

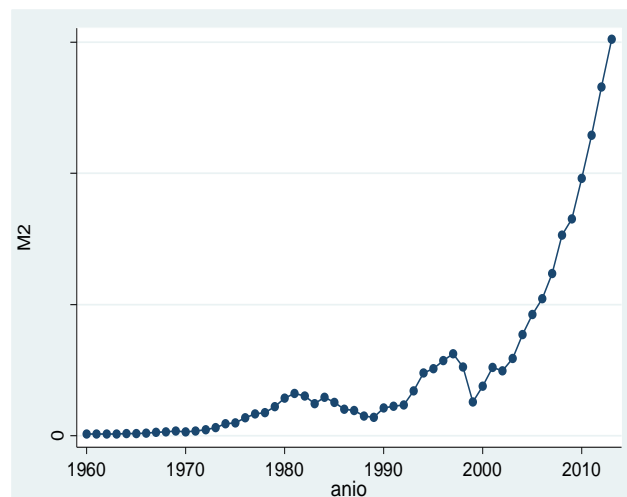
Para conocer la evolución de M2 mediante un gráfico solamente de línea, se ejecuta el siguiente comando: **line** M2 anio

Y se obtiene:



Ahora, para conocer la evolución de M2 mediante un gráfico de líneas y puntos, se ejecuta el siguiente comando: **twoway connected M2 año**

Y se obtiene:



Interpretación:

Desde 1960 a 1975 la cantidad de dinero (M2) de Ecuador ha permanecido estable, de 1976 hasta 1998 ha sufrido pequeñas variaciones. Pero en el período de 1999 al 2013 M2 ha presentado una tendencia muy positiva “disparándose hacia arriba”, período en el cual el país ha estado dolarizado.

### 3.2.2.2 Evolución de dos o más series de tiempo.

#### Ejercicio 6:

Se puede comparar la evolución de dos o más series de tiempo en un solo gráfico. Esto con el fin de averiguar el comportamiento de una variable en comparación con otras u otras a

medida que pasa el tiempo. Además estas gráficas ayudan a realizar un análisis conjunto, ya que la aceleración o desaceleración de una variable podría reflejar el crecimiento o disminución de otra.

Comandos a utilizar:

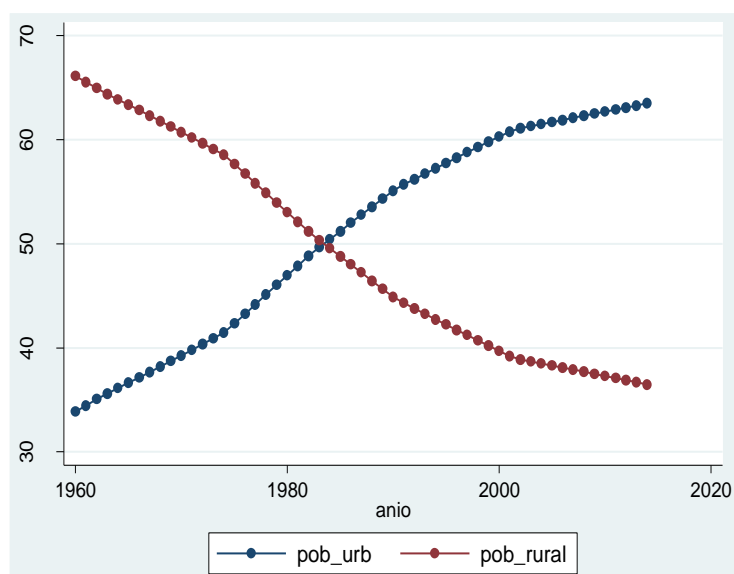
Rutina utilizada:  
**twoway (connected pob\_urb anio) (connected pob\_rural anio)**

Desarrollo:

Se quiere graficar en un mismo gráfico la evolución de la población urbana y rural del Ecuador en el período 1960-2014 (Tabla 3.2.6). Como lo se mencionó en el ejercicio 5, este tipo de gráficos se los puede hacer de dos maneras: solamente con líneas, o con líneas y puntos. En este caso se graficará de la segunda manera. Para ello se ejecuta el siguiente comando:

**twoway (connected pob\_urb anio) (connected pob\_rural anio)**

Y se obtiene el siguiente gráfico:



Interpretación:

En Ecuador la población urbana superó a la población rural en el año 1984. A partir de dicho año la brecha entre población urbana y rural ha ido creciendo con el paso del tiempo. La población rural ha migrado a las ciudades en busca de una mejor calidad de vida.

### 3.2.3 Gráficos de barras.

Los gráficos de barras pueden ser simples o agrupados. Además las barras de estos gráficos pueden ser de forma vertical u horizontal. A continuación se realizan gráficos de barras tanto simples como agrupados.



### 3.2.3.1 Gráfico de barras simple.

#### Ejercicio 7:

Los gráficos de barras es otro tipo de gráficos que representan un conjunto de datos mediante columnas o filas rectangulares. Es decir estos gráficos de barras pueden ser horizontales y verticales. Estos comparan las alturas de las barras de cada categoría.

Comandos a utilizar:

- **graph bar:** Permite realizar gráficos de barras verticales.  
Su estructura es:  
**graph bar** variableY, **over** (variable\_categorica)
- **graph hbar:** Permite realizar gráficos de barras horizontales.  
Su estructura es:  
**graph hbar** variableY, **over**(variable\_categorica)
- **if:** Es un condicional, que nos permite utilizar solo los datos que están bajo una condición (sujeta a las necesidades del usuario).  
Ejemplos:  
**if** variable==valor  
**if** variable<valor  
**if** variable>valor
- **over:** Permite agrupar los datos de acuerdo a su categoría.  
Su estructura es:  
**over**(variable\_categorica)

Rutina utilizada:

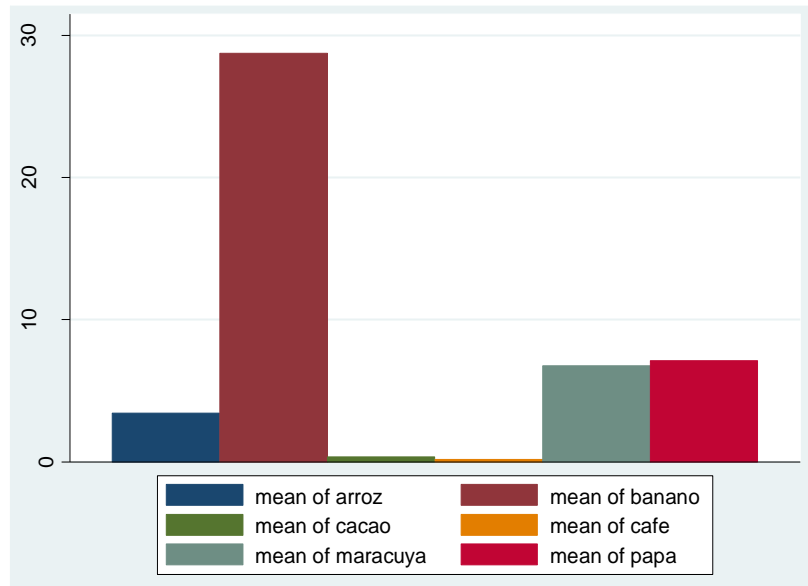
```
graph bar arroz banano cacao cafe maracuya papa  
graph bar arroz banano cacao cafe maracuya papa if anio==2010  
graph bar rendimiento, over(producto)  
graph hbar rendimiento, over(producto)
```

Desarrollo:

Haciendo uso de la Tabla 3.2.7 se va a comparar el rendimiento promedio de algunos productos agrícolas más importantes en el sector agrícola ecuatoriano en el período 2000-2012. Por consiguiente, se ejecuta el comando:

**graph bar** arroz banano cacao cafe maracuya papa

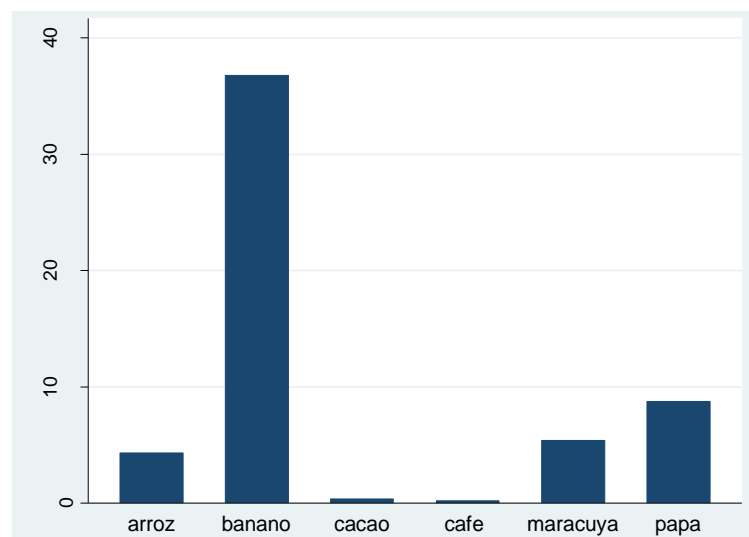
Como cada variable presenta el rendimiento del producto en cada año, se tienen 13 datos de cada producto y por ello Stata calcula el rendimiento promedio de cada uno para presentarlo en el gráfico de barras que se observa a continuación:



Ahora, si se quiere solamente comparar el rendimiento de estos productos en un determinado año, por ejemplo del año 2010, hacemos uso del comando **if**. Entonces, el comando a ejecutar sería el siguiente: **graph bar** arroz banano cacao cafe maracuya papa **if** anio==2010

Por otro lado, si tenemos datos de serie transversal como los de la Tabla 3.2.7. Hacemos uso del siguiente comando: **graph bar** rendimiento, **over**(producto)

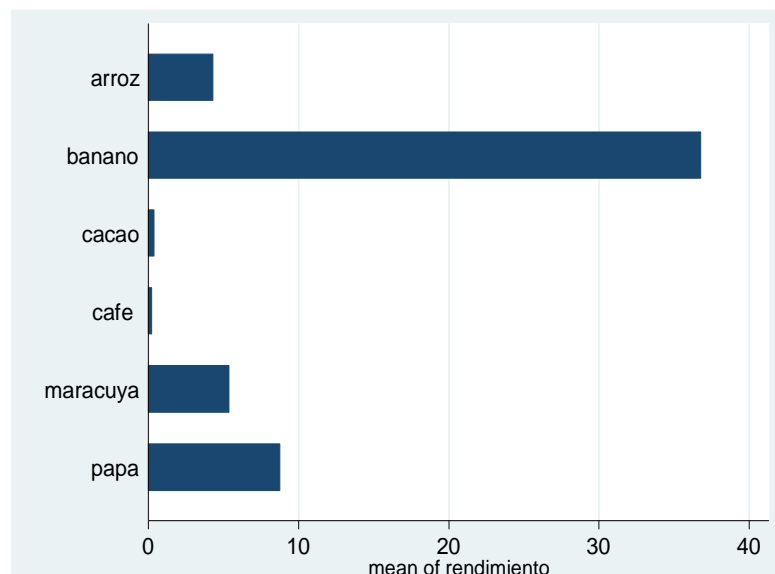
El “rendimiento” es la variable Y y el “producto” es la variable categórica. Por lo tanto, se obtendrá el presente gráfico de barras:



Finalmente, para realizar el gráfico con barras horizontales. Se utiliza el siguiente comando:

**graph hbar** rendimiento, **over**(producto)

Y se obtiene el gráfico de barras así:



Interpretación:

El banano es el producto agrícola con mayor rendimiento. La papa y la maracuyá rinden casi la tercera parte del rendimiento del banano. Seguido por el arroz, luego el cacao y por último el café con un bajo rendimiento.

### 3.2.3.2 Gráfico de barras agrupado.

#### Ejercicio 8:

Un tipo particular de gráficos de barras son los agrupados. Estos presentan dos o más conjuntos de datos en el mismo gráfico.

Comandos a utilizar:

**blabel:** Permite agregar etiquetas a los datos, es decir el valor de cada barra.

*Su estructura es:*

**blabel** (total)

Rutina utilizada:

**graph bar** arroz banano cacao cafe maracuya papa, **over** (anio)

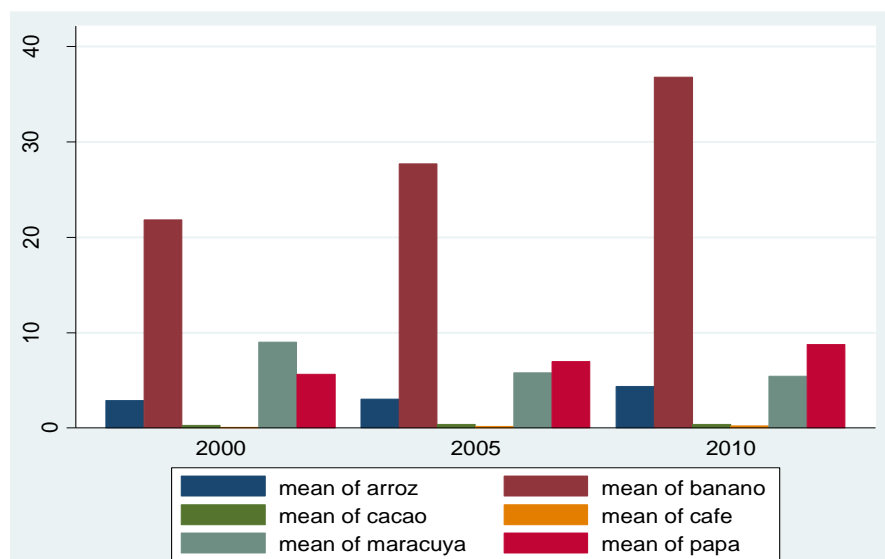
**graph bar** arroz banano cacao cafe maracuya papa, **over** (anio) **blabel** (total)

Desarrollo:

Se tienen los datos del rendimiento de algunos productos agrícolas importantes del Ecuador de los años 2000, 2005, y 2010 (Tabla 3.2.8). Con estos datos se comparará el rendimiento de los productos en los tres años mencionados. Para ello se realizará un gráfico de barras agrupado usando el siguiente comando:

**graph bar** arroz banano cacao cafe maracuya papa, **over** (anio)

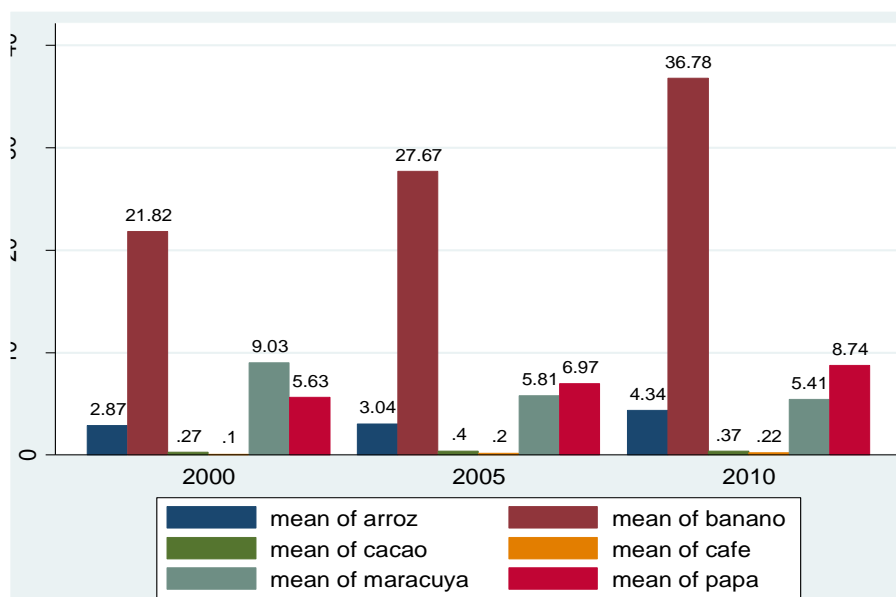
Y se obtiene:



Se observa que este gráfico, así como los hechos en el ejercicio anterior, no presentan los valores de las barras, lo que de alguna manera dificulta conocer el valor exacto del rendimiento de cada producto. Para evitarnos este problema, agregamos las etiquetas de los datos agregando el comando **blabel** así:

**graph bar** arroz banana cacao cafe maracuya papa, **over** (anio) **blabel** (total)

De esta manera se obtiene el gráfico de barras con los respectivos valores en cada una de las barras.



Interpretación:

El rendimiento de banano es mucho mayor en el año 2010, en todos los años ha sobrepasado con mucho a los demás productos en análisis. La maracuyá tuvo un rendimiento mayor en el año 2000 respecto al año 2005 y 2010. La papa y el arroz han aumentado su rendimiento para el año 2010, pero este incremento no ha sido muy amplio. El cacao y el café han mantenido su bajo nivel de rendimiento en los tres años de comparación.

### 3.2.4 Gráficos de sectores.

#### Ejercicio 9:

Son comúnmente conocidos como gráficos de pastel, ya que “representan proporcionalmente la frecuencia de los valores contenidos de una o más variables” (Escobar, Fernández, & Bernardi, 2012, p.153).

Comandos a utilizar:

- **graph pie:** Permite realizar gráficos de sectores.  
Su estructura es:  
**graph pie, over** (variable\_categorica)
- **title:** Permite poner el título a un gráfico.  
Su estructura es:  
**title**("título")
- **by:** Permite presentar los datos en subconjuntos.  
Su estructura es:  
**by** (variable\_categorica)

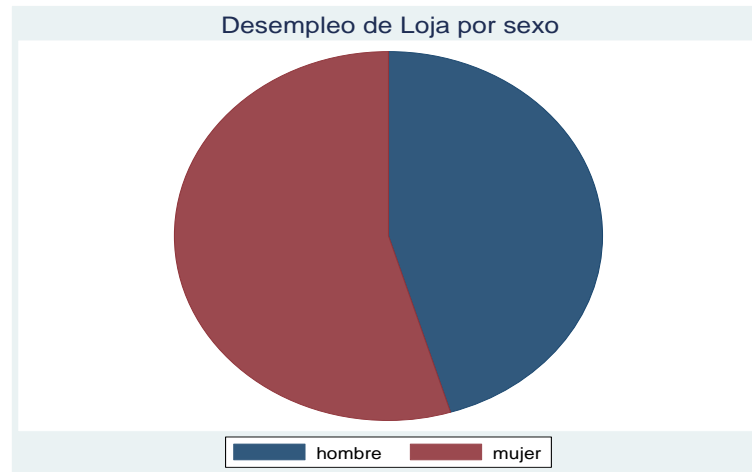
Rutina utilizada:

```
graph pie, over (sexo)  
graph pie, over (sexo) title("Desempleo de Loja por sexo")  
graph pie, over (sexo) by (motivo_desempleo)
```

Desarrollo:

Se quiere averiguar la proporción de hombres y mujeres desempleados en la ciudad de Loja mediante un gráfico de sectores o pastel (Tabla 3.2.9). Para ello se hace uso del siguiente comando: **graph pie, over** (sexo). Antes de ejecutarlo le vamos a agregar el comando **title** para ponerle un título al gráfico. Entonces el comando a ejecutar quedaría de la siguiente manera: **graph pie, over** (sexo) **title**("Desempleo de Loja por sexo")

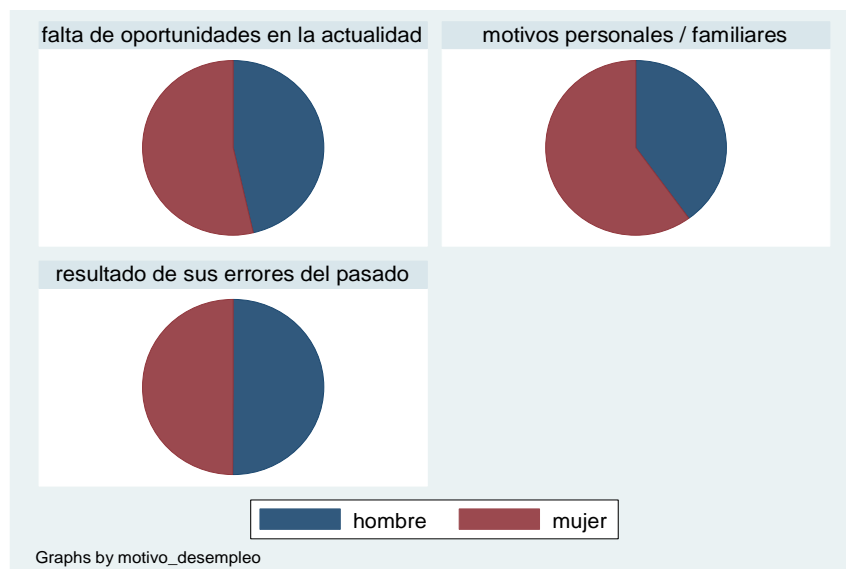
Y se obtiene el siguiente gráfico de pastel:



Ahora, se va a agregarle al gráfico una nueva variable que es “motivo\_desempleo”, con el fin de averiguar la proporción de mujeres y hombres desempleados por cada motivo de desempleo. Por lo tanto, se va a ejecutar el siguiente comando:

**graph pie, over (sexo) by (motivo\_desempleo)**

Y se obtiene un gráfico con múltiples gráficos de sectores, en donde cada gráfico pertenece a cada categoría de la variable añadida.



Interpretación:


La mayor parte de las personas desempleadas en ciudad de Loja son mujeres. Principalmente lo son por motivos personales y familiares, seguido de la falta de oportunidades y por resultado de sus errores del pasado. Este último motivo de desempleo es igual de frecuente para mujeres y hombres de la ciudad.

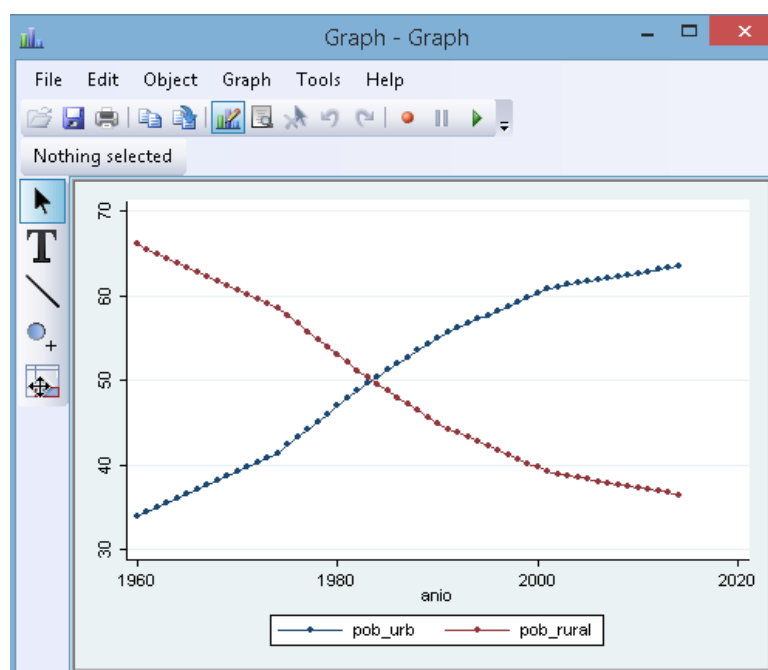
### 3.2.5 Editor de gráficos.

#### Ejercicio 10:

Stata 13 incorpora un editor de gráficos. Una vez que se ha producido el gráfico el usuario puede hacer modificaciones al mismo, sin necesidades de ejecutar complejos comandos.

Desarrollo:

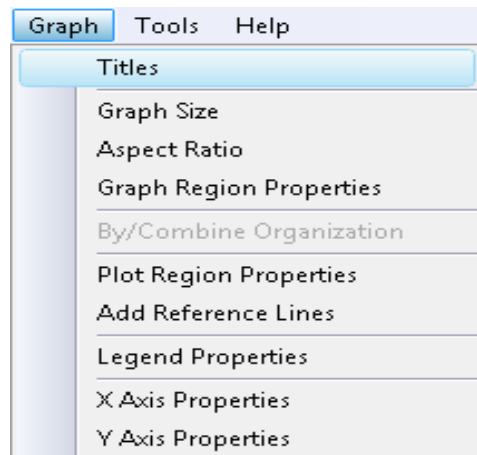
Supongamos que se quiere modificar el gráfico realizado en el ejercicio 6. Por consiguiente, para empezar a editar el gráfico, se da clic en el ícono . Y aparece la pantalla del editor de gráficos de esta manera:



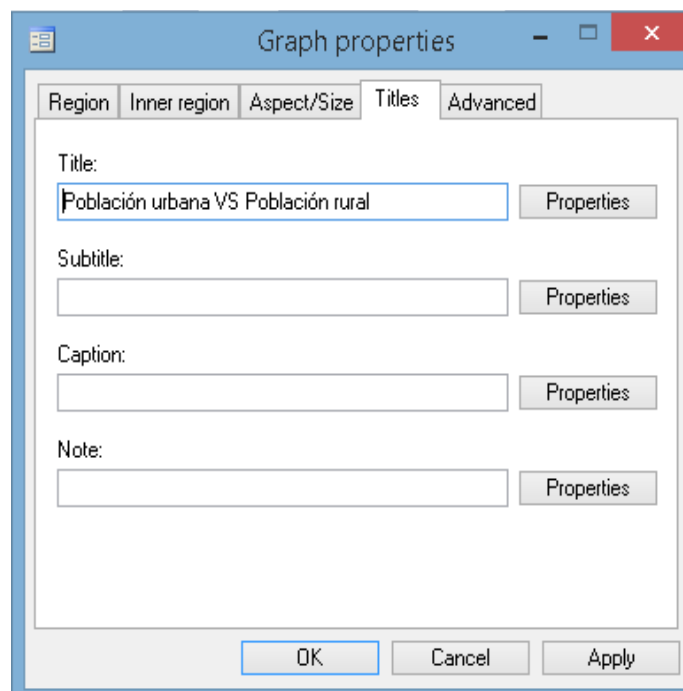
Solamente “se trata de buscar el elemento que se quiera cambiar, señalarlo con el cursor, a partir de lo cual se marca automáticamente en rojo, averiguar si lo que se desea cambiar está contenido en el nuevo menú horizontal que surge debajo de la barra de íconos, pulsar el botón derecho el ratón, porque generalmente en la última línea del menú contextual emergente se encuentra el acceso a las propiedades” (Escobar, Fernández, & Bernardi, 2012, p.191).

También se puede hacer uso de los íconos de la izquierda de la ventanilla, para seleccionar el gráfico (generalmente para copiarlo), o para agregar texto, líneas o marcar al gráfico.

Por otro lado, haciendo uso del menú *Graph*, se puede cambiar gran parte del aspecto del gráfico realizado mediante las opciones que despliega este menú.



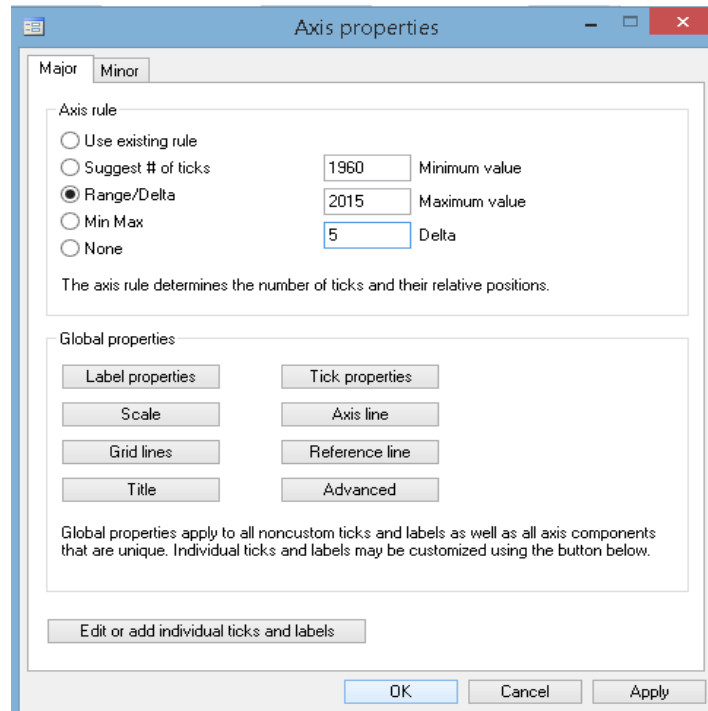
Por ejemplo, para agregarle un título al gráfico seleccionamos *Graph/ Titles* y aparece la siguiente ventanilla:



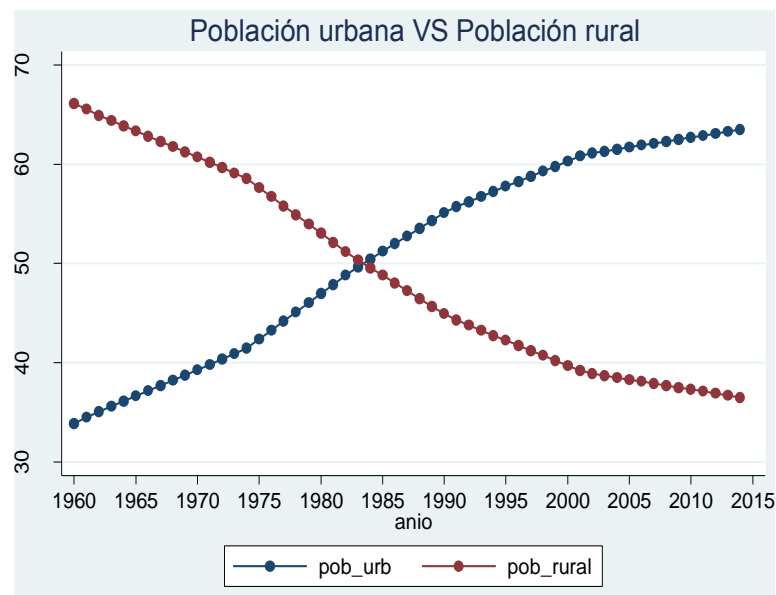
Una vez escrito el título del gráfico presionamos aceptar.

Ahora, se quiere cambiar el rango de la variable “años”. Para ello, se selecciona la variable “años”/ clic derecho/ *axis properties*. Y aparece la siguiente ventanilla:





En esta ventanilla, en *Range/Delta* agregamos los valores correspondientes y presionamos aceptar. Con ello, el gráfico modificado queda así:



Es importante señalar que no se pueden abarcar todas las posibilidades de modificación, ya que existen muchísimos aspectos que pueden ser modificados.

### 3.3 Análisis de regresión con dos variables: algunas ideas básicas

El análisis de regresión se relaciona en gran medida con la estimación o predicción de la media (de la población) o valor promedio de la variable dependiente,  $Y$ , con base en los valores conocidos o fijos de las variables explicativas,  $X$  (Gujarati & Porter, 2010).

El presente apartado introduce ideas básicas de lo que es estimar un modelo econométrico con dos variables (una dependiente  $Y$ , y una independiente  $X$ ), para ello se utiliza el modelo de consumo keynesiano en donde la variable dependiente  $Y$  es el consumo y la variable independiente  $X$  es el PIB, estos datos se encuentran en la Tabla 3.3.1. Adicionalmente se muestran dos tipos de funciones: poblacional y muestral a partir de datos hipotéticos de consumo e ingreso familiar (Tabla 3.3.2 y Tabla 3.3.3). Finalmente se utilizan datos del  $CO_2$  y PIB per cápita (Tabla 3.3.4) para explicar la linealidad de una regresión.

### 3.3.1 Estimación del modelo econométrico.

#### Ejercicio 1:

Un modelo econométrico es una representación en símbolos matemáticos de relaciones económicas, en donde una o más variables independientes ( $X$ s) explican el comportamiento de una variable dependiente ( $Y$ ). La forma básica de un modelo econométrico es la siguiente:

$$Y = \beta_1 + \beta_2 X + \mu \quad (3.3.1)$$

Donde,  $\beta_1$  y  $\beta_2$  son los parámetros o estimadores,  $X$  es la variable independiente,  $Y$  es la variable dependiente, y  $\mu$  es el término estocástico.

Comandos a utilizar:

**regress:** Permite estimar un modelo econométrico de regresión lineal. También se puede utilizar su forma abreviada **reg**

Su estructura es:  
**regress** varY varX

Rutina utilizada:  
**regress** consumo pib

Desarrollo:

Se van a estimar los parámetros de la función keynesiana de consumo para Ecuador, haciendo uso de los datos de la Tabla 3.3.1 (en millones de dólares). Para lo cual se ejecuta el siguiente comando: **regress** consumo pib. Y como resultado se tiene la siguiente regresión:

Source	SS	df	MS	Number of obs =	55
Model	7.2707e+09	1	7.2707e+09	F( 1, 53) =	18563.77
Residual	20758037.9	53	391661.093	Prob > F =	0.0000
				R-squared =	0.9972
				Adj R-squared =	0.9971
Total	7.2915e+09	54	135027096	Root MSE =	625.83

consumo	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
pib	.8029628	.0058934	136.25	0.000	.7911422 .8147833
_cons	433.8072	177.3339	2.45	0.018	78.12042 789.494

Sabiendo que,  $\beta_1$  es el coeficiente de la constante ( $\_cons$ ) y  $\beta_2$  es el coeficiente de la variable independiente (pib), se reemplazan los valores obtenidos en la forma básica de una regresión lineal (3.3.1). Así:

$$\widehat{consumo} = 433.81 + 0.80 (pib) + \mu \quad (3.3.2)$$

Es importante señalar que en la función de consumo keynesiano  $\beta_1$  es el consumo autónomo y  $\beta_2$  es la propensión marginal a consumir (entre 0 y 1).

Por último, a través del modelo de consumo obtenido (3.3.2) se pretende predecir el valor de la media del gasto de consumo para 2010 si el pib para 2010 fue de 49,035.71 millones de dólares. Por lo tanto, se reemplaza el valor del pib en el modelo y se tiene:

$$Consumo_t = 433.81 + 0.80 (pib_t) + \mu$$

$$Consumo_{2010} = 433.81 + 0.80 (pib_{2010}) + \mu$$

$$Consumo_{2010} = 433.81 + 39,228.57 + \mu$$

$$Consumo_{2010} = 39,662.38 \text{ millones de dólares}$$

El promedio del gasto de consumo previsto para el 2010 es de alrededor de 39,662.38 millones de dólares. El valor real del gasto de consumo registrado en 2010 fue de 41,132.00 millones de dólares. Por consiguiente, el error de predicción es de alrededor 14,69.62 millones de dólares, que representa aproximadamente el 3.7% del valor real del pib para 2010.

Interpretación:

Ante el aumento de 1 millón de dólares en el pib ecuatoriano, el consumo del país aumentará en 0.8 millones de dólares. Por lo tanto, la propensión marginal a consumir en Ecuador es alrededor de 0.8 lo que apoya a la teoría de Keynes.

### 3.3.2 Función de regresión poblacional.

#### Ejercicio 2:

Para estimar una función de regresión poblacional se necesitan los datos de toda una población objeto de estudio. Los resultados poblacionales son más costosos, de larga duración, y difíciles de aplicar que los estudios aplicados a una muestra o parte de la población. Sin embargo si se aplican eficientemente se obtendrán resultados más reales y verdaderos que los estudios sobre una muestra. El levantamiento de información de toda una población se la denomina comúnmente como “censo”.

Comandos a utilizar:

Rutina utilizada:  
**twoway (scatter consumof ingresosf) (lfit consumof ingresosf)**

Desarrollo:

Los datos de la Tabla 3.3.2 referentes a la población total de 60 familias de una comunidad hipotética, así como a su ingreso mensual ( $X$ ) y su gasto de consumo mensual ( $Y$ ), en dólares, se dividen en 10 grupos de ingresos (de 120 dólares a 300). Por consiguiente, hay 10 valores fijos de  $X$ , y los correspondientes valores  $Y$  para cada valor  $X$ . Por lo tanto hay 10 subpoblaciones  $Y$ .

La siguiente tabla proporciona la media o promedio mensual de consumo familiar correspondiente a cada uno de los 10 niveles de ingresos.

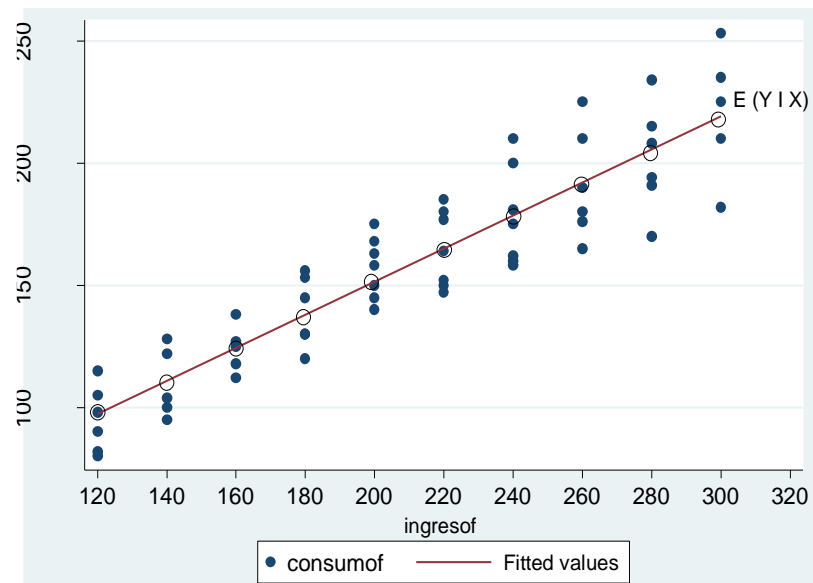
$Y \downarrow \quad X \rightarrow$	120	140	160	180	200	220	240	260	280	300
	80	95	112	120	140	147	158	165	170	182
	82	100	118	130	145	150	160	176	191	210
	90	104	125	130	150	152	162	180	194	225
	98	122	127	145	158	164	175	190	208	235
	105	128	138	153	163	177	181	210	215	253
	115	-	-	156	168	180	200	225	234	-
	-	-	-	-	175	185	210	-	-	-
<b>Total</b>	570	549	620	834	1099	1155	1246	1146	1212	1105
<b>Media condicional de <math>Y</math>, <math>E(Y X)</math></b>	95	109,8	124	139	157	165	178	191	202	221

A nivel de ingreso de 120 dólares le corresponde una media de consumo igual a 95 dólares, así como a nivel de ingreso de 300 dólares le corresponde una media de consumo igual a 221 dólares. Hay 10 valores medios para las 10 subpoblaciones. Estos valores también se los conoce como valores esperados condicionales.

Por otra parte, existe el valor esperado incondicional, el mismo que es el resultado de sumar todos los valores del consumo y dividirlo para el total de valores. En este caso el valor esperado incondicional sería:  $\$9536 / 60 = \$158.93$ .

La representación gráfica de los valores medios condicionales se la evidencia en una línea de regresión poblacional (LRP) o curva de regresión poblacional (CRP). Para ello, se ejecuta el siguiente comando: **twoway (scatter consumof ingresosf) (lfit consumof ingresosf)**

Luego de hacer ciertas modificaciones mediante el editor de gráficos (añadir los puntos medios, e introducir el texto “E (Y I X)”), se obtiene:



La media condicional de  $E(Y | X_i)$  es función de  $X_i$ . Simbólicamente:

$$E(Y | X_i) = f(X_i) \quad (3.3.3)$$

De este ejemplo, se puede suponer que la función de regresión poblacional (FRP) es una función lineal de  $X_i$ , de tipo

$$E(Y | X_i) = \beta_1 + \beta_2 (X_i) \quad (3.3.4)$$

(3.3.4) es una función de regresión poblacional lineal (FRPL), donde,  $\beta_1$  y  $\beta_2$  son los parámetros no conocidos pero fijos que se denominan coeficiente de intersección y de pendiente respectivamente.

Interpretación:

El valor esperado condicional de consumo para un nivel de ingreso de 120 dólares es de 95 dólares. El valor esperado condicional de consumo para un nivel de ingreso de 140 dólares es de 109.80 dólares. La misma interpretación es para las 10 subpoblaciones determinadas por los distintos niveles de ingreso.

### 3.3.3 Función de regresión muestral.

#### Ejercicio 3:

La muestra es una representación de la población. Por consiguiente, para estimar una función de regresión muestral (FRM) se necesitan los datos de una parte de la población objeto de estudio. Es decir, se utiliza una muestra representativa de la población, con el fin de estimar la función de regresión poblacional (FRP) con base en información muestral.

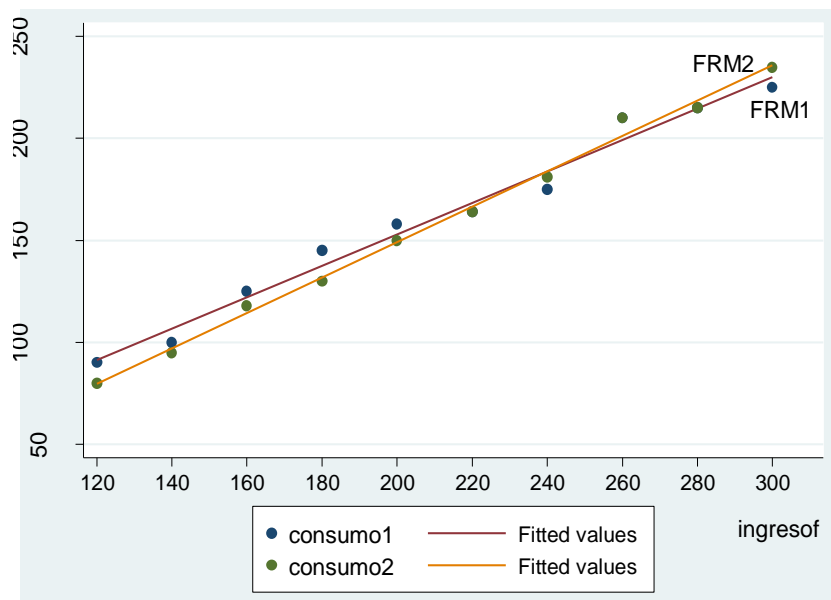
Comandos a utilizar:

Rutina utilizada:  
**twoway (scatter consumo1 ingreso1) (lfit consumo1 ingreso1) (scatter consumo2 ingreso2) (lfit consumo2 ingreso2)**

Desarrollo:

A manera de ejemplo se va a suponer que no se conocía la población exacta de la Tabla 3.3.2 y se toman dos muestras aleatorias (Tabla 3.3.3). Ahora, se grafica en un mismo plano las dos muestras para poder apreciar la desviación entre estas dos muestras. Para ello se ejecuta el siguiente comando: **twoway (scatter consumo1 ingreso1) (lfit consumo1 ingreso1) (scatter consumo2 ingreso2) (lfit consumo2 ingreso2)**

Y se obtiene:



Las líneas FRM1 Y FRM2 se conocen como líneas de regresión muestral. “Se supone que representan la línea de regresión poblacional, pero, debido a fluctuaciones muestrales, son, en el mejor de los casos, sólo una aproximación de la verdadera RP” (Gujarati & Porter, 2010, p.44).

Como señala Gujarati & Porter (2010) el objetivo principal del análisis de regresión es estimar la FRP

$$Y_i = \beta_1 + \beta_2 X_i + \mu_i \quad (3.3.5)$$

Con base en la FRM

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\mu}_i \quad (3.3.6)$$

Interpretación:

Comparando las dos FRM (1 y 2), se evidencia que la FRM1 es diferente a la FRM2, pero esta diferencia no es tan grande, pues las dos líneas FRM presentan un similar comportamiento. Tanto la FRM1 como la FRM2 son aproximaciones de la verdadera FRP.

### 3.3.4 Linealidad.

#### **Ejercicio 4:**

El término lineal se interpreta de dos formas: linealidad en las variables, y linealidad en los parámetros.

- Linealidad en las variables: Según Gujarati & Porter (2010) el significado de linealidad es aquel en que la esperanza condicional de  $Y$  es una función lineal de  $X_i$ . Geométricamente, la curva de regresión en este caso es una recta. En esta interpretación, una función de regresión como  $E(Y | X_i) = \beta_1 + \beta_2 X_i^2$  no es una función lineal porque la variable  $X$  aparece elevada a una potencia o índice 2.
- Linealidad en los parámetros: Según Gujarati & Porter (2010) la linealidad se presenta cuando la esperanza condicional de  $Y$ ,  $E(Y | X_i)$ , es una función lineal de los parámetros, los  $\beta$ , puede ser o no lineal en la variable  $X$ . Una función de regresión como  $E(Y | X_i) = \beta_1 + 3\beta_2^2$  es no lineal respecto del parámetro  $\beta_2$ .

Sin embargo, el término de regresión “lineal” siempre significará una regresión lineal en los parámetros, es decir los parámetros  $\beta$  se elevan sólo a la primera potencia, aunque sea no lineal en las variables  $X$ .

Comandos a utilizar:

- **predict:** Permite predecir los valores de la Y estimada luego de haber ejecutado una regresión.  
Su estructura es:  
**predict** Yestimada (predecir Y estimada)
- **connect (1):** Permite trazar una línea (de cualquier orden) que represente la relación entre una variable X y una Y.
- **exp:** Permite sacar el antilogaritmo o variable exponencial de los datos de una variable determinada.  
Su estructura es:  
**gen** nueva\_variable= **exp**(variable)

Rutina utilizada:

*/\*Función cuadrática\*/*

**generate** pibp2= pibp^2

**regress** co2 pibp pibp2

**predict** Y2

**scatter** Y2 pibp, **connect (1)**

*/\*Función cúbica\*/*

**generate** pibp2= pibp^2

(ya está definida en la función cuadrática)

**generate** pibp3= pibp^3

**regress** co2 pibp pibp2 pibp3

**predict** Y3

**scatter** Y3 pibp, **connect (1)**

*/\*Función exponencial\*/*

**regress** co2 pibp

**predict** Yest

**generate** Yexp= **exp**(Yest)

**scatter** Yexp pibp, **connect (1)**

Desarrollo:

En el presente ejercicio se graficará algunos tipos de funciones lineales en los parámetros, aunque no lo sean en las variables X. Utilizando los datos de CO<sub>2</sub> (Y) y PIB per cápita (X) de la Tabla 3.3.4, primeramente se grafica la función cuadrática:

$$Y = \beta_1 + \beta_2 X + \beta_3 X^2 \quad (3.3.7)$$

Para ello es necesario generar la variable pibp al cuadrado, estimar la regresión, predecir el valor de Y estimada (Y cuadrática), y finalmente graficar. Por lo tanto se ejecutan los siguientes comandos

**generate** pibp2= pibp^2

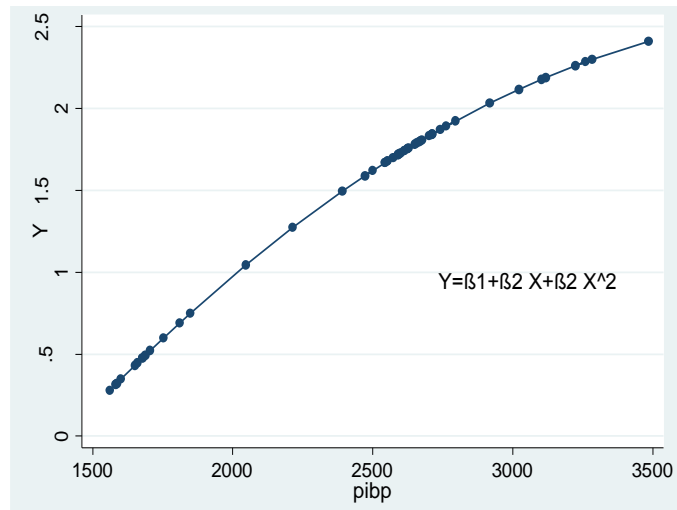
**regress** co2 pibp pibp2

**predict** Y2



**scatter Y2 pibp, connect (1)**

Y se obtiene el gráfico de la función cuadrática que se muestra a continuación:



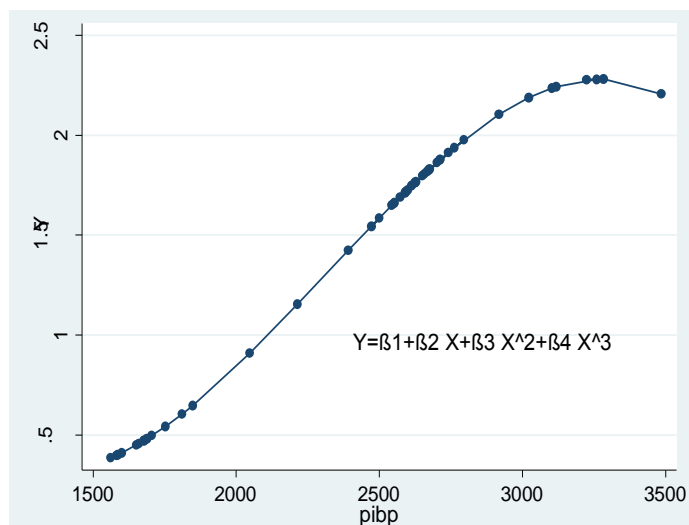
Segundo, para graficar la función cúbica:

$$Y = \beta_1 + \beta_2 X + \beta_3 X^2 + \beta_4 X^3 \quad (3.3.8)$$

Se genera la variable pibp al cubo, se estima la regresión, se predice el valor de Y estimada (Y cúbica), y finalmente se grafica. Por lo tanto se ejecutan los siguientes comandos:

```
generate pibp3= pibp^3  
regress co2 pibp pibp2 pibp3  
predict Y3  
scatter Y3 pibp, connect (1)
```

Y se obtiene gráfica de la función cúbica que se muestra a continuación:



Tercero, para graficar la función exponencial:

$$Y = e^{\beta_1 + \beta_2 X} \quad (3.3.9)$$

Se estima la regresión normal básica para predecir los valores de la  $Y$  estimada, luego se crea la variable exponencial de la  $Y$  estimada, se estima la regresión exponencial, y finalmente se grafica. Por lo tanto se ejecutan los siguientes comandos:

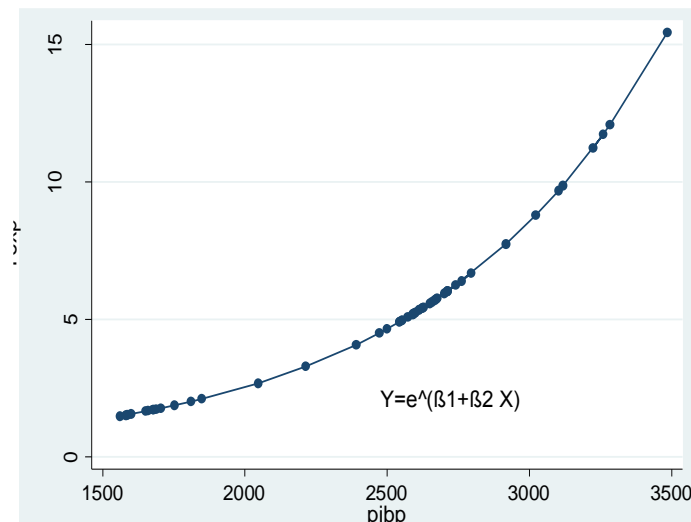
```
regress co2 pibp
```

```
predict Yest
```

```
generate Yexp= exp(Yest)
```

```
scatter Yexp pibp, connect (1)
```

Y se obtiene gráfica de la función exponencial que se muestra a continuación:



### 3.4 Análisis de regresión con dos variables: problema de estimación

Existen dos métodos para estimar una regresión, estos son: Mínimos Cuadrados Ordinarios (MCO) y Máxima Verosimilitud (MV). “El método de MCO es el más común en el análisis de regresión, sobre todo por ser mucho más intuitivo y matemáticamente más sencillo que el método de máxima verosimilitud” (Gujarati & Porter, 2010, p.55). Sin embargo, este método implica 10 supuestos que se deben cumplir.

Se inicia aplicando MCO a datos hipotéticos de ingreso y consumo contenidos en la Tabla 3.4.1. Luego a partir de información económica (PIB y las Exportaciones) contenida en la Tabla 3.4.2 se determina y analiza el coeficiente de determinación  $r^2$ , el coeficiente de correlación  $r$ , se aplican pruebas de significancia a los parámetros  $\beta$  y por último se aplican pruebas de normalidad de los residuos.

### 3.4.1 Método de Mínimos Cuadrados Ordinarios (MCO).

#### Ejercicio 1:

Un método muy eficaz y popular para el análisis de regresión es el Método de Mínimos Cuadrados (MCO), el cual permite determinar la FRM de manera que quede lo más cerca posible de la Y observada. El MCO supone que la sumatoria del cuadrado de los residuos sea lo más pequeña posible. Este método permite encontrar los estimadores o parámetros de un modelo de regresión lineal expresados en cantidades. Sin embargo, las propiedades estadísticas de los estimadores MCO se mantienen sólo con ciertos supuestos sobre la forma como se generaron los datos (Gujarati & Porter, 2010).

Según Gujarati & Porter (2010) los supuestos de un MCRL son los siguientes:

1. Ser un modelo de regresión lineal.
2. Valores fijos de X independientes del término de error.
3. El valor medio de la perturbación  $\mu_i$  es igual a cero.
4. Homoscedasticidad o varianza constante de  $\mu_i$ .
5. No hay autocorrelación entre las perturbaciones.
6. Número de observaciones n debe ser mayor que el número de parámetros por estimar.
7. Variabilidades en las regresoras (X).
8. Multicolinealidad.
9. Sesgo de especificación.
10. No normalidad de las perturbaciones (p. 85)

Comandos a utilizar:

- **predict:** Permite predecir los valores de los residuos de una regresión.  
Su estructura es:  
**predict r, resid** (predecir residuos)
- **total:** Realiza la sumatoria de los valores de determinada variable con su respectiva desviación estándar total.  
Su estructura es:  
**total variable**

Rutina utilizada:  
**regress Y X**  
**predict r, resid**  
**generate r2=r^2**  
**total r2**

Desarrollo:

Utilizando los datos de la Tabla 3.4.1 (datos hipotéticos), donde  $Y$  es el consumo y  $X$  el ingreso, se quiere obtener los estimadores MCO del modelo de regresión lineal. Mediante la ejecución del comando: **regress Y X**, se obtienen directamente los valores de los estimadores del modelo. Por consiguiente, Stata muestra la regresión estimada así:

Source	SS	df	MS	Number of obs = 60		
Model	100705.18	1	100705.18	F( 1, 58) = .		
Residual	55.8031382	58	.962123073	Prob > F = 0.0000		
				R-squared = 0.9994		
				Adj R-squared = 0.9994		
Total	100760.983	59	1707.81328	Root MSE = .98088		

Y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
X	.6418162	.0019838	323.53	0.000	.6378452	.6457873
_cons	22.5546	.4402947	51.23	0.000	21.67325	23.43594

Con ello ya se tienen los estimadores MCO que son:  $\beta_1 = 22.5546$  y  $\beta_2 = 0.6418$ . Y por lo tanto el modelo de regresión lineal queda representado de la siguiente manera:

$$\hat{Y}_i = 22.5546 + 0.6418X_i + u_i \quad (3.4.1)$$

Ahora, para averiguar si la sumatoria del cuadrado los residuos es pequeña, se predicen los residuos, se los eleva al cuadrado y se realiza la sumatoria de sus valores mediante los siguientes comandos:

**predict r, resid**

**generate r2=r^2**

**total r2**

Total estimation	Number of obs = 60		
	Total	Std. Err.	[95% Conf. Interval]
r2	55.80314	7.977965	39.83927 71.76701

Se tiene que la sumatoria del cuadrado de los residuos es igual a 55.80, siendo este un valor pequeño.

Interpretación:

Basándose en los resultados de (3.4.1), el estimador  $\beta_2 = 0.6418$  significa que, si el ingreso aumenta en 1 dólar, el consumo aumenta en 64 centavos. Mientras que el estimador  $\beta_1 = 22.5546$  significa que, si aunque el ingreso sea cero, el consumo es alrededor de 22.55 dólares.

### 3.4.2 Coeficiente de determinación $r^2$ .

#### Ejercicio 2:

El coeficiente de determinación  $r^2$  “mide la bondad de ajuste de la ecuación de regresión, es decir, da la proporción o porcentaje de la variación total en la variable dependiente  $Y$  explicada por la variable explicativa  $X$ ” (Gujarati & Porter, 2010, p.196). Este valor es siempre positivo, y va de 0 a 1 (en términos absolutos)<sup>9</sup>.

Por otro lado, se tiene el coeficiente de correlación  $r$ , que mide el grado de asociación lineal entre las dos variables. Además el coeficiente de correlación es la raíz cuadrada del coeficiente de determinación. Este valor puede ser negativo o positivo, y va de -1 a 1 (en términos absolutos).

Comandos a utilizar:

**correlate**: Proporciona una matriz con los valores del coeficiente de correlación  $r$ , entre un conjunto de variables.

Su estructura es:

**correlate** var1 var2... var\_n

Rutina utilizada:

**correlate** PIB Exp

**regress** PIB EXP

**scatter** PIB EXP

Desarrollo:

Utilizando los datos de la Tabla 3.4.2, se averiguará el coeficiente de correlación y el coeficiente de determinación entre las variables PIB y exportaciones. Primeramente se va a obtener el coeficiente de correlación  $r$ , con el siguiente comando: **correlate** PIB EXP. Y se obtiene la siguiente matriz de correlación:

	PIB	EXP
PIB	1.0000	
EXP	0.9810	1.0000

<sup>9</sup> En términos relativos el coeficiente de determinación  $r^2$  y el coeficiente de correlación  $r$  va de 0% a 100%

Segundo, el valor del coeficiente de determinación entre el PIB y EXP se puede obtener de dos maneras: por medio del cuadrado de  $r$ , o mediante la estimación de la regresión lineal ejecutando el comando: **regress** PIB EXP.

$$r^2 = (r)^2 = (0.9810)^2 = 0.9624$$

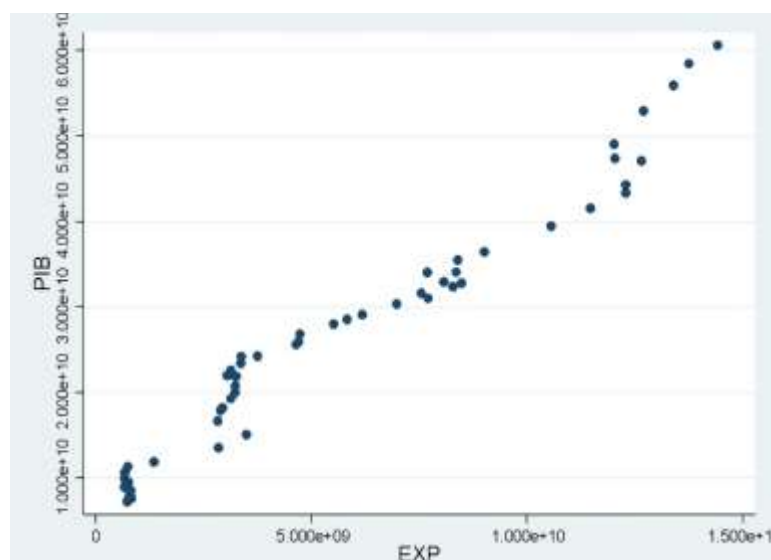
Source	SS	df	MS	Number of obs =	55
Model	1.0852e+22	1	1.0852e+22	F( 1, 53) =	1355.35
Residual	4.2438e+20	53	8.0071e+18	Prob > F	= 0.0000
				R-squared	= 0.9624
				Adj R-squared	= 0.9617
Total	1.1277e+22	54	2.0883e+20	Root MSE	= 2.8e+09

PIB	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
EXP	3.306538	.0898147	36.82	0.000	3.126392	3.486683
_cons	7.98e+09	6.31e+08	12.66	0.000	6.72e+09	9.25e+09

$$\hat{Y}_t = 7.98 * 10^9 + 3.31X_t + u_t \quad (3.4.2)$$

$$R^2 = 0.9624$$

En la regresión estimada (3.4.2) el valor del coeficiente de determinación  $r^2$  es igual a 0.9624. El siguiente gráfico muestra una fuerte relación lineal positiva entre el PIB y las exportaciones de Ecuador, con la ejecución del comando: **scatter** PIB EXP



Interpretación:

Las exportaciones explican al PIB en un 96%. Además el PIB y las exportaciones tienen un grado de relación lineal del 98%, es decir están fuertemente relacionadas. Cabe mencionar que un  $r^2$  alto no garantiza que un modelo este bien especificado, ya que debe cumplir con otros supuestos de un MCRL.

### 3.4.3 Prueba de significancia de los coeficientes de regresión.

#### Ejercicio 3:

La prueba de significancia trata de verificar la verdad o falsedad de una hipótesis nula utilizando los resultados muestrales obtenidos. “La idea básica de las pruebas de significancia es la de un estadístico de prueba (un estimador) y su distribución muestral según la hipótesis nula” (Gujarati & Porter, 2010, p.115)

*Toma de decisión de prueba de significancia*<sup>10</sup>: Considerando un intervalo de confianza del 95% con su respectivo nivel de significancia del 5%, para aceptar o rechazar la  $H_0$  (hipótesis nula) de que el estimador  $\beta$  es estadísticamente igual a 0, se debe tomar en cuenta el valor de su probabilidad, si ésta es mayor a 5% se acepta la  $H_0$ , de lo contrario se la rechaza llegando a concluir que el estimador no es estadísticamente igual a 0, aceptando así  $H_1$  (hipótesis alternativa). Adicionalmente también se considera el valor absoluto de  $t$ , si éste valor es menor a 2 se acepta  $H_0$ , en caso contrario se rechaza  $H_0$  y se acepta  $H_1$ .

$$H_0: \beta_n = 0$$

$$H_1: \beta_n \neq 0$$

Generalmente lo que se espera es rechazar  $H_0$ , donde el estimador no es estadísticamente igual a 0, ya que de esta manera el estimador será estadísticamente significativo.

Prueba de significancia de un estimador $\beta$ para un intervalo de confianza del 95%		
$p > 5\%$ ; $t < 2$	Acepta $H_0$	Es estadísticamente no significativo
$p < 5\%$ ; $t > 2$	Rechaza $H_0$	Es Estadísticamente significativo

Comandos a utilizar:

Rutina utilizada:  
**regress PIB EXP**

Desarrollo:

A manera de ejemplo se comprobará la significancia de los estimadores  $\beta$  del ejercicio 2 (Tabla 3.4.2). Para ello se estimada la regresión en donde  $Y$  es el PIB y  $X$  son las exportaciones.

<sup>10</sup> La prueba de significancia puede variar dependiendo del intervalo de confianza (90% al 99%), y del nivel de significancia (1% al 10%)

Source	SS	df	MS	Number of obs = 55		
Model	1.0852e+22	1	1.0852e+22	F( 1, 53)	=	1355.35
Residual	4.2438e+20	53	8.0071e+18	Prob > F	=	0.0000
				R-squared	=	0.9624
				Adj R-squared	=	0.9617
Total	1.1277e+22	54	2.0883e+20	Root MSE	=	2.8e+09

PIB	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Exp	3.306538	.0898147	36.82	0.000	3.126392	3.486683
_cons	7.98e+09	6.31e+08	12.66	0.000	6.72e+09	9.25e+09

$$\hat{Y}_t = 7.98 * 10^9 + 3.31X_t + u_t \quad (3.4.2) = (3.4.3)$$

$$R^2 = 0.9624$$

Con un intervalo de confianza del 95% y un nivel de significancia del 5%, se puede observar que el intercepto  $\beta_1 = 7.98e+09$  tiene un valor de  $t = 36.82$  y una  $p=0.000$ . Y la pendiente  $\beta_2=3.306538$  tiene un valor de  $t = 12.66$  y una  $p=0.00$ . Cabe mencionar que el nivel de significancia  $\alpha$  puede tomar valores entre 0% y 10%, sin embargo Stata considera un  $\alpha=5\%$  (que está predeterminado).

Interpretación:

En un intervalo de confianza del 95%, tanto el intercepto  $\beta_1$  como la pendiente  $\beta_2$  de la regresión PIB y EXP son estadísticamente significativos, ya que sus valores absolutos de  $t$  son mayores a 2 y sus probabilidades son menores a 5%. Por lo tanto, se rechaza  $H_0$  y se acepta  $H_1$ .

#### 3.4.4 Pruebas de normalidad.

##### Ejercicio 4:

Para saber si los residuos de una regresión están distribuidos normalmente se realizan pruebas de normalidad. Gujarati & Porter (2010) consideran tres tipos de pruebas de normalidad:

1) histograma de residuos: Si la curva de distribución normal en forma de campana es similar al comportamiento de los residuos (mostrado en el histograma) entonces los residuos estarán normalmente distribuidos.

2) gráfico de probabilidad normal (GPN): Si los datos están cerca de la línea del GPN entonces los residuos estarán normalmente distribuidos.



3) prueba Jarque-Bera: Si el valor de la asimetría es cercano a 0, el valor de la curtosis es cercano a 3 y se tiene una probabilidad razonablemente alta entonces los residuos estarán normalmente distribuidos. Los residuos tendrán una distribución normal perfecta cuando su asimetría es 0, su curtosis es 3, y por lo tanto su Jarque-Bera es 0.

El test de normalidad incluye el siguiente conjunto de hipótesis:

Ho: Los residuos están normalmente distribuidos

H1: Los residuos no están normalmente distribuidos

Comandos a utilizar:

- **histogram:** Permite graficar histogramas con y sin curva de distribución normal.  
Su estructura es:  
**histogram** residuos (histograma de residuos)  
**histogram** residuos, **normal** (histograma con curva de distribución normal)
- **pnorm:** Permite realizar gráficos de probabilidad normal.  
Su estructura es:  
**pnorm** residuos
- **summarize:** Muestra los principales datos estadísticos de una variable (#observaciones, media, desviación estándar, mínimo y máximo). También se puede utilizar su forma abreviada *sum*.  
Su estructura es:  
**summarize** variable
- **detail:** Es una opción del comando *summarize* que muestra los valores de la media, desviación estándar, varianza, asimetría y curtosis de una variable.  
Su estructura es:  
**summarize** variable, **detail**
- **sktest:** Permite realizar el test de normalidad. Muestra los valores de las probabilidades de la asimetría y curtosis.  
Su estructura es:  
**sktest** residuos

Rutina utilizada:

**generate** PIB EXP

**predict** r, resid

**histogram** r

(histograma)

**histogram** r, **normal**

(histograma con curva de distribución normal)

**pnorm** r

(gráfico de probabilidad normal)

**summarize** r **detail**

(valores de asimetría y curtosis)

**sktest** r

(probabilidades de asimetría y curtosis)

Desarrollo:

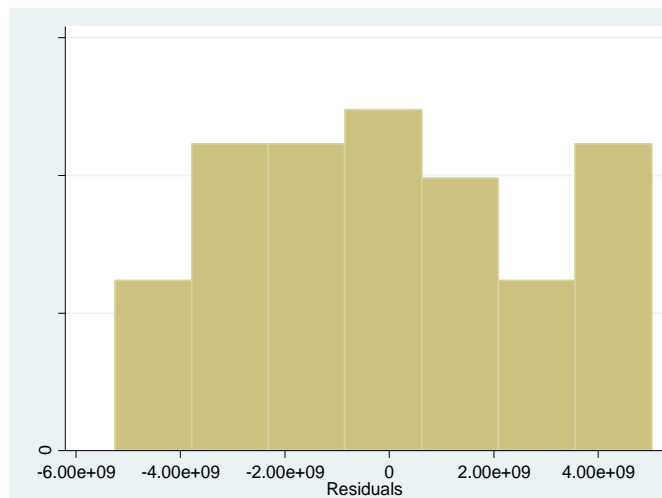
Utilizando nuevamente los datos de la Tabla 3.4.2, se va a averiguar si los residuos, de la regresión PIB sobre EXP, están normalmente distribuidos mediante la aplicación de los 3 tipos

de pruebas de normalidad. Antes de realizar cualquiera de las 3 pruebas de normalidad es necesario estimar la regresión y predecir los residuos. Para ello se ejecutan los siguientes comandos:

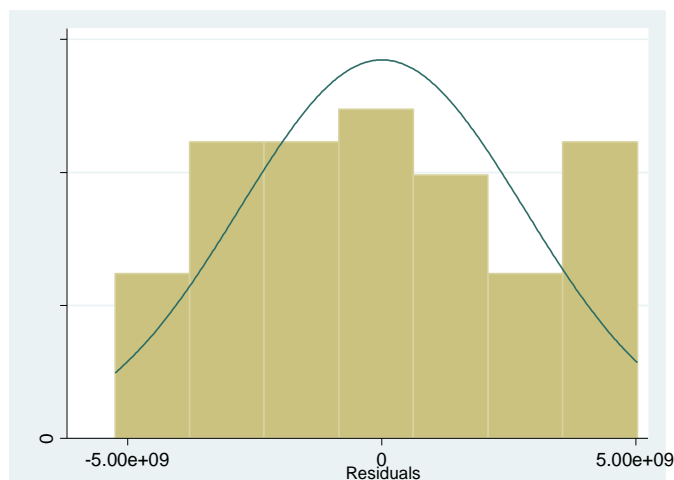
**generate** PIB EXP

**predict** r, resid

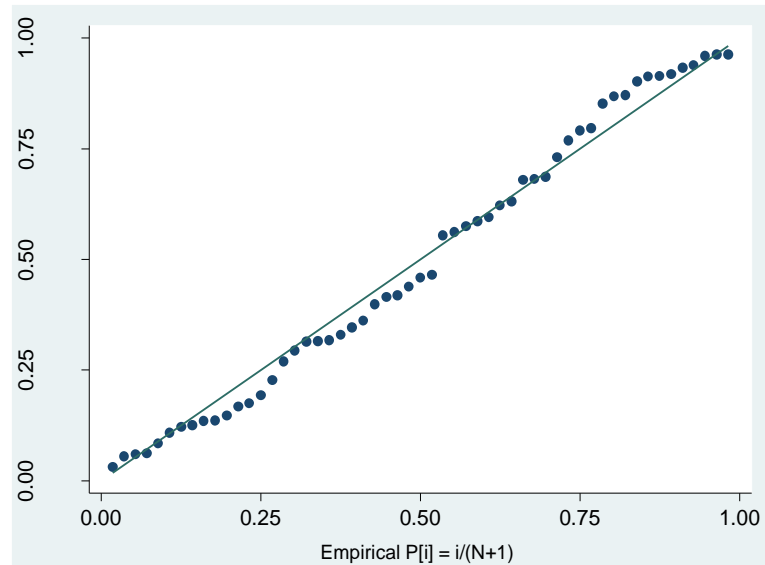
Para realizar la prueba de normalidad mediante el histograma de residuos, se ejecuta el siguiente comando: **histogram** r. Y se presenta el siguiente histograma:



Sin embargo, este histograma no presenta la curva de distribución normal, y para agregarla se añade al comando la palabra *normal*, así: **histogram** r, **normal**. Al ejecutar este comando se obtiene el siguiente histograma:



Para realizar la prueba de normalidad mediante el gráfico de probabilidad normal se ejecuta el siguiente comando: **pnorm** r. Y se obtiene el siguiente gráfico:



Para realizar la prueba de normalidad con los valores de asimetría y curtosis se ejecuta el siguiente comando: **summarize r, detail**. Y se obtiene lo siguiente:

Fitted values					
	Percentiles	Smallest			
1%	1.02e+10	1.02e+10			
5%	1.02e+10	1.02e+10			
10%	1.04e+10	1.02e+10	Obs		55
25%	1.73e+10	1.03e+10	Sum of Wgt.		55
50%	2.04e+10		Mean		2.65e+10
		Largest	Std. Dev.		1.42e+10
75%	3.58e+10	5.00e+10			
90%	4.86e+10	5.23e+10	Variance		2.01e+20
95%	5.23e+10	5.34e+10	Skewness		.5635494
99%	5.57e+10	5.57e+10	Kurtosis		2.036395

Con ello se tienen los valores de asimetría y curtosis de 0,56 y 2,03 respectivamente. Adicional a ello se pueden obtener los valores de las probabilidades de estos valores, y la probabilidad conjunta. Si la probabilidad de este test es mayor al 5% se acepta  $H_0$ : Los residuos están normalmente distribuidos, y si esta probabilidad es menor se rechaza  $H_0$ . Para ello se ejecuta el siguiente comando: **sktest r**, obteniendo los siguientes resultados:

Skewness/Kurtosis tests for Normality					
Variable	Obs	Pr (Skewness)	Pr (Kurtosis)	adj chi2 (2)	joint Prob>chi2
r	55	0.6803	0.0169	5.60	0.0607

Se tiene una probabilidad de 6.07%, la misma que es mayor a 5%.

Interpretación:

Al observar el histograma se evidencia (en primera instancia) que los residuos no están distribuidos normalmente, ya que curva de distribución normal en forma de campana sobre el histograma no es similar al comportamiento de los residuos. Sin embargo el gráfico de probabilidad normal muestra que los datos están muy cercanos a la línea del GPN, por lo tanto se podría concluir que los residuos están normalmente distribuidos (aceptando  $H_0$ ).

Por otro lado, los valores de la asimetría y curtosis, los cuales son iguales a 0.56 (valor cercano a 0) y 2.03 (valor medianamente cercano a 3) respectivamente, con una probabilidad conjunta del 6.07% muestran que los residuos están normalmente distribuidos (se acepta  $H_0$ ), debido a que la probabilidad del 6.07% es mayor a 5%. Por lo tanto se concluye que los residuos efectivamente tienen una distribución normal.

### **3.5 Extensiones del modelo de regresión lineal con dos variables**

Para abordar las extensiones del modelo de regresión normal con dos variables se inicia explicando y estimando una regresión a través de origen, luego se ejemplifican modelos con distintas unidades de medición, continuando con un modelo sobre variables estandarizadas y por último se consideran las formas funcionales del modelo de regresión lineal, como son: modelo log-lineal, modelos semilogarítmicos, modelo recíproco y modelo recíproco logarítmico. Para abarcar estos temas se desarrollan 8 ejercicios que hacen uso de los datos de la Tabla 3.5.1 a la Tabla 3.5.6. Algunos de los datos son a nivel cantonal o provincial del año 2010, y otros son a nivel nacional de un período determinado según sea el caso.

#### **3.5.1 Regresión a través del origen.**

##### **Ejercicio 1:**

Existen modelos en los cuales el intercepto está ausente, es decir  $\beta_1=0$ , a estos modelos se los conoce como “regresiones a través de origen”, y adquieren la siguiente forma:

$$Y_i = \beta_2 X_i + u_i \quad (3.5.1)$$

Es oportuno indicar que “el  $r^2$  convencional no es apropiado en regresiones que no consideren el intercepto. Pero se puede calcular para tales modelos, lo que se conoce como el  $r^2$  simple” (Gujarati & Porter, 2010, p.150). En ocasiones  $r^2$  simple puede ser negativo, lo cual no ocurre en el modelo convencional (con intercepto).

El  $r^2$  simple se define como:

$$r^2_{simple} = \frac{(\sum X_i Y_i)^2}{\sum X_i^2 \sum Y_i^2} \quad (3.5.2)$$

Comandos a utilizar:

Rutina utilizada:  
**regress** tc lprod  
**regress** tc lprod, **noconst** (se escribe **noconst** al final para indicar que la regresión a estimar va sin intercepto)

Desarrollo:

Se va a estimar el modelo de convergencia absoluta de la productividad para el caso del Ecuador, utilizando los datos de la Tabla 3.5.1. Al estimar el modelo convencional (con intercepto) ejecutando el comando: **regress** tc lprod. Se tienen los siguientes resultados:

Source	SS	df	MS	Number of obs = 410		
Model	.028022562	1	.028022562	F( 1, 408) = 5.50		
Residual	2.0782928	408	.005093855	Prob > F = 0.0195		
Total	2.10631536	409	.005149915	R-squared = 0.0133		
				Adj R-squared = 0.0109		
				Root MSE = .07137		

tc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lprod	.0461757	.0196871	2.35	0.019	.0074748	.0848765
_cons	-.005013	.0163688	-0.31	0.760	-.0371908	.0271647

$$\hat{tc}_i = -0.005 + 0.0462lprod_i + u_i \quad (3.5.3)$$

$$t = (-0.31) \quad (2.35)$$

Los resultados la regresión (3.5.3) muestran que el intercepto  $\beta_1 = -0.005$  no es estadísticamente diferente de cero, es decir no es significativo aunque el coeficiente de la pendiente  $\beta_2 = 0.0462$  es muy significativo. Esto indica que puede ser que los datos se ajusten de mejor manera en una regresión a través del origen. Para averiguarlo se estima la regresión a través del origen mediante la ejecución del siguiente comando: **regress** tc lprod, **noconst**

Y se obtienen los siguientes resultados:

Source	SS	df	MS	Number of obs = 410		
Model	.460044188	1	.460044188	F( 1, 409) =	90.51	
Residual	2.07877057	409	.005082569	Prob > F =	0.0000	
				R-squared =	0.1812	
				Adj R-squared =	0.1792	
Total	2.53881475	410	.006192231	Root MSE =	.07129	

tc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lprod	.0402878	.0042346	9.51	0.000	.0319634	.0486122

$$\hat{tc}_i = 0.04029lprod_i + u_i \quad (3.5.4)$$

Los resultados de la regresión a través del origen (3.5.4) muestran que el coeficiente de la pendiente  $\beta_2 = 0.00376$  es muy significativo, ya que el valor de su probabilidad es muy pequeña. Por lo tanto, se evidencian que los datos se ajustan mejor a una regresión a través del origen.

Interpretación:

El modelo de convergencia absoluta de la productividad del Ecuador se ajusta a una regresión a través del origen (el intercepto se ausenta), ya que el intercepto es estadísticamente no significativo. Según los resultados de este modelo (3.5.4) si el logaritmo de la productividad aumenta en un punto porcentual, la tasa de crecimiento aumentará alrededor de 0.04 puntos porcentuales.

### 3.5.2 Escalas y unidades de medición.

#### Ejercicio 2:

A menudo se presenta la situación en que unos investigadores utilizan información medida en miles de millones de dólares y otros expresan estos datos en millones de dólares. O también unos expresan los datos en miles de dólares y otros en dólares. En fin, la escala de medición es distinta. En estos casos es pertinente utilizar factores de escala  $w$ . Así:

$$Y_i^* = w_1 Y_i \quad (3.5.5)$$

$$X_i^* = w_2 X_i \quad (3.5.6)$$

Donde  $Y_i^*$  y  $X_i^*$  son  $Y_i$  y  $X_i$  reescaladas. Por lo tanto, si  $Y_i$  y  $X_i$  se miden en miles de millones de dólares y se quiere expresarlas en millones de dólares, se tendrá  $Y_i^* = 1000Y_i$  y  $X_i^* = 1000X_i$ . En este caso  $w_1 = w_2 = 1000$  (Gujarati & Porter, 2010).

Comandos a utilizar:

Rutina utilizada:  
**reg** FBKFmm PIBmm  
**reg** FBKFm PIBm  
**reg** FBKFmm PIBm  
**reg** FBKFm PIBmm

Desarrollo:

Se consideran los datos de la Tabla 3.5.2, para examinar los resultados de las regresiones entre la FBKF y PIB en distintas escalas de medición.

Si las escalas de la FBKF y del PIB están en miles de millones de dólares, se ejecuta el siguiente comando: **reg** FBKFmm PIBmm. Y se obtiene:

Source	SS	df	MS	Number of obs = 50		
Model	419.365932	1	419.365932	F( 1, 48) = 233.25		
Residual	86.302182	48	1.79796212	Prob > F = 0.0000		
				R-squared = 0.8293		
				Adj R-squared = 0.8258		
Total	505.668114	49	10.3197574	Root MSE = 1.3409		

FBKFmm	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
PIBmm	.2118544	.0138717	15.27	0.000	.1839633	.2397454
_cons	.4827288	.4364628	1.11	0.274	-.3948385	1.360296

$$\widehat{FBKFmm}_t = 0.4827 + 0.2118PIBmm_t + u_t \quad (3.5.7)$$

$$ee = (0.43646) \quad (0.01387) \quad R^2 = 0.8293$$

Si las escalas de la FBKF y del PIB están en millones de dólares, se ejecuta el siguiente comando: **reg** FBKFm PIBm. Y se obtiene:

Source	SS	df	MS	Number of obs = 50		
Model	419365932	1	419365932	F( 1, 48) = 233.25		
Residual	86302182	48	1797962.12	Prob > F = 0.0000		
				R-squared = 0.8293		
				Adj R-squared = 0.8258		
Total	505668114	49	10319757.4	Root MSE = 1340.9		

FBKFm	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
PIBm	.2118544	.0138717	15.27	0.000	.1839633	.2397454
_cons	482.7288	436.4628	1.11	0.274	-394.8385	1360.296

$$\widehat{FBKF}m_t = 482.73 + 0.2118PIBm_t + u_t \quad (3.5.8)$$

$$ee = (436.46) \quad (0.0139) \quad R^2 = 0.8293$$

Se observa que el intercepto y su error estándar de esta regresión es 1000 veces los valores correspondientes a la primera regresión ( $w_1 = 1000$  al pasar de miles de millones a millones de dólares), pero el coeficiente de la pendiente y su error estándar son los mismos.

Si FBKF está en miles de millones de dólares y el PIB está en millones de dólares, se ejecuta el siguiente comando: **reg** FBKFm PIBm. Y se obtiene:

Source	SS	df	MS	Number of obs = 50		
Model	419.365932	1	419.365932	F( 1, 48) =	233.25	
Residual	86.302182	48	1.79796212	Prob > F =	0.0000	
				R-squared =	0.8293	
				Adj R-squared =	0.8258	
Total	505.668114	49	10.3197574	Root MSE =	1.3409	

FBKFm	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
PIBm	.0002119	.0000139	15.27	0.000	.000184	.0002397
_cons	.4827288	.4364628	1.11	0.274	-.3948385	1.360296

$$\widehat{FBKF}mm_t = 0.48273 + 0.002118PIBm_t + u_t \quad (3.5.9)$$

$$ee = (0.43646) \quad (0.000139) \quad R^2 = 0.8293$$

El coeficiente de la pendiente y su error estándar es (1/1000) de su valor de la primera regresión, pues solamente se modificó la escala de  $X$  (PIB).

Si FBKF está en millones de dólares y el PIB está en miles de millones de dólares, se ejecuta el siguiente comando: **reg** FBKFm PIBmm. Y se obtiene:

Source	SS	df	MS	Number of obs = 50		
Model	419365932	1	419365932	F( 1, 48) =	233.25	
Residual	86302182	48	1797962.12	Prob > F =	0.0000	
				R-squared =	0.8293	
				Adj R-squared =	0.8258	
Total	505668114	49	10319757.4	Root MSE =	1340.9	

FBKFm	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
PIBmm	211.8544	13.87174	15.27	0.000	183.9633	239.7454
_cons	482.7288	436.4628	1.11	0.274	-394.8385	1360.296



$$\widehat{FBKFmm}_t = 0.48273 + 0.002118PIBm_t + u_t \quad (3.5.10)$$

$$ee = (436.46) \quad (13.8717) \quad R^2 = 0.8293$$

Tanto el intercepto como el coeficiente de la pendiente y sus respectivos errores estándar son 1000 veces los valores de la primera regresión.

Un punto importante a considerar es que  $R^2$  es invariante a las escalas de medición, pues en todas las regresiones  $R^2$  es igual a 0.8293.

Interpretación:

*Primera regresión:* Si el PIB cambia 1000 millones de dólares, la FBKF cambiará en promedio 0.21185 miles de millones de dólares.

*Segunda regresión:* Si el PIB cambia un millón de dólares, la FBKF cambiará en promedio 0.21185 millones de dólares.

*Tercera regresión:* Si el PIB cambia un millón de dólares, la FBKF cambiará en promedio 0.00021185 miles de millones de dólares.

*Cuarta regresión:* Si el PIB cambia 1000 millones de dólares, la FBKF cambiará en promedio 211.85 millones de dólares.

### 3.5.3 Regresión sobre variables estandarizadas.

#### Ejercicio 3:

A saber, las escalas de medición de la variable  $Y$  como de la variable  $X$  influyen en la interpretación de los coeficientes de regresión. Esto se evita utilizando variables estandarizadas para  $Y$  y  $X$  (Gujarati & Porter, 2010).

Las variables estandarizadas se definen de la siguiente manera:

$$Y_i^* = \frac{Y_i - \bar{Y}}{S_Y} \quad (3.5.11)$$

$$X_i^* = \frac{X_i - \bar{X}}{S_X} \quad (3.5.12)$$

Donde,  $\bar{Y}$  y  $\bar{X}$  son las medias muestrales de  $Y$  y  $X$ ,  $S_Y$  y  $S_X$  son las desviaciones estándar de  $Y$  y  $X$ , y  $Y_i^*$  y  $X_i^*$  son las variables estandarizadas. “Una propiedad interesante de una variable estandarizada es que el valor de su media siempre es cero y que su desviación estándar siempre es 1” (Gujarati & Porter, 2010, p.157).

Además, en una regresión que involucra variables estandarizadas para  $Y$  y  $X$  el término de intercepto será 0, siendo ésta una regresión a través del origen. En tal caso, la unidad de medición de las variables estandarizadas es en desviaciones estándar.

La regresión sobre variables estandarizadas se expresa de la siguiente manera:

$$Y_i^* = \beta_2^* X_i^* + \mu_i^* \quad (3.5.13)$$

Comandos a utilizar:

**egen / std:** Permiten generar variables estandarizadas.

Su estructura es:

**egen** variablestan= **std**(variable)

Rutina utilizada:

/\*Generar variables estadarizadas paso a paso\*/

**reg** FBKFmm PIBmm

**sum** FBKFmm

**gen** FBKFstan=(FBKFmm-6.486546)/3.212438

**sum** PIBmm

**gen** PIBstan= (PIBmm-28.33936)/13.80897

/\*Generar variables estadarizadas de manera directa\*/

**egen** FBFFstan= **std**(FBKFmm)

**egen** PIBstan= **std**(PIBmm)

/\*Estimar regresión sobre variables estandarizadas\*/

**reg** FBKFstan PIBstan, **noconst**

Desarrollo:

Para mostrar la diferencia entre una regresión estándar y una regresión sobre variables estandarizadas se continuará con los datos del ejercicio anterior (Tabla 5.3.2). Primero se estima la regresión estándar (normal) con el siguiente comando: **reg** FBKFmm PIBmm. Y se obtiene:

Source	SS	df	MS	Number of obs = 50		
Model	419.365932	1	419.365932	F( 1, 48) = 233.25		
Residual	86.302182	48	1.79796212	Prob > F = 0.0000		
				R-squared = 0.8293		
				Adj R-squared = 0.8258		
Total	505.668114	49	10.3197574	Root MSE = 1.3409		

FBKFmm	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
PIBmm	.2118544	.0138717	15.27	0.000	.1839633	.2397454
_cons	.4827288	.4364628	1.11	0.274	-.3948385	1.360296

$$\widehat{FBKFmm}_t = 0.4827 + 0.2118PIBmm_t + u_t \quad (3.5.7) = (3.5.14)$$

$$R^2 = 0.8293$$

Ahora, se va a estimar la regresión sobre las variables estandarizadas. Como se sabe que una variable estandarizada es el resultado de restar el valor de su media de sus valores individuales y esto dividido para su desviación estándar, se procede a determinar los datos que no se tienen, como el valor de las medias y las desviaciones estándar. Para ello se hace uso del comando **sum**, para luego proceder a generar las variables estandarizadas. Así:

```
sum FBKFmm
gen FBKFstan=(FBKFmm-6.486546)/3.212438
sum PIBmm
gen PIBstan= (PIBmm-28.33936)/13.80897
```

Como el comando **sum** arrojó los valores de la media y desviación estándar de cada variable, se pudieron generar las variables estandarizadas correspondientes.

También se pueden generar directamente variables estandarizadas, así:

```
egen FBKFstan= std(FBKFmm)
egen PIBstan= std(PIBmm)
```

Finalmente, se estima la regresión sobre las variables estandarizadas generadas, con el siguiente comando: **reg FBKFstan PIBstan, noconst**. Obteniendo los siguientes resultados:

Source	SS	df	MS	Number of obs = 50		
Model	40.6371874	1	40.6371874	F( 1, 49) = 238.10		
Residual	8.362811	49	.170669612	Prob > F = 0.0000		
Total	48.9999984	50	.979999968	R-squared = 0.8293		
				Adj R-squared = 0.8258		
				Root MSE = .41312		

FBKFstan	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
PIBstan	.9106761	.0590174	15.43	0.000	.7920762	1.029276

$$\widehat{FBKFmm}_i^* = 0.9107 PIBmm_i^* + \mu_i^* \quad (3.5.15)$$

$$R^2 = 0.8293$$

Interpretación:

De acuerdo a los resultados de la regresión (3.5.14) si el PIB se incrementa en mil millones de dólares, la FBKF aumentará, en promedio, 0.21 miles de millones de dólares. En cambio de acuerdo a los resultados de la regresión sobre variables estandarizadas (3.5.15) si el PIB

(estandarizado) se incrementara una desviación estándar, en promedio, la FBKF (estandarizada) aumentaría casi 0.91 desviaciones estándar.

#### 3.5.4 Modelo log-lineal.

##### Ejercicio 4:

Los modelos log-lineal o también conocidos como modelos doble log son modelos de regresión muy comunes útiles para estandarizar variables. La diferencia con los modelos estándar está en que tanto la variable dependiente  $Y$  como la o las variables independientes  $X$  están expresadas en logaritmos, es decir toman la siguiente forma:

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + \mu_i \quad (3.5.16)$$

$$\ln Y_i = \alpha + \beta_2 \ln X_i + \mu_i \quad (3.5.17)$$

Donde  $\ln$ =logaritmo natural,  $\alpha = \ln \beta_1$ . Este modelo es lineal en los parámetros  $\alpha$  y  $\beta_2$  y lineal en los logaritmos de las variables  $Y$  y  $X$ , y se estima por regresión MCO. Una característica relevante de este tipo de modelos es que el coeficiente de la pendiente  $\beta_2$  mide la elasticidad de  $Y$  respecto de  $X$ , es decir el cambio porcentual de  $Y$  ante un cambio porcentual de  $X$  (Gujarati & Porter, 2010).

Comandos a utilizar:

**log:** Permite extraer el logaritmo natural de una variable determinada.

Ejemplo:

**gen** nueva\_variable=**log**(variable)

Rutina utilizada:

**gen** lcons\_hog=**log**(cons\_hog)

**gen** lcons\_tot=**log**(cons\_tot)

**reg** lcons\_hog lcons\_tot

Desarrollo:

Se va a estimar un modelo log-lineal utilizando los datos de la Tabla 3.5.3. Para ello, antes de estimar el modelo es necesario generar las variables logarítmicas de las variables originales de  $Y$  (gasto de consumo de los hogares) y  $X$  (gasto de consumo total). Por lo tanto se ejecutan los siguientes comandos:

**gen** lcons\_hog=**log**(cons\_hog)

**gen** lcons\_tot=**log**(cons\_tot)

Una vez generadas las variables en logaritmo se procede a estimar el modelo doble log ejecutando el comando: **reg lcons\_hog lcons\_tot**. Y se obtienen los siguientes resultados:

Source	SS	df	MS	Number of obs = 55		
Model	20.1785882	1	20.1785882	F( 1, 53) =24798.54		
Residual	.04312614	53	.000813701	Prob > F = 0.0000		
				R-squared = 0.9979		
				Adj R-squared = 0.9978		
Total	20.2217143	54	.374476192	Root MSE = .02853		

lcons_hog	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lcons_tot	1.015079	.0064459	157.48	0.000	1.00215	1.028008
_cons	-.5439009	.1524221	-3.57	0.001	-.849621	-.2381809

$$\widehat{lcons\_hog}_t = -0.5439 + 1.0151lcons\_tot_t + \mu_t \quad (3.5.18)$$

$$ee = (0.1524) \quad (0.0006)$$

$$t = (-3.57) \quad (157.48) \quad R^2 = 0.9979$$

Los resultados muestran una elasticidad  $\beta_2$  del gasto de consumo de hogares respecto del gasto de consumo total igual a 1.015.

Interpretación:

Si el gasto de consumo total del Ecuador aumenta en 1%, en promedio, el gasto de consumo de los hogares aumentará alrededor del 1.015%. En consecuencia, los gastos de consumo de los hogares son muy sensibles a los cambios en el gasto de consumo total.

### 3.5.5 Modelos semilogarítmicos: log- lin y lin-log.

#### Ejercicio 5:

- **Modelo log-lin:** Estos modelos generalmente se utilizan para medir la tasa de crecimiento de ciertas variables, en la mayoría de los casos variables económicas, y se expresa de la siguiente manera:

$$\ln Y_t = \beta_1 + \beta_2 t + \mu_t \quad (3.5.19)$$

Este modelo es lineal, ya que  $\beta_1$  y  $\beta_2$  son lineales. Se diferencia de un modelo de regresión lineal estándar en el sentido de que “la variable dependiente es el logaritmo de Y y la variable independiente es el “tiempo”, que adquiere valores de 1, 2, 3, etc” (Gujarati & Porter, 2010, p.162).

Comandos a utilizar:

Rutina utilizada:  
**gen** lpob=**log**(pob)  
**reg** lpob tiempo

Desarrollo:

Considerando los datos de población de 1960 al 2014 de la Tabla 3.5.4, se va a estimar una regresión log-lin de la población con respecto al tiempo. Primeramente se genera la variable del logaritmo de la población con la ejecución del comando: **gen** lpob=**log**(pob). Luego se estima la regresión log-lin con el siguiente comando: **reg** lpob tiempo.

Y se obtienen los siguientes resultados de la regresión log-lin:

Source	SS	df	MS	Number of obs = 55		
Model	7.62308859	1	7.62308859	F( 1, 53) = 5533.98		
Residual	.073007768	53	.001377505	Prob > F = 0.0000		
				R-squared = 0.9905		
				Adj R-squared = 0.9903		
Total	7.69609636	54	.142520303	Root MSE = .03711		

lpob	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
tiempo	.0234522	.0003153	74.39	0.000	.0228199	.0240845
_cons	15.37088	.0101472	1514.80	0.000	15.35053	15.39124

$$\widehat{lpob}_t = 15.37 + 0.0234t + \mu_t \quad (3.5.20)$$

$$ee = (0.0101) \quad (0.0003)$$

$$t = (1514.8) \quad (74.39) \quad R^2 = 0.9905$$

El valor del intercepto 15.37088 es igual al logaritmo de la población al inicio del período de análisis, es decir que  $15.37088 = \log(4545548)$ . Además,  $\beta_2 = 0.0234522$  es la tasa instantánea (en un momento dado) y no la tasa compuesta (durante un período). La tasa compuesta se la calcula de la siguiente manera:

$$tasa\ compuesta = (\text{antilogaritmo}\beta_2 - 1) \quad (3.5.21)$$

Por lo tanto, la tasa compuesta es igual al  $\text{antilog}(0.0234522) - 1 = 0.023729$ . La tasa compuesta de crecimiento de la población 2,37% es un poco mayor que la tasa de crecimiento instantánea 2,34%.

Interpretación:

Durante el período de 1960 a 2014, la población se incrementó con una tasa del 2.3% aproximadamente.

### **Ejercicio 6:**

- **Modelo lin-log:** Se utiliza este tipo de modelo cuando se quiere averiguar el cambio absoluto en la variable dependiente  $Y$  debido a un cambio porcentual en la variable independiente  $X$  (Gujarati & Porter, 2010). Se escribe como:

$$Y_i = \beta_1 + \beta_2 \ln X_i + \mu_i \quad (3.5.22)$$

Comandos a utilizar:

Rutina utilizada:  
**gen** lcons\_tot=log(cons\_tot)  
**reg** cons\_hog lcons\_tot

Desarrollo:

Se va a estimar un modelo lin-log con los datos de la Tabla 3.5.3. En primer lugar se genera la variable del logaritmo de la variable  $X$  (gasto de consumo total) ejecutando el comando:

**gen** lcons\_tot=log(cons\_tot).

Luego se procede a estimar el modelo lin-log con el comando: **reg** cons\_hog lcons\_tot.

Y se obtienen los siguientes resultados:

Source	SS	df	MS	Number of obs = 55		
Model	4.8071e+21	1	4.8071e+21	F( 1, 53) = 556.48		
Residual	4.5783e+20	53	8.6384e+18	Prob > F = 0.0000		
Total	5.2649e+21	54	9.7498e+19	R-squared = 0.9130		
				Adj R-squared = 0.9114		
				Root MSE = 2.9e+09		

cons_hog	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lcons_tot	1.57e+10	6.64e+08	23.59	0.000	1.43e+10	1.70e+10
_cons	-3.52e+11	1.57e+10	-22.43	0.000	-3.84e+11	-3.21e+11

$$\widehat{cons\_hog}_t = -3.52 \times 10^{11} + 1.57 \times 10^{10} lcons\_tot_t + \mu_t \quad (3.5.22)$$

$$ee = (1.57 \times 10^{10}) \quad (6.64 \times 10^8)$$

$$t = (-22.43) \quad (23.59) \quad R^2 = 0.9130$$

El cambio absoluto en  $Y$  es  $(0.01) * \beta_2$ . En la regresión (3.5.22) el cambio absoluto en el gasto de consumo de los hogares es  $(0.01) * 15\,700\,000\,000 = 157\,000\,000$  dólares. En conclusión el cambio absoluto de  $Y$  es igual a 157 millones de dólares.

Interpretación:

Un incremento del 1% en el gasto de consumo total, en promedio, producirá un aumento alrededor de 157 millones de dólares en el gasto de consumo de los hogares.

### 3.5.6 Modelo recíproco.

#### Ejercicio 7:

Los modelos recíprocos toman la siguiente forma:

$$Y_i = \beta_1 + \beta_2 \left( \frac{1}{X_i} \right) + \mu_i \quad (3.5.23)$$

El modelo (3.5.23) no es lineal en la variable  $X$ , pero si lo es en los parámetros  $\beta_1$  y  $\beta_2$ , por lo tanto se lo considera un modelo de regresión lineal. “A medida que  $\beta_2 \left( \frac{1}{X_i} \right)$  se acerca a cero,  $Y$  se aproxima al valor límite o asintótico  $\beta_1$ . Por consiguiente, estos modelos contienen un valor asintótico límite que tomará la variable dependiente  $Y$  cuando el valor de la variable  $X$  aumente indefinidamente” (Gujarati & Porter, 2010, p.166). En otras palabras,  $\beta_1$  es el límite máximo que puede cambiar  $Y$  frente a un cambio de  $X$ . Si  $X$  crece indefinidamente  $Y$  no va a bajar menos de  $\beta_1$ . Además es importante señalar que un valor positivo de  $\beta_2$  implicará que la tasa de cambio de la variable  $Y$  respecto a la variable  $X$  es negativa, y viceversa.

Comandos a utilizar:

Rutina utilizada:  
**gen** recPIBp=1/ PIBp  
**reg** mort recPIBp

Desarrollo:

Se va a estimar un modelo recíproco entre la mortalidad (muertes por cada 1000 personas) y el pib per cápita (Tabla 3.5.5). Para ello es necesario generar una variable en forma recíproca de la variable  $X$  (PIB per cápita), por lo tanto se ejecuta el comando: **gen** recPIBp=1/ PIBp.

Ahora si se puede estimar el modelo recíproco ejecutando el comando: **reg** mort recPIBp

Y se obtienen los siguientes resultados:



Source	SS	df	MS	Number of obs =	30
Model	5.28801793	1	5.28801793	F( 1, 28) =	15.22
Residual	9.72846687	28	.347445245	Prob > F =	0.0005
				R-squared =	0.3521
				Adj R-squared =	0.3290
Total	15.0164848	29	.517809821	Root MSE =	.58944

mort	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
recPIBp	11584.62	2969.467	3.90	0.001	5501.942	17667.3
_cons	1.582743	1.054927	1.50	0.145	-.5781761	3.743662

$$\widehat{mort}_t = 1.5827 + 11584.62 \left( \frac{1}{PIBp_t} \right) + \mu_t \quad (3.5.24)$$

$$ee = (1.0549) \quad (2969.47)$$

$$t = (1.50) \quad (3.90) \quad R^2 = 0.3521$$

En la regresión estimada (3.5.24) el valor del intercepto  $\beta_1 = 1.582743$  es el valor asintótico, y el valor positivo de  $\beta_2 = 11584.62$  implica una relación negativa.

Interpretación:

Si el PIB per cápita aumenta indefinidamente, la mortalidad se acerca a su valor asintótico de casi 2 muertes por cada 1000 personas. Además existe una relación negativa entre la tasa de mortalidad y el PIB per cápita, si el PIB per cápita aumenta la tasa de mortalidad disminuirá.

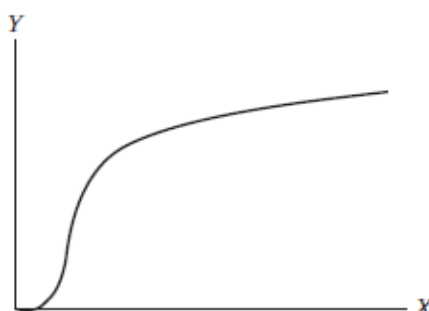
### 3.5.7 Modelo recíproco logarítmico.

#### Ejercicio 8:

El modelo recíproco logarítmico o log hipérbola adopta la siguiente forma:

$$\ln Y_i = \beta_1 + \beta_2 \left( \frac{1}{X_i} \right) + \mu_i \quad (3.5.25)$$

En este tipo de modelos “Y se incrementa con una tasa creciente (es decir, la curva es convexa al inicio) y luego aumenta con una tasa decreciente (la curva se convierte en cóncava” (Gujarati & Porter, 2010, p.172). Así:



Comandos a utilizar:

Rutina utilizada:  
**gen** lprod=log(prod)  
**gen** rec\_empleo= 1/empleo  
**reg** lprod rec\_empleo

Desarrollo:

Se va a estimar un modelo recíproco logarítmico utilizando los datos de la Tabla 3.5.6. Primeramente es necesario generar la variable del logaritmo de la producción (Y) y de crear la variable recíproca del empleo (X), para ello se ejecutan los siguientes comandos:

**gen** lprod=log(prod)  
**gen** rec\_empleo= 1/empleo

Ahora, se procede a estimar el modelo recíproco logarítmico ejecutando el comando:

**reg** lprod rec\_empleo

Y se obtienen los siguientes resultados:

Source	SS	df	MS	Number of obs = 221		
Model	.086660302	1	.086660302	F( 1, 219) = 0.04		
Residual	489.236749	219	2.23395776	Prob > F = 0.8440		
Total	489.32341	220	2.22419732	R-squared = 0.0002		
				Adj R-squared = -0.0044		
				Root MSE = 1.4946		

lprod	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
rec_empleo	-15.23371	77.34514	-0.20	0.844	-167.6698	137.2024
_cons	11.52707	.1344851	85.71	0.000	11.26202	11.79212

$$lprod_t = 11.527 - 15.2337 \left( \frac{1}{empleo_t} \right) + \mu_t \quad (3.5.26)$$

$$ee = (0.1345) \quad (77.3451)$$

$$t = (85.71) \quad (-0.20) \quad R^2 = 0.0002$$

Interpretación:

Si el empleo aumenta indefinidamente, el logaritmo de la producción se acerca a su valor asintótico de aproximadamente 12. Además el valor negativo de  $\beta_2 = -15.23$  implica una relación positiva entre el logaritmo de la producción y el empleo, sin embargo el valor de este parámetro es estadísticamente no significativo, pues su probabilidad es bastante elevada, dando indicios de que los datos no se ajustan a un modelo recíproco logarítmico.

### 3.6 Análisis de regresión múltiple

Los modelos de regresión múltiple son una ampliación de un modelo de regresión simple que contiene únicamente dos variables (una regresada  $Y$  y una regresora  $X$ ), pues ya no sólo cuenta con una variable  $X$  sino con dos o más variables  $X$  para explicar la variación de  $Y$ . Por lo tanto es pertinente analizar el problema de la estimación y la inferencia de los modelos de regresión múltiple. Este análisis se lo ha hecho mediante 13 ejercicios utilizando 6 tablas de datos (Tabla 3.6.1- Tabla 3.6.6). Los 7 primeros ejercicios tratan de explicar el problema de la estimación de los modelos de regresión múltiple abarcando temas como: modelo con tres variables, estimación de coeficientes de regresión parcial, regresión múltiple sobre variables estandarizadas, efecto simultáneo, comparaciones de dos valores  $R^2$ , la función de producción Cobb-Douglas, y los modelos de regresión polinomial. Y los últimos 6 ejercicios explican el problema de la inferencia de los modelos de regresión múltiple abarcando pruebas de: hipótesis sobre coeficientes de regresión individual, significancia general de la regresión muestral, igualdad de dos coeficientes de regresión, restricciones de igualdades lineales, estabilidad estructural (prueba de Chow), y la prueba de la forma funcional de la regresión.

#### 3.6.1 Modelo con tres variables.

##### Ejercicio 1:

Partiendo de la función de regresión poblacional (FRP) de dos variables, se agrega una variable  $X$  así:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \mu_i \quad (3.6.1)^{11}$$

Donde,  $Y$  es la variable dependiente y  $X_2$  y  $X_3$  son las variables independientes. En este modelo  $\beta_1$ , que representa al intercepto, da efecto promedio sobre  $Y$  de todas las variables que están excluidas en el modelo, por otro lado  $\beta_2$  y  $\beta_3$  representan los coeficientes de regresión parcial, y  $\mu_i$  es el término de perturbación estocástica (Gujarati & Porter, 2010).

El significado de los coeficientes de regresión parcial es el siguiente:  $\beta_2$  mide el cambio de  $Y$  con la influencia de  $X_2$  dejando constante la influencia de  $X_3$ , mientras que  $\beta_3$  mide el cambio de  $Y$  con la influencia de  $X_3$  dejando constante la influencia de  $X_2$ .

Comandos a utilizar:

Rutina utilizada:  
**reg MI PIBPC TAM**

---

<sup>11</sup>En el caso de se consideren datos que sean series de tiempo, el subíndice de las variables será  $t$

Desarrollo:

Se tienen los datos de la mortalidad infantil (MI), PIB per cápita (PIBPC) y la tasa de alfabetización de mujeres (TAM) en la Tabla 3.6.1. Utilizando estos datos se va a estimar una regresión múltiple de tres variables, en donde a priori se espera que tanto el PIBPC y la TAM tengan un efecto negativo en la MI. Cabe mencionar que MI es el número de muertes por cada 1000 niños menores a 5 años. Para averiguar aquello se estima la regresión con el siguiente comando: **reg MI PIBPC TAM**. Y se obtienen los siguientes resultados:

Source	SS	df	MS	Number of obs = 55		
Model	125770.949	2	62885.4746	F( 2, 52)	=	353.20
Residual	9258.41806	52	178.046501	Prob > F	=	0.0000
				R-squared	=	0.9314
				Adj R-squared	=	0.9288
Total	135029.367	54	2500.54384	Root MSE	=	13.343

MI	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
PIBPC	-.0390213	.0056543	-6.90	0.000	-.0503675	-.0276751
TAM	-4.536024	.5362936	-8.46	0.000	-5.612176	-3.459872
_cons	553.4028	33.44535	16.55	0.000	486.2898	620.5158

Por lo tanto la regresión múltiple queda representada de la siguiente manera:

$$MI_t = 553.40 - 0.039PIBPC_{2t} - 4.54TAM_{3t} + \mu_t \quad (3.6.2)$$

Donde,  $\beta_1$  es el intercepto con un valor igual a 553.40,  $\beta_2$  y  $\beta_3$  son los coeficientes de regresión parcial con  $-0.039$  y  $-4.54$  respectivamente. Al mirar los signos de los coeficientes de regresión parcial se confirma que efectivamente el efectos del PIBPC y TAM sobre MI es negativo. Además se tiene un valor de  $r^2$  de 0.93 bastante alto que quiere decir que casi el 93% de la variación en la mortalidad infantil se explica mediante el PIBPC y la TAM

Interpretación:

Si se mantiene constante la influencia de la TAM, si el PIBPC incrementa en 1 dólar, la MI disminuye 0.039 unidades. Por otro lado, si se mantiene constante la influencia del PIBPC, si la TAM incrementa en 1 punto porcentual (1%), la MI disminuye en 4.54 unidades.

En términos económicos, manteniendo constante la influencia de la tasa de alfabetización de mujeres, si el PIB per cápita aumenta en 1000 dólares, en promedio, el número de muertes de niños menores a 5 años disminuiría, en promedio, 39 por cada 1000 nacimientos vivos. Y manteniendo constante la influencia del PIB per cápita, si la tasa de alfabetización de mujeres

aumenta 1 punto porcentual, en promedio, el número de muertes de niños menores a 5 años disminuiría, en promedio, 4.54 por cada 1000 nacimientos vivos.

### 3.6.2 Otra manera de estimar los coeficientes de regresión parcial.

#### Ejercicio 2:

Para conservar constante la influencia de una regresora se pueden estimar los coeficientes de regresión parcial mediante tres pasos, si la variable regresora que se desea que se mantenga constante su influencia es  $X_3$ , estos 3 pasos se detallarían de la siguiente manera:

- Paso 1: Se estima  $Y$  sobre  $X_3$ , y se obtienen los residuos  $\hat{\mu}_{1i}$ .
- Paso 2: Se estima  $X_2$  sobre  $X_3$ , y se obtienen los residuos  $\hat{\mu}_{2i}$ .
- Paso 3: Se estima (sin término de intercepto)  $\hat{\mu}_{1i}$  sobre  $\hat{\mu}_{2i}$ .

El valor del coeficiente de  $\hat{\mu}_{2i}$  es el coeficiente de regresión parcial de  $X_2$  manteniendo constante la influencia de  $X_3$ . Si se quiere obtener el coeficiente de regresión parcial de  $X_3$  manteniendo constante la influencia de  $X_2$  se debe volver a realizar el mismo procedimiento. Esta manera es un poco más larga que la realizada en el ejercicio 1, sin embargo ambas llevan a los mismos resultados.

Comandos a utilizar:

```
Rutina utilizada:  
reg MI TAM  
predict r1, resid  
reg PIBPC TAM  
predict r2, resid  
reg r1 r2, noconst
```

Desarrollo:

Para ejemplificar lo ya expuesto anteriormente se utilizarán los mismos datos del ejercicio 1 (Tabla 3.6.1.) y comprobar que se obtienen los mismos resultados de las dos maneras para estimar los verdaderos valores de los coeficientes de regresión parcial. En este ejercicio se considerará constante la influencia de la variable TAM para conocer el verdadero efecto que tiene la variación de PIBPC en el valor de MI.

El primer paso consiste en estimar  $\widehat{MI}_t = \beta_1 + \beta_2 TAM_{2t} + \mu_t$ , es decir MI sobre TAM con el comando: **reg** MI TAM y luego obtener los residuos  $\hat{\mu}_{1t}$  con el comando: **predict** r1, **resid**.

Se obtienen los resultados de la regresión de MI sobre TAM:

Source	SS	df	MS	Number of obs = 55		
Model	117291.346	1	117291.346	F( 1, 53) = 350.46		
Residual	17738.0214	53	334.679649	Prob > F = 0.0000		
				R-squared = 0.8686		
				Adj R-squared = 0.8662		
Total	135029.367	54	2500.54384	Root MSE = 18.294		

MI	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
TAM	-7.618487	.4069588	-18.72	0.000	-8.434743	-6.802231
_cons	709.7742	33.72775	21.04	0.000	642.1249	777.4235

Como segundo paso se estima  $\widehat{PIBPC}_t = \beta_1 + \beta_2 TAM_{2t} + \mu_t$ , es decir PIBPC sobre TAM con el comando: **reg PIBPC TAM** y luego se obtienen los residuos  $\hat{\mu}_{2t}$  con el comando:

**predict r2, resid.**

Se obtienen los resultados de la regresión de PIBPC sobre TAM:

Source	SS	df	MS	Number of obs = 55		
Model	12610170.4	1	12610170.4	F( 1, 53) = 120.01		
Residual	5568935.4	53	105074.253	Prob > F = 0.0000		
				R-squared = 0.6937		
				Adj R-squared = 0.6879		
Total	18179105.8	54	336650.108	Root MSE = 324.15		

PIBPC	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
TAM	78.99439	7.210808	10.95	0.000	64.53134	93.45744
_cons	-4007.333	597.6141	-6.71	0.000	-5205.995	-2808.671

El tercer paso consiste en estimar  $\hat{\mu}_{1t} = \beta_1 \hat{\mu}_{2t}$ , es decir  $\hat{\mu}_{1t}$  sobre  $\hat{\mu}_{2t}$  con el comando: **reg r1 r2, noconst.**

Se obtienen los resultados de la regresión sin intercepto de  $\hat{\mu}_{1t}$  sobre  $\hat{\mu}_{2t}$ :

Source	SS	df	MS	Number of obs = 55		
Model	8479.60327	1	8479.60327	F( 1, 54) = 49.46		
Residual	9258.41825	54	171.45219	Prob > F = 0.0000		
				R-squared = 0.4780		
				Adj R-squared = 0.4684		
Total	17738.0215	55	322.509482	Root MSE = 13.094		

r1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
r2	-.0390213	.0055486	-7.03	0.000	-.0501456	-.027897

Por lo tanto reemplazando los valores se tiene:

$$\hat{\mu}_{1t} = -0.039\hat{\mu}_{2t} \quad (3.6.3)$$

En (3.6.3) el valor  $-0.039$  es el coeficiente de regresión parcial  $\beta_2$  de la regresora PIBPC manteniendo constante la influencia de la regresora TAM. Con ello se comprueba que este valor es el mismo obtenido en el ejercicio 1. Además si se quiere obtener el coeficiente de regresión parcial de TAM se puede realizar este mismo procedimiento.

Interpretación:

Manteniendo constante la influencia de la tasa de alfabetización de mujeres, si el PIB per cápita aumenta en 1000 dólares, en promedio, el número de muertes de niños menores a 5 años disminuiría, en promedio, 39 por cada 1000 nacimientos vivos. La interpretación de este ejercicio es la misma que la del ejercicio 1 ya que se obtuvo el mismo resultado pero de distinta manera.

### 3.6.3 Regresión múltiple sobre variables estandarizadas.

#### Ejercicio 3:

En el apartado 3.5.3 Regresiones sobre variables estandarizadas, ya se analizó este tema, sin embargo es necesario mencionar que este tipo de análisis se puede extender a regresiones múltiples o multivariadas. Hay que recordar que las variables en este tipo de regresiones se expresan en unidades de desviación estándar. Estas variables se definen de la siguiente manera:

$$Y_i^* = \frac{Y_i - \bar{Y}}{s_Y} \quad (3.6.4)$$

Comandos a utilizar:

```
Rutina utilizada:
/*Generar variables estandarizadas paso a paso*/
sum MI
gen MIstan=(MI-80.06364)/50.00544
sum PIBPC
gen PIBPCstan= (PIBPC-2521.994)/580.2156
sum PIBPC
gen TAMstan= (TAM-82.65559)/6.117407
/*Generar variables estandarizadas directamente*/
egen MIstan= std(MI)
egen PIBPCstan= std(PIBPC)
egen TAMstan= std(TAM)

/*Estimar regresión múltiple sobre variables estandarizadas*/
reg MIstan PIBPCstan TAMstan, noconst
```

Desarrollo:

Continuando con el uso los datos de la Tabla 3.6.1, se procede a estimar una regresión múltiple sobre variables estandarizadas así:

$$MI_t^* = \beta_2^* PIBPC_t^* + \beta_3^* TAM_t^* + \mu_t^* \quad (3.6.5)$$

Por lo tanto es necesario generar la variable estandarizada de cada una de las 3 variables del modelo (3.6.5), para lo cual se hace uso del comando **sum** el cual brinda el valor de la media y la desviación estándar de la variable, y luego se generan las variables estandarizadas como lo muestra la siguiente rutina de comandos:

```
sum MI
gen MIstan=(MI-80.06364)/50.00544
sum PIBPC
gen PIBPCstan= (PIBPC-2521.994)/580.2156
sum TAM
gen TAMstan= (TAM-82.65559)/6.117407
```

Por ejemplo para generar la variable estandarizada de MI al ejecutar el comando **sum MI**, tomo los valores de la media 80.06364 y de la desviación estándar 50.00544, y genero la variable, lo mismo se hace con todas las variables que se han estandarizado.

Se pueden generar variables estadarizadas directamente con los comandos:

```
egen MIstan= std(MI)
egen PIBPCstan= std(PIBPC)
egen TAMstan= std(TAM)
```

Una vez generadas las 3 variables estandarizadas se procede a estimar la regresión mediante el comando: **reg MIstan PIBPCstan TAMstan, noconst.** Y se obtienen los siguientes resultados:

Source	SS	df	MS	Number of obs = 55		
Model	50.2974349	2	25.1487175	F( 2, 53) = 359.99		
Residual	3.70256164	53	.069859654	Prob > F = 0.0000		
Total	53.9999966	55	.98181812	R-squared = 0.9314		
				Adj R-squared = 0.9288		
				Root MSE = .26431		

MIstan	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
PIBPCstan	-.452766	.0649855	-6.97	0.000	-.5831104	-.3224216
TAMstan	-.5549137	.0649855	-8.54	0.000	-.6852581	-.4245693



Por lo tanto la regresión quedaría expresada de la siguiente manera:

$$MI_t^* = -0.4528PIBPC_t^* + -0.5549TAM_t^* + \mu_t^* \quad (3.6.6)$$

Interpretación:

Si se conserva constante la influencia de la tasa de alfabetización de mujeres (TAM), ante un incremento igual a una desviación estándar del PIB per cápita (PIBPC), en promedio, disminuirá 0.45 desviación estándar en la mortalidad infantil (MI). Y por otro lado, manteniendo constante la influencia del PIB per cápita, ante un incremento igual a una desviación estándar de la tasa de alfabetización de mujeres, en promedio, producirá una disminución de 0.55 desviación estándar en la mortalidad infantil.

### 3.6.4 Efecto simultáneo.

#### Ejercicio 4:

Cuando se habla de efecto simultáneo no es más que el efecto sobre la variable dependiente de un cambio unitario en más de una regresora. Por ejemplo esto ocurriría si  $X_2$  y  $X_3$  incrementarían de manera simultánea (al mismo tiempo), ya no sólo  $X_2$  o sólo  $X_3$ . Por lo tanto, para averiguar el efecto simultáneo de un aumento unitario de  $X_2$  y  $X_3$  en  $Y$  se suman los coeficientes de  $X_2$  y  $X_3$ .

$$\text{Efecto simultáneo} = \beta_2 + \beta_3 \quad (3.6.7)$$

Sin embargo qué ocurriría cuando  $X_2$  y  $X_3$  tienen un aumento diferente, por ejemplo,  $X_2$  aumenta dos unidades y  $X_3$  aumenta una unidad, en este caso se multiplica 2 por el coeficiente de  $X_2$  más el coeficiente de  $X_3$ , así:

$$\text{Efecto simultáneo} = 2(\beta_2) + \beta_3 \quad (3.6.8)$$

Es importante mencionar que el efecto neto es  $\beta_2$  o  $\beta_3$  según sea el caso, es decir si se mantiene constante la influencia de  $X_3$  el efecto neto de un cambio unitario de  $X_2$  en  $Y$  será  $\beta_2$ , como se vio en los ejercicios 1 y 2.

Comandos a utilizar:

Rutina utilizada:  
**reg MI PIBPC TAM**

Desarrollo:

Utilizando el mismo modelo de la mortalidad infantil (Tabla 3.6.1) se tratará de averiguar el efecto simultáneo de un cambio unitario del PIBPC y de la TAM en MI. Para ello es necesario

estimar la regresión  $MI_t = \beta_1 + \beta_2 PIBPC_{2t} + \beta_3 TAM_{3t} + \mu_t$ , con el siguiente comando:  
**reg MI PIBPC TAM.** Y se obtiene:

Source	SS	df	MS	Number of obs = 55		
Model	125770.949	2	62885.4746	F( 2, 52) = 353.20		
Residual	9258.41806	52	178.046501	Prob > F = 0.0000		
				R-squared = 0.9314		
				Adj R-squared = 0.9288		
Total	135029.367	54	2500.54384	Root MSE = 13.343		

MI	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
PIBPC	-.0390213	.0056543	-6.90	0.000	-.0503675	-.0276751
TAM	-4.536024	.5362936	-8.46	0.000	-5.612176	-3.459872
_cons	553.4028	33.44535	16.55	0.000	486.2898	620.5158

$$MI_t = 553.40 - 0.039PIBPC_{2t} - 4.54TAM_{3t} + \mu_t \quad (3.6.9)$$

De acuerdo con los resultados de la regresión (3.6.9) el efecto neto es  $\beta_2 = -0.039$  y  $\beta_3 = -4.54$ , pero el efecto simultáneo de un cambio unitario del PIBPC y de la TAM es la suma de  $\beta_2$  y  $\beta_3$ , así:

$$\text{Efecto simultáneo} = \beta_2 + \beta_3$$

$$\text{Efecto simultáneo} = -0.039 - 4.54 = -4.58$$

Ahora, si el PIBPC aumenta 2 unidades, y la TAM aumenta solamente 1 unidad, ¿qué efecto producirían estos cambios en MI?. Nótese que el cambio de las dos regresoras es diferente por tanto para responder a la pregunta se hace lo siguiente:

$$\text{Efecto simultáneo} = 2(\beta_2) + \beta_3$$

$$\text{Efecto simultáneo} = 2(-0.039) - 4.54 = -4.62$$

Interpretación:

Ante un cambio simultáneo de 1 unidad del PIB per cápita y de la tasa de alfabetización de mujeres, el número de muertes de niños menores a 5 años disminuiría en casi 4,58 muertes. Mientras que si se produce un cambio simultáneo de 2 unidades del PIB per cápita y 1 unidad de la tasa de alfabetización de mujeres, el número de muertes de niños menores a 5 años disminuiría en casi 4.62 muertes.

### 3.6.5 Comparación de dos valores de $R^2$ .

#### Ejercicio 5:

Si se presentan dos modelos, los cuales tienen el mismo tamaño de la muestra  $n$  pero la variable regresada se expresa en distinta forma, no se podrían comparar los términos  $R^2$  calculados de dichos modelos. Por ejemplo para los modelos:

$$\ln Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \mu_i \quad (3.6.10)$$

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \mu_i \quad (3.6.11)$$

“No pueden compararse los términos  $R^2$  calculados. La razón es la siguiente: por definición,  $R^2$  mide la proporción de la variación en la variable dependiente explicada por la(s) variable(s) explicativa(s). Por consiguiente, en (3.6.10) el  $R^2$  mide la proporción de la variación en  $\ln Y$  explicada por  $X_2$  y  $X_3$ , mientras que en (3.6.11) mide la proporción de la variación en  $Y$ , y las dos no son la misma variable” (Gujarati & Porter, 2010, p.203). Entonces, ¿Es posible comparar las  $R^2$  de estos modelos? ¿Cómo hacerlo? Si es posible hacer esta comparación, a continuación se ejemplificará la manera de comparar dos valores de  $R^2$ .

Comandos a utilizar:

```
Rutina utilizada:
reg ING EMP
gen lnING=log(ING)
gen lnEMP=log(EMP)
reg lnING lnEMP
/*Comparar (3.6.13) con (3.6.12)*/
reg lnING lnEMP
predict INGf
gen antilogINGf=exp(INGf)
reg antilogINGf ING
/*Comparar (3.6.12) con (3.6.13)*/
reg ING EMP
predict INGf1
gen lnINGf1=log(INGf1)
gen lnING=log(ING)
reg lnINGf1 lnING
```

(Ya está definida anteriormente)

Desarrollo:

Utilizando los datos de la Tabla 3.6.2 que contienen datos del ingreso cantonal (ING) y el número de empresas (EMP), se van a comparar los  $R^2$  de los siguientes modelos:

$$\widehat{ING}_i = -521.97 + 0,5114EMP_i + \mu_i \quad R^2 = 0.9156 \quad (3.6.12)$$

$$\widehat{\ln ING}_i = -5.1815 + 1,34624\ln EMP_i + \mu_i \quad R^2 = 0.8976 \quad (3.6.13)$$

Los resultados presentados en (3.6.12) y (3.6.13) se los obtuvieron con un procedimiento previo mediante la ejecución de los siguientes comandos:

```
reg ING EMP
gen lnING=log(ING)
gen lnEMP=log(EMP)
reg lnING lnEMP
```

Ahora, para comparar los dos valores  $R^2$  obtenidos en los modelos (3.6.12) y (3.6.13) y conocer así cuál es el mejor modelo, primero se va a comparar (3.6.13) con (3.6.12), siguiendo el siguiente procedimiento:

1. Obtener Y estimada (INGf) del modelo (3.6.13).  

```
reg lnING lnEMP
predict INGf
```
2. Generar la variable antilogarítmica de INGf  

```
gen antilogINGf=exp(INGf)
```
3. Estimar la regresión antilogINGf sobre INGf, y obtener  $R^2$ .  

```
reg antilogINGf ING
```

Source	SS	df	MS	Number of obs = 221		
Model	1.5878e+09	1	1.5878e+09	F( 1, 219) = 4972.74		
Residual	69925936	219	319296.511	Prob > F = 0.0000		
Total	1.6577e+09	220	7535017.24	R-squared = 0.9578		
				Adj R-squared = 0.9576		
				Root MSE = 565.06		

antilogINGf	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ING	.5314971	.0075371	70.52	0.000	.5166426	.5463516
_cons	64.75534	38.33437	1.69	0.093	-10.79616	140.3068

Se tiene un  $R^2 = 0.9576$ , el cual es mayor al  $R^2 = 0.9156$  del modelo (3.6.12), por lo tanto se elige el modelo (3.6.13) ya que su  $R^2$  es el mayor.

Segundo se va a comparar (3.6.12) con (3.6.13), siguiendo el siguiente procedimiento:

1. Obtener Y estimada (INGf1) del modelo (3.6.12).  

```
reg ING EMP
predict INGf1
```
2. Generar la variable logarítmica de INGf  

```
gen lnINGf1=log(INGf1)
```

3. Generar la variable logarítmica de ING  
**gen lnING=log(ING)**
4. Estimar la regresión lnINGf1 sobre lnING  
**reg lnINGf1 lnING**

Source	SS	df	MS	Number of obs = 70		
Model	151.804041	1	151.804041	F( 1, 68) = 262.52		
Residual	39.3208149	68	.578247277	Prob > F = 0.0000		
Total	191.124856	69	2.76992545	R-squared = 0.7943		
				Adj R-squared = 0.7912		
				Root MSE = .76043		

lnINGf1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnING	.9225606	.056939	16.20	0.000	.8089405	1.036181
_cons	1.319421	.3280006	4.02	0.000	.6649063	1.973936

Se tiene un  $R^2 = 0.7912$ , el cual es menor al  $R^2 = 0.8976$  del modelo (3.6.13), por lo tanto se elige el modelo (3.6.13) ya que su  $R^2$  es el mayor.

Interpretación:

A pesar de que el  $R^2$  del modelo (3.6.12) es mayor al  $R^2$  del modelo (3.6.13) se elige al modelo (3.6.13), ya que al realizar las comparaciones respectivas de estos dos valores el  $R^2$  del modelo (3.6.13) resultó mayor. En primera instancia talvez se hubiese elegido erróneamente al modelo (3.6.12) por su  $R^2$  mayor, pero con los procedimientos comparativos se ha determinado que se debe elegir al modelo (3.6.13). Por lo tanto, al interpretar los resultados del modelo elegido (3.6.13) se tiene que, a medida que el número de empresas aumenta en 1%, en promedio, el ingreso cantonal aumentará el 0.92%. Además el valor de  $R^2 = 0.8976$  significa que el 89% de la variación del ingreso cantonal en el Ecuador es explicado por la variación en el logaritmo de la cantidad de empresas en dichos cantones.

### 3.6.6 La función de producción Cobb-Douglas: más sobre la forma funcional.

#### Ejercicio 6:

Este apartado muestra cómo convertir relaciones no lineales en lineales mediante transformaciones necesarias y así poder trabajar dentro del marco de MCRL. El modelo de producción Cobb-Douglas demuestra transformaciones con una extensión multivariada del modelo log-lineal de dos variables (Gujarati & Porter, 2010).

Como lo mencionan Gujarati & Porter (2010) la función de producción Cobb-Douglas, en su forma estocástica se expresa como:

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{\mu_i} \quad (3.6.14)$$

Donde,  $Y$  es la producción,  $X_2$  el insumo trabajo,  $X_3$  el insumo capital,  $\mu$  es el término de perturbación estocástica, y  $e$  la base del logaritmo natural. Es claro que de la ecuación (3.6.14) la relación entre la producción  $Y$  y los dos insumos ( $X_2$  y  $X_3$ ) no es lineal, por lo tanto es necesario transformarla, así:

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + \mu_i \quad (3.6.15)$$

$$= \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + \mu_i$$

Comandos a utilizar:

Rutina utilizada:  
**gen lnY=log(Y)**  
**gen lnL=log(L)**  
**gen lnK=log(K)**  
**reg lnY lnL lnK**

Desarrollo:

Se pretende estimar la función de producción Cobb-Douglas utilizando los datos de producción ( $Y$ ), trabajo ( $L$ ) y capital ( $K$ ) de la Tabla 3.6.3. Basándose en el modelo (3.6.15) el modelo a estimar se expresaría de la siguiente manera:

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln L_{2i} + \beta_3 \ln K_{3i} + \mu_i \quad (3.6.16)$$

Para estimar este modelo de producción es necesario crear primeramente las variables logarítmicas de  $Y$ ,  $L$ , y  $K$ , y luego sí se estima la regresión log lineal o doble log. Por consiguiente se ejecutan los siguientes comandos:

**gen lnY=log(Y)**  
**gen lnL=log(L)**  
**gen lnK=log(K)**  
**reg lnY lnL lnK**

Obteniendo los siguientes resultados del modelo de producción Cobb-Douglas:

Source	SS	df	MS	Number of obs =	221
Model	382.077756	2	191.038878	F( 2, 218) =	501.45
Residual	83.0525632	218	.380975061	Prob > F =	0.0000
				R-squared =	0.8214
				Adj R-squared =	0.8198
Total	465.130319	220	2.11422872	Root MSE =	.61723

lnY	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnL	.9021417	.0614754	14.67	0.000	.7809794	1.023304
lnK	.1577797	.0405109	3.89	0.000	.0779366	.2376228
_cons	7.087062	.3798796	18.66	0.000	6.338355	7.835769

$$\ln Y_i = 7.0871 + 0.9021 \ln L_{2i} + 0.1578 \ln K_{3i} + \mu_i \quad (3.6.17)$$

$$R^2 = 0.8214$$

Al sumar  $(\beta_2 + \beta_3)$  del modelo (3.6.17) se tiene que:  $0.9021 + 0.1578 = 1.0599$ , este valor es mayor a 1, por lo tanto existen rendimientos crecientes a escala.

Interpretación:

En la producción del Ecuador durante el 2010, las elasticidades de la producción respecto del trabajo y el capital fueron 0.9021 y 0.1578 respectivamente. En otras palabras, manteniendo constante el insumo capital un incremento del 1% en el insumo trabajo, en promedio, producirá un aumento del 0.9% en la producción, del mismo modo, manteniendo constante el insumo trabajo un incremento del 1% en el insumo capital, en promedio, producirá un aumento del 0.16% en la producción. Además, si el trabajo y el capital aumentan 1% simultáneamente, la producción nacional aumentará, en promedio, 1.06%, teniendo así rendimientos crecientes a escala en el sector productivo del Ecuador.

### 3.6.7 Modelos de regresión polinomial.

#### Ejercicio 7:

Los modelos de regresión polinomial son una clase de modelos de regresión múltiple. Este tipo de modelos pueden ser de grado 2, 3, 4, ..., k, es decir modelos cuadráticos, cúbicos, etc. Las funciones de costo y producción son un ejemplo claro de un modelo de regresión polinomial.

La regresión polinomial de grado k general puede escribirse así:

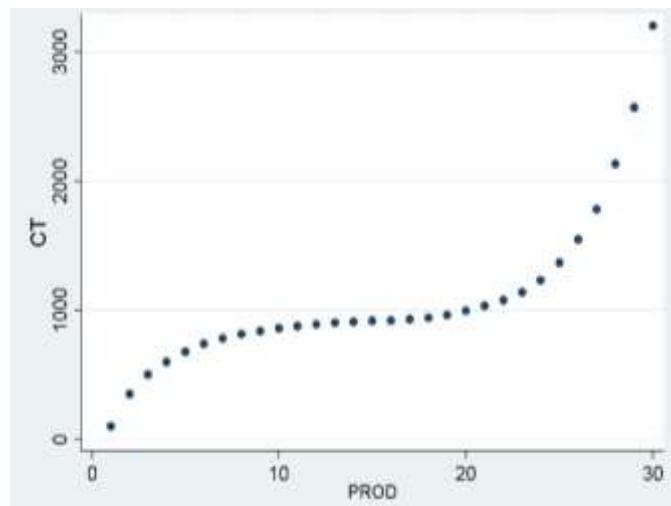
$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \dots + \beta_k X_i^k + \mu_i \quad (3.6.18)$$

Comandos a utilizar:

```
Rutina utilizada:  
scatter CT PROD  
gen PROD2=PROD^2  
gen PROD3=PROD^3  
reg CT PROD PROD2 PROD3
```

Desarrollo:

Considerando datos hipotéticos del costo total (CT) y de la producción (PROD) de la Tabla 3.6.4, se trata de determinar ¿qué tipo de modelo de regresión se ajusta a estos datos?, para ello se grafica un diagrama de dispersión con el comando: **scatter** CT PROD



Observando el diagrama de dispersión se evidencia que la relación entre el costo total y la producción es similar a una curva en S. Este tipo de curva en S del costo total se representa por la siguiente regresión polinómica cúbica o de tercer grado:

$$CT_i = \beta_0 + \beta_1 PROD_i + \beta_2 PROD_i^2 + \beta_3 PROD_i^3 + \mu_i \quad (3.6.19)$$

Antes de estimar este modelo de regresión polinomial es necesario crear las variables de segundo y tercer grado de la variable PROD, para ello se ejecutan los siguientes comandos:

```
gen PROD2=PROD^2  
gen PROD3=PROD^3
```

Una vez generadas todas las variables del modelo, se procede a estimarlo aplicando el método de MCO mediante el comando: **reg** CT PROD PROD2 PROD3. Y se obtiene:



Source	SS	df	MS	Number of obs = 30		
Model	11284995.3	3	3761665.09	F( 3, 26) =	466.66	
Residual	209581.684	26	8060.834	Prob > F =	0.0000	
				R-squared =	0.9818	
				Adj R-squared =	0.9797	
Total	11494577	29	396364.723	Root MSE =	89.782	

CT	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
PROD	265.5397	20.52757	12.94	0.000	223.3447	307.7348
PROD2	-20.87652	1.525338	-13.69	0.000	-24.0119	-17.74114
PROD3	.5156667	.0323769	15.93	0.000	.449115	.5822185
_cons	-166.3177	74.69993	-2.23	0.035	-319.8656	-12.76977

$$CT_i = -166.32 + 265.54PROD_i - 20.88PROD_i^2 + 0.52PROD_i^3 + \mu_i \quad (3.6.20)$$

$$R^2 = 0.9818$$

Se muestra que el modelo de regresión polinomial de tercer grado es el modelo que se ajusta a los datos de costo total y producción, ya que los resultados del modelo (3.6.20) muestra que los coeficientes parciales de las regresoras son todos significativos, y se tiene un  $R^2$  alto.

Interpretación:

A medida que la producción crece, el costo total aumenta con rendimientos decrecientes, es decir el aumento de costo total es cada vez más pequeño, pero, después de un nivel dado de producción, el costo total vuelve a aumentar. Además el valor de  $R^2 = 0.9818$  significa que el 98.18% de la variación en el costo total es explicado por la variación de la producción, producción al cuadrado y producción al cubo.

### 3.6.8 Pruebas de hipótesis sobre coeficientes de regresión individual.

#### Ejercicio 8:

Se puede utilizar la prueba  $t$  para demostrar una hipótesis sobre cualquier coeficiente de regresión parcial. Por ejemplo si se quiere realizar una prueba de hipótesis sobre el coeficiente  $\beta_2$ , se tiene el siguiente conjunto de hipótesis:

$$H_0: \beta_2 = 0$$

$$H_1: \beta_2 \neq 0$$

Decisión: Si el si el valor el valor absoluto de  $t$  calculado excede el valor de  $t$  crítico en el nivel de significancia escogido, se rechaza  $H_0$ , por lo tanto se acepta  $H_1$ .

La  $t$  calculada se encuentra definido como:

$$t = \frac{\hat{\beta}_2 - \beta_2}{ee(\hat{\beta}_2)} \quad (3.6.21)$$

Donde,  $\hat{\beta}_2$  es el coeficiente de regresión parcial estimado,  $\beta_2$  es el valor hipotético del coeficiente de regresión parcial a comprobar (en este caso es 0), y  $ee(\hat{\beta}_2)$  es la desviación estándar de  $\hat{\beta}_2$ .

Por otro lado, el valor de  $t$  crítico viene dado por el valor encontrado en la tabla  $t$  del Anexo 3, el cual se basa de los valores de: la muestra  $n$ , los grados de libertad  $gl = n - \# \text{ parámetros o estimadores } \beta$ , y el nivel de significancia  $\alpha$ .

Comandos a utilizar:

Rutina utilizada:  
**gen MI PIBPC TAM**

Desarrollo:

Retornando al uso de los datos de la Tabla 3.6.1, se quiere comprobar si el coeficiente de regresión parcial de la regresora TAM es estadísticamente igual a 0<sup>12</sup>. Primeramente es necesario estimar la regresión de mortalidad infantil sobre PIBPC y TAM con el comando: **gen MI PIBPC TAM**. Y se obtienen los siguientes resultados:

Source	SS	df	MS	Number of obs = 55		
Model	125770.949	2	62885.4746	F( 2, 52) = 353.20		
Residual	9258.41806	52	178.046501	Prob > F = 0.0000		
				R-squared = 0.9314		
				Adj R-squared = 0.9288		
Total	135029.367	54	2500.54384	Root MSE = 13.343		

MI	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
PIBPC	-.0390213	.0056543	-6.90	0.000	-.0503675	-.0276751
TAM	-4.536024	.5362936	-8.46	0.000	-5.612176	-3.459872
_cons	553.4028	33.44535	16.55	0.000	486.2898	620.5158

$$MI_t = 553.40 - 0.039PIBPC_{2t} - 4.54TAM_{3t} + \mu_t \quad (3.6.22)$$

<sup>12</sup> El valor que se somete a comprobación mediante la prueba de hipótesis, no solamente debe ser 0 sino que puede tomar cualquier valor.

Se formula el conjunto de hipótesis para averiguar si el coeficiente de regresión parcial de TAM es estadísticamente igual a cero, es decir si al mantener constante el PIB per cápita (PIBPC), la tasa de alfabetización de las mujeres (TAM) no tiene influencia (lineal) sobre la mortalidad infantil (MI). Por consiguiente el conjunto de hipótesis es:

$$H_0: \beta_3 = 0$$

$$H_1: \beta_3 \neq 0$$

Para averiguar el valor de  $t$  calculada tomamos los valores de  $\hat{\beta}_3 = -4.54$ ,  $\beta_3 = 0$  y  $ee\hat{\beta}_3 = 0.5363$ . Por lo tanto el cálculo de  $t$  se detalla de la siguiente manera:

$$t = \frac{-4.54 - 0}{0.5363}$$

$$t = -8.4654$$

Ahora, para saber el valor de  $t$  crítico se tiene:  $n=55$ ,  $gl = 55-3 = 52$  y se considera un  $\alpha=5\%=0.05$ . Se acude a la tabla  $t$  del Anexo 3, y como no hay  $gl=52$  se toma el valor más cercano, es decir  $gl=60$ , por lo tanto  $t=2$ .

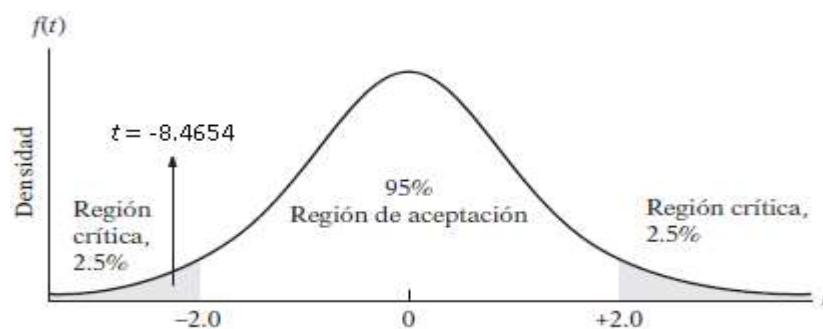
Al comparar los valores de  $t$  calculada y  $t$  crítica se tiene que:

$$|t \text{ calculado}| > t \text{ crítica}$$

$$|-8.4654| > 2$$

$$8.4654 > 2$$

El valor de  $t$  calculada = -8.4654 cae en la zona de rechazo o crítica como se muestra a continuación:



Por lo tanto, se rechaza  $H_0$  y se acepta  $H_1$ .

Interpretación:

Al rechazar  $H_0$  y aceptar  $H_1$  se afirma que al mantener constante el PIB per cápita, la tasa de alfabetización de las mujeres sí tiene influencia (lineal) sobre la mortalidad infantil. En el caso de que se aceptare  $H_0$  se debería eliminar la variable tasa de alfabetización de mujeres ya no tendría influencia (lineal) sobre la mortalidad infantil.

### 3.6.9 Pruebas de significancia general de la regresión muestral.

#### **Ejercicio 9:**

La prueba de significancia general trata de comprobar conjunta o simultáneamente si  $\beta_2$  y  $\beta_3$  son iguales a cero, es decir si  $Y$  está o no relacionada linealmente con  $X_2$  y  $X_3$  a la vez, en donde se trata de probar la siguiente hipótesis:

$$H_0: \beta_2 = \beta_3 = 0$$

Esta hipótesis conjunta se prueba con la técnica del análisis de varianza mediante la prueba  $F$ , esta prueba al igual que la vista en el ejercicio 7 se basa en comparar dos valores que en este caso son la  $F$  calculada y la  $F$  crítica.

La  $F$  calculada se define de la siguiente manera:

$$F = \frac{SCE / gl}{SCR / gl} = \frac{SCE / (k-1)}{SCR / (n-k)} \quad (3.6.22)$$

Y la  $F$  crítica se basa en los valores del *numerador* =  $k - 1$ , y del *denominador* =  $n - k$ , es decir  $F_\alpha (k - 1, n - k)$ . Este valor se lo encuentra en la tabla  $F$  (Anexo 4).

Decisión: “Si el valor de la  $F$  calculada excede al valor de la  $F$  crítica de la tabla  $F$  en el nivel de significancia escogido  $\alpha$ , se rechaza  $H_0$ , de otra forma no se rechaza  $H_0$ ” (Gujarati & Porter, 2010, p.239).

Comandos a utilizar:

Rutina utilizada:  
**gen MI PIBPC TAM**

Desarrollo:

Continuando con el análisis del mismo conjunto de datos del ejercicio 8 (Tabla 3.6.1) se va a probar la hipótesis conjunta:

$$H_0: \beta_2 = \beta_3 = 0$$

Se estima la regresión de MI sobre PIBPC y TAM con el comando: **gen MI PIBPC TAM**, y se obtienen los siguientes resultados:

Source	SS	df	MS	Number of obs = 55		
Model	125770.949	2	62885.4746	F( 2, 52) = 353.20		
Residual	9258.41806	52	178.046501	Prob > F = 0.0000		
Total	135029.367	54	2500.54384	R-squared = 0.9314		
				Adj R-squared = 0.9288		
				Root MSE = 13.343		

MI	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
PIBPC	-.0390213	.0056543	-6.90	0.000	-.0503675	-.0276751
TAM	-4.536024	.5362936	-8.46	0.000	-5.612176	-3.459872
_cons	553.4028	33.44535	16.55	0.000	486.2898	620.5158

Con estos resultados se tiene que  $SCE=125770.949$  con  $gl = 2$  (resulta de  $k - 1 = 3 - 1 = 2$ ),  $SCR= 9258.4181$  con  $gl = 52$  (resulta de  $n - k = 55 - 3 = 52$ ). Aplicando la fórmula (3.6.22) se tiene:

$$F = \frac{SCE/gl}{SCR/gl} = \frac{SCE/(k-1)}{SCR/(n-k)}$$

$$F = \frac{125770.949/(3-1)}{9258.4181/(55-3)}$$

$$F = 353.20$$

Este resultado de  $F$  calculada ya se encuentra calculado por Stata, como se observa en la segunda fila de la parte superior izquierda del recuadro de resultados de la ecuación estimada. Por otro lado es necesario averiguar el valor de la  $F$  crítica por lo tanto se tiene que:  $F_{\alpha}(k-1, n-k) = F_{\alpha}(3-1, 55-3) = F_{\alpha}(2, 52)$ . Con estos resultados hay que dirigirse a la Tabla  $F$ , en cual  $gl$  para el numerador es 2 y  $gl$  para el denominador es 52, como la tabla no contiene el valor de 52 se elige el valor de 60 ya que es el más cercano, por lo tanto se determina que el valor de  $F$  crítica es 3.15.

Finalmente comparando los valores de la  $F$  calculada y la  $F$  crítica se tiene que:

$$F_{\text{calculada}} > F_{\text{crítica}}$$

$$353.20 > 3.15$$

Por lo tanto, se rechaza  $H_0$ .

Interpretación:

Al rechazar  $H_0$  se afirma que el PIB per cápita y la tasa de alfabetización de mujeres, conjuntamente, sí tienen influencia (lineal) sobre la mortalidad infantil.

### 3.6.10 Prueba de igualdad de dos coeficientes de regresión.

#### Ejercicio 10:

Esta prueba trata de comprobar si dos coeficientes  $\beta$  son iguales, por ejemplo, en la regresión múltiple (3.6.23), se desea comprobar la hipótesis de que  $\beta_3$  y  $\beta_4$  son iguales.

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \mu_i \quad (3.6.23)$$

El conjunto de hipótesis (nula y alternativa) quedaría así:

$$H_0: \beta_3 = \beta_4 \quad \text{o} \quad (\beta_3 - \beta_4) = 0$$

$$H_0: \beta_3 \neq \beta_4 \quad \text{o} \quad (\beta_3 - \beta_4) \neq 0$$

En el caso de que se cumpla la hipótesis nula ( $H_0$ ) al ser iguales  $\beta_3 = \beta_4$ , la diferencia entre estos dos coeficientes será 0. Esto significaría que las elasticidades de  $X_3$  y  $X_4$  son las mismas. Para comprobar esta hipótesis se hace uso de la prueba  $t$ , donde la  $t$  calculada se define de la siguiente manera:

$$t = \frac{\hat{\beta}_3 - \hat{\beta}_4}{\sqrt{\text{var}(\hat{\beta}_3) + \text{var}(\hat{\beta}_4) - 2\text{cov}(\hat{\beta}_3, \hat{\beta}_4)}} \quad (3.6.24)$$

Decisión: “Si el valor absoluto de la variable  $t$  calculada de (3.6.24) excede el valor  $t$  crítico en el nivel de significancia designado para los  $gl$  dados, se puede rechazar la hipótesis nula; de lo contrario, no se rechaza” (Gujarati & Porter, 2010, p.247).

Comandos a utilizar:

**covariance:** Proporciona la matriz de varianza-covarianza de los estimadores determinados. VCE puede ser utilizado después de cualquier instrucción de estimación.

Rutina utilizada:

**gen** PROD2=PROD^2

**gen** PROD3=PROD^3

**reg** CT PROD PROD2 PROD3

**vce, covariance**

Desarrollo:

Retornando al uso de los datos hipotéticos del costo total (CT) y de la producción (PROD) de la Tabla 3.6.4. A partir de los resultados de la regresión (3.6.20) se pretende comprobar o rechazar la hipótesis nula de que  $\beta_3 = \beta_4$ . Por lo tanto es necesario primeramente estimar nuevamente la regresión (3.6.19), mediante los siguientes comandos:

**gen** PROD2=PROD^2

**gen** PROD3=PROD^3

**reg** CT PROD PROD2 PROD3

Y se obtienen los siguientes resultados:

Source	SS	df	MS	Number of obs = 30		
Model	11284995.3	3	3761665.09	F( 3, 26)	=	466.66
Residual	209581.684	26	8060.834	Prob > F	=	0.0000
				R-squared	=	0.9818
				Adj R-squared	=	0.9797
Total	11494577	29	396364.723	Root MSE	=	89.782

CT	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
PROD	265.5397	20.52757	12.94	0.000	223.3447	307.7348
PROD2	-20.87652	1.525338	-13.69	0.000	-24.0119	-17.74114
PROD3	.5156667	.0323769	15.93	0.000	.449115	.5822185
_cons	-166.3177	74.69993	-2.23	0.035	-319.8656	-12.76977

$$CT_i = -166.32 + 265.54PROD_i - 20.88PROD_i^2 + 0.52PROD_i^3 + \mu_i \quad (3.6.20)$$

$$ee = (74.69993) \quad (20.52757) \quad (1.525338) \quad (0.0323769) \quad R^2 = 0.981$$

En los resultados presentados se tienen los valores de los coeficientes de regresión parcial estimados y su respectivos errores estándar, sin embargo es necesario que se calcule también el valor de la covarianza entre  $\beta_3$  y  $\beta_4$  para poder calcular la  $t$  calculada mediante la fórmula (3.6.24). Para calcular la covarianza requerida se ejecuta el comando: **vce, covariance**. Obteniendo así la siguiente matriz de varianza-covarianza:

e(V)	PROD	PROD2	PROD3	_cons
PROD	421.38115			
PROD2	-30.430403	2.3266555		
PROD3	.61438849	<b>-.04874435</b>	.00104827	
_cons	-1368.9642	89.711891	-1.7158012	5580.0796

Sabiendo que, el coeficiente de  $PROD2 (X_3)$  es  $\hat{\beta}_3$  y el coeficiente de  $PROD3 (X_4)$  es  $\hat{\beta}_4$ , se tiene que:  $cov(\hat{\beta}_3, \hat{\beta}_4) = -0.04874435$ . Con ello, ya se puede calcular el valor de  $t$  con la fórmula (3.6.24)<sup>13</sup>, así:

$$t = \frac{\hat{\beta}_3 - \hat{\beta}_4}{\sqrt{var(\hat{\beta}_3) + var(\hat{\beta}_4) - 2cov(\hat{\beta}_3, \hat{\beta}_4)}}$$

$$t = \frac{-20.87652 - 0.5156667}{\sqrt{(1.525338)^2 + (0.0323769)^2 - 2(-0.04874435)}} \quad (3.6.25)$$

$$t = \frac{-21.3921867}{\sqrt{2.326656014 + 0.001048264 + 0.0974887}}$$

$$t = \frac{-21.3921867}{1.55730311} = -13.7367$$

Ahora bien, para averiguar el valor de  $t$  crítico se tiene:  $n=30$ ,  $gl = 30-4 = 26$  y se considera un  $\alpha=5\%=0.05$ . Se acude a la tabla  $t$  del Anexo 3, por lo tanto  $t=2.056$ .

Al comparar los valores de  $t$  calculada y  $t$  crítica se tiene que:  $|t \text{ calculado}| > t \text{ crítica}$ , es decir  $8.4654 > 2.056$ , por lo tanto, se rechaza  $H_0$  y se acepta  $H_1$ .

Interpretación:

Al rechazar  $H_0$  y aceptar  $H_1$  se concluye que los coeficientes de  $PROD^2$  y  $PROD^3$  en la función cúbica de costo no son iguales, es decir sus elasticidades no son las mismas.

### 3.6.11 Mínimos cuadrados restringidos: pruebas de restricciones de igualdades lineales.

#### Ejercicio 11:

“Hay ocasiones en las cuales la teoría económica puede sugerir que los coeficientes en un modelo de regresión satisfacen algunas restricciones de igualdades lineales. Por ejemplo, considere la función de producción Cobb-Douglas” (Gujarati & Porter, 2010, p.248). En la función de producción Cobb-Douglas (3.6.15) si existen rendimientos constantes a escala, la teoría económica sugeriría que  $\beta_1 + \beta_2 = 1$ , lo cual es considerado como una restricción de igualdad lineal (Gujarati & Porter, 2010).

<sup>13</sup> Recuerde que la varianza es igual al cuadrado del error estándar:  $var(\hat{\beta}_3) = (ee)^2$



Para conocer si la restricción es válida (hipótesis nula,  $H_0$ ) se aplica la prueba  $F$ , la cual hace uso de los valores brindados por la estimación de la regresión no restringida (NR) y la regresión restringida (R), esta última regresión incorpora la restricción en la ecuación a estimar mediante el método de mínimos cuadrados restringidos.

La prueba  $F$  se define como:

$$F = \frac{(SCR_R - SCR_{NR})/m}{(SCR_{NR})/(n-k)} \quad (3.6.26)$$

Donde,  $SCR_R$  es la suma de cuadrados de los residuos de la regresión restringida,  $SCR_R$  es la suma de cuadrados de los residuos de la regresión no restringida,  $m$  es el número de restricciones lineales,  $k$  es el número de parámetros en la regresión no restringida, y  $n$  es el número de observaciones.

Decisión: Si el valor de la  $F$  calculada es mayor al valor de la  $F$  crítica ( $F_\alpha(k-1, n-k)$ ), a un nivel de significancia predeterminado, se rechaza la hipótesis nula, es decir que la restricción impuesta no es válida, casi contrario sí sería válida.

Comandos a utilizar:

```
Rutina utilizada:
/*Regresión no restringida*/
gen lnY=log(Y)
gen lnL=log(L)
gen lnK=log(K)
reg lnY lnL lnK
/*Regresión restringida*/
gen YparaL= Y/ L
gen KparaL= K/ L
gen lnYparaL= log(YparaL)
gen lnKparaL= log(KparaL)
reg lnYparaL lnKparaL
```

Desarrollo:

Haciendo uso de los datos de la Tabla 3.6.3, que contiene datos de las variables de producción (Y), trabajo (L) y capital (K), se trata de averiguar si la restricción de igualdad lineal  $\beta_2 + \beta_3 = 1$  establecida por el modelo de producción de Cobb-Douglas es válida o no lo es. Para ello primeramente se estima la siguiente regresión sin restricción o no restringida (NR):

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln L_{2i} + \beta_3 \ln K_{3i} + \mu_i \quad (3.6.16) = (3.6.27)$$

Con los comandos:

gen lnY=log(Y)  
gen lnL=log(L)  
gen lnK=log(K)  
reg lnY lnL lnK

Obteniendo los siguientes resultados:

Source	SS	df	MS	Number of obs = 221		
Model	382.077756	2	191.038878	F( 2, 218) = 501.45		
Residual	83.0525632	218	.380975061	Prob > F = 0.0000		
Total	465.130319	220	2.11422872	R-squared = 0.8214		
				Adj R-squared = 0.8198		
				Root MSE = .61723		

lnY	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnL	.9021417	.0614754	14.67	0.000	.7809794	1.023304
lnK	.1577797	.0405109	3.89	0.000	.0779366	.2376228
_cons	7.087062	.3798796	18.66	0.000	6.338355	7.835769

Donde la  $SCR_{NR}=83.0525632$ .

Ahora bien, para estimar la regresión con restricción o restringida se usa el método de mínimos cuadrados restringidos, el cual trata de incorporar la restricción  $\beta_2+\beta_3=1$  a la ecuación (3.6.27). De esta manera se tiene que la restricción:

$$\beta_2+\beta_3=1 \quad (3.6.28)$$

$$\beta_2 = 1 - \beta_3 \quad (3.6.29)$$

Al incorporar la restricción (3.6.29) en (3.6.27) se tiene:

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln L_{2i} + \beta_3 \ln K_{3i} + \mu_i$$

$$\ln Y_i = \ln \beta_1 + (1 - \beta_3) \ln L_{2i} + \beta_3 \ln K_{3i} + \mu_i$$

$$\ln Y_i = \ln \beta_1 + \ln L_{2i} + \beta_3 (\ln K_{3i} - \ln L_{2i}) + \mu_i$$

$$\ln Y_i - \ln L_{2i} = \ln \beta_1 + \beta_3 (\ln K_{3i} - \ln L_{2i}) + \mu_i$$

$$\ln(Y_i/L_{2i}) = \beta_0 + \beta_3 \ln(K_{3i}/L_{2i}) + \mu_i \quad (3.6.30)$$

La regresión (3.6.30) es la regresión restringida (R), por tanto para estimarla se ejecutan los siguientes comandos:

```
gen YparaL= Y/ L
gen KparaL= K/ L
gen lnYparaL= log( YparaL)
gen lnKparaL= log(KparaL)
reg lnYparaL lnKparaL
```

Y se obtienen los siguientes resultados:

Source	SS	df	MS	Number of obs =	221
Model	7.80270539	1	7.80270539	F( 1, 219) =	20.32
Residual	84.0858109	219	.383953475	Prob > F =	0.0000
Total	91.8885163	220	.417675074	R-squared =	0.0849
				Adj R-squared =	0.0807
				Root MSE =	.61964

lnYparaL	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lnKparaL	.1762048	.0390872	4.51	0.000	.0991696 .25324
_cons	7.502842	.2849497	26.33	0.000	6.941248 8.064437

Donde la  $SCR_R=84.0858109$ .

Ahora sí se puede obtener la  $F$  calculada, sabiendo que el valor de  $m$  es 1 porque hay solamente una restricción lineal, el valor de  $n$  es 221 (# de observaciones) y el valor de  $k$  es 3 (# de parámetros), se tiene:

$$F = \frac{(SCR_R - SCR_{NR})/m}{(SCR_{NR})(n-k)} \quad (3.6.26)=(3.6.31)$$

$$F = \frac{(84.0858109 - 83.0525632)/1}{(83.0525632)(221 - 3)}$$

$$F = \frac{1.0332477}{18105.45878} = 0.0000571$$

El valor de la  $F$  crítica en este caso está dada por  $F_\alpha(k-1, n-k) = F_\alpha(3-1, 221-3) = F_\alpha(2, 218)$ . Con estos resultados hay que dirigirse a la Tabla  $F$  (Anexo 4), en cual  $g/$  para el numerador es 2 y  $g/$  para el denominador es 218, como la tabla no contiene el valor de 210 se elige el valor de 200, por lo tanto se determina que el valor de la  $F$  crítica, al 5% de nivel de significancia, es 3.04.

$$F_{\text{calculada}} < F_{\text{crítica}}$$

$$0.0000571 < 3.14$$

Por lo tanto, se acepta  $H_0$ .

Interpretación:

Al aceptar  $H_0$  se verifica que la restricción de igualdad lineal  $\beta_2 + \beta_3 = 1$  es válida para el modelo de producción de Cobb-Douglas para Ecuador, ya que quizás la economía ecuatoriana se caracterizó por rendimientos constantes a escala en el año 2010, de tal manera que se puede utilizar tanto la regresión no restringida (3.6.27) como la regresión restringida (3.6.30). Interpretando la regresión (3.6.30), ante un incremento del 1% en la razón capital/trabajo, en promedio, la productividad del trabajo aumentaría casi el 1%.

### 3.6.12 Prueba de estabilidad estructural: La prueba de Chow.

#### Ejercicio 12:

A menudo al utilizar regresiones que implican series de tiempo, puede suceder el caso de que se presente un cambio estructural en la relación entre la variable  $Y$  y las regresoras  $X$ . Este cambio estructural indica que los valores de los parámetros de modelo no permanecen constantes a lo largo del tiempo o de un periodo, el mismo que puede ocurrir por fuerzas externas como: cambios en políticas, caída de precios, cambio de moneda, cambios en salarios mínimos, recesión económica, entre otras (Gujarati & Porter, 2010).

Para saber si realmente existe o no un cambio estructural se puede aplicar la prueba de Chow cuyo procedimiento es el siguiente:

1. Estimar la regresión suponiendo que no hay cambio estructural, y obtener la  $SCR_R$  (suma de los cuadrados residual restringida).
2. Estimar una regresión solamente con los datos hasta el año anterior al cambio estructural, y obtener la  $SCR_1$ .
3. Estimar una regresión con los datos a partir del año que ocurre el cambio estructural, y obtener la  $SCR_2$ .
4. Obtener la  $SCR_{NR}$  (suma de cuadrados residual no restringida) mediante la suma de  $SCR_1 + SCR_2$ .
5. Aplicar la prueba  $F$ .

$$F = \frac{(SCR_R - SCR_{NR})/k}{(SCR_{NR})/(n_1 + n_2 - 2k)} \quad (3.6.31)$$

En este caso la  $H_0$  es que no hay cambio estructural (hay estabilidad paramétrica).

6. La  $F$  crítica o de la tabla (Anexo 4) viene dada así:  $F[k, (n_1+n_2-2k)]$ . Se compara los valores de  $F$  calculado y  $F$  de la tabla.

Decisión: si el valor  $F$  calculado es mayor al  $F$  crítico en el nivel de significancia escogido, se rechaza  $H_0$ , por lo tanto si existe un cambio estructural.

Comandos a utilizar:

**in:** Va al final de un comando y especifica el rango de observaciones que el usuario quiere utilizar. Se permite en la mayoría de los comandos de Stata.

Ejemplo:

**reg** var1 var2...var\_n **in** #obs\_inicial/#obs\_final

Rutina utilizada:

**gen** lnPIB=log(PIB)

**reg** lnPIB DES

**reg** lnPIB DES **in** 1/20

**reg** lnPIB DES **in** 21/35

Desarrollo:

Para ejemplificar la prueba de Chow se utilizarán los datos de la Tabla 3.6.5, la misma que contiene datos del Pib y del desempleo. Se va a comprobar si hay o no un cambio estructural en la relación logaritmo del Pib sobre el desempleo, tomando en cuenta que el año 2000 se produjo la dolarización en el Ecuador. El conjunto de hipótesis es la siguiente:

$H_0$ : No existe cambio estructural (estabilidad paramétrica)

$H_1$ : Existe cambio estructural

Primeramente es necesario crear la variable logarítmica del Pib mediante el comando: **gen** lnPIB=log(PIB). Ahora si se procede a realizar la prueba de Chow paso a paso.

Paso 1: Se estima la regresión (período: 1980-2014):

$$\ln PIB = \alpha_1 + \alpha_2 DES_t + \mu_t \quad (3.6.32)$$
$$n = 35$$

Se ejecuta el comando: **reg** lnPIB DES, y se obtienen los siguientes resultados:

Source	SS	df	MS	Number of obs = 35		
Model	.205711185	1	.205711185	F( 1, 33) = 2.21		
Residual	3.07799897	33	.093272696	Prob > F = 0.1470		
Total	3.28371015	34	.09657971	R-squared = 0.0626		
				Adj R-squared = 0.0342		
				Root MSE = .30541		

lnPIB	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
DES	-.026881	.0181007	-1.49	0.147	-.0637071	.009945
_cons	10.61007	.1459643	72.69	0.000	10.3131	10.90704

Donde la  $SCR_R=3.07799897$ .

Paso 2: Se estima la regresión (período: 1980-1999):

$$\ln PIB = \lambda_1 + \lambda_2 DES_t + \mu_{1t} \quad (3.6.33)$$

$$n = 20$$

Con el comando: **reg lnPIB DES in 1/20**. Se obtiene:

Source	SS	df	MS	Number of obs = 20		
Model	.04895864	1	.04895864	F( 1, 18) = 2.11		
Residual	.417885055	18	.023215836	Prob > F = 0.1637		
Total	.466843695	19	.024570721	R-squared = 0.1049		
				Adj R-squared = 0.0551		
				Root MSE = .15237		

lnPIB	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
DES	.0205263	.0141348	1.45	0.164	-.0091697	.0502224
_cons	10.02859	.1140459	87.93	0.000	9.788987	10.26819

Donde la  $SCR_1=0.417885055$ .

Paso 3: Se estima la regresión (período: 2000 -2014):

$$\ln PIB = \gamma_1 + \gamma_2 DES_t + \mu_{2t} \quad (3.6.34)$$

$$n = 15$$

Con el comando: **reg lnPIB DES in 21/35**. Se obtiene:

Source	SS	df	MS	Number of obs =	15
Model	.423640485	1	.423640485	F( 1, 13) =	45.59
Residual	.120801958	13	.009292458	Prob > F =	0.0000
Total	.544442444	14	.038888746	R-squared =	0.7781
				Adj R-squared =	0.7611
				Root MSE =	.0964

lnPIB	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
DES	-.0503162	.007452	-6.75	0.000	-.0664153    -.034217
_cons	11.07052	.0600493	184.36	0.000	10.94079    11.20025

Donde la  $SCR_2=0.120801958$ .

Paso 4:  $SCR_{NR} = SCR_1 + SCR_2 = 0.417885055 + 0.120801958 = 0.545904635$

Paso 5: Aplicar la prueba F.

$$F = \frac{(SCR_R - SCR_{NR})/k}{(SCR_{NR})/(n_1 + n_2 - 2k)} \quad (3.6.31) = (3.6.35)$$

$$F = \frac{(3.07799897 - 0.545904635)/2}{(0.545904635)/(20 + 15 - 2(2))}$$

$$F = \frac{1.266047168}{0.017609826} = 71.89$$

Paso 6:  $F[k, (n_1+n_2-2k)] = F[2, (20+15-2(2))] = F(2, 31)$ . Con estos resultados hay que dirigirse a la Tabla  $F$  (Anexo 4), en cual  $g/$  para el numerador es 2 y  $g/$  para el denominador es 31, como la tabla no contiene el valor de 31 se elige el valor de 30, por lo tanto se determina que el valor de la  $F$  crítica, al 5% de nivel de significancia, es 3.32.

$$F_{calculada} > F_{crítica}$$

$$71.89 > 3.32$$

Por lo tanto, se rechaza  $H_0$ .

Interpretación:

Al rechazar  $H_0$  se afirma que sí existe un cambio estructural en la relación del logaritmo del Pib con el desempleo para Ecuador en el período 1980-2014, debido al fenómeno de la dolarización en el año 2000, el cual afectó fuertemente al país.

### 3.6.13 Prueba de la forma funcional de la regresión: elección entre modelos de regresión lineal y log-lineal.

#### Ejercicio 13:

Para elegir entre un modelo de regresión lineal y un modelo de regresión log-lineal se puede hacer uso de la prueba MWD, la misma que formula las siguientes hipótesis:

$H_0$ : Modelo lineal

$H_1$ : Modelo log-lineal

Como lo menciona Gujarati & Porter (2010), la prueba MWD comprende los siguientes pasos:

1. Estimar el modelo lineal y obtener los valores  $Y$  estimados ( $Yf$ ).
2. Estimar el modelo log-lineal y obtener los valores  $\ln Y$  estimados ( $\ln f$ ).
3. Obtener  $Z1 = (\ln Yf - \ln f)$
4. Estimar la regresión de  $Y$  sobre las  $X$  y  $Z1$ . Rechazar  $H_0$  si el coeficiente de  $Z1$  es estadísticamente significativo.
5. Obtener  $Z2 = (\text{antilog} \ln f - Yf)$
6. Estimar la regresión de  $\ln Y$  sobre las  $\ln X$  y  $Z1$ . Rechazar  $H_1$  si el coeficiente de  $Z2$  es estadísticamente significativo.

Comandos a utilizar:

```
Rutina utilizada:  
reg MIG DES  
predict Yf  
gen lnMIG=log(MIG)  
gen lnDES=log(DES)  
reg lnMIG lnDES  
predict lnf  
gen lnYf=log(Yf)  
gen Z1= lnYf- lnf  
reg MIG DES Z1  
gen antiloglnf=exp(lnf)  
gen Z2= antiloglnf-Yf  
reg lnMIG lnDES Z2
```

Desarrollo:

Para ejemplificar la prueba MWD se va a usar los datos de la Tabla 3.6.6, que contiene datos de la migración neta y el número de desempleados. Por consiguiente se va a elegir entre el modelo lineal y log lineal, por lo tanto se tienen las siguientes hipótesis:



$$H_0: \text{Modelo lineal:} \quad MIG_t = \alpha_1 + \alpha_2 DES_t + \mu_t \quad (3.6.36)$$

$$H_1: \text{Modelo log-lineal:} \quad \ln MIG_t = \beta_1 + \beta_2 \ln DES_t + \mu_t \quad (3.6.37)$$

Ahora bien, se prosigue a realizar los pasos que implica la prueba MWD. Primero, para estimar el modelo lineal y obtener los valores MIG estimados (Yf), se ejecuta el comando: **reg MIG DES**

$$MIG_t = -37.45601 + 0.2126583 DES_t + \mu_t \quad (3.6.38)$$

$$R^2 = 0.6973$$

Y se predicen los valores de MIG estimados con el comando: **predict Yf**.

Segundo, se estima el modelo log-lineal y se obtienen los valores lnY estimados (lnf), para ello es necesario generar previamente las variables logarítmicas de las variables migración (MIG) y desempleo (DES), por lo tanto se ejecutan los siguientes comandos:

**gen lnMIG=log(MIG)**

**gen lnDES=log(DES)**

**reg lnMIG lnDES**

$$\ln MIG_t = -4.779291 + 1.411355 \ln DES_t + \mu_t \quad (3.6.39)$$

$$R^2 = 0.7023$$

Y se predicen los valores de MIG estimados con el comando: **predict lnf**.

Tercero, se genera la variable Z1= (lnYf - lnf), con el comando: **gen Z1= lnYf- lnf**.

Cuarto, se estima la regresión de MIG sobre DES y Z1, obteniendo lo siguientes resultados:

Source	SS	df	MS	Number of obs = 31		
Model	42749.6925	2	21374.8462	F( 2, 28) = 37.17		
Residual	16103.7389	28	575.133531	Prob > F = 0.0000		
Total	58853.4313	30	1961.78104	R-squared = 0.7264		
				Adj R-squared = 0.7068		
				Root MSE = 23.982		

MIG	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
DES	.2682755	.0355508	7.55	0.000	.195453	.3410981
Z1	-32.71491	22.79262	-1.44	0.162	-79.40347	13.97365
_cons	-60.15018	14.58303	-4.12	0.000	-90.02217	-30.2782

De estos resultados se observa que la variable Z1 es no significativa a un nivel de confianza del 95% ( $|t| < 2$  y  $p > 5\%$ ), por lo tanto se acepta  $H_0$ .

Quinto, se genera la variable  $Z2 = (\text{antilog} \ln f - Yf)$ , para ello es necesario primeramente generar la variable antilogarítmica de  $\ln f$  haciendo uso del comando **exp**, por consiguiente se ejecutan los siguientes comandos:

**gen antiloglnf=exp(lnf)**

**gen Z2= antiloglnf- Yf**

Sexto, se estima la regresión de  $\ln \text{MIG}$  sobre  $\ln \text{DES}$  y  $Z2$ , obteniendo los siguientes resultados:

Source	SS	df	MS	Number of obs = 34		
Model	15.1881572	2	7.59407858	F( 2, 31) = 46.26		
Residual	5.08854237	31	.164146528	Prob > F = 0.0000		
				R-squared = 0.7490		
				Adj R-squared = 0.7329		
Total	20.2766995	33	.61444544	Root MSE = .40515		

$\ln \text{MIG}$	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
$\ln \text{DES}$	-1.963964	1.412658	-1.39	0.174	-4.8451	.9171721
$Z2$	-.1322671	.0550378	-2.40	0.022	-.2445175	-.0200167
_cons	14.31842	7.997007	1.79	0.083	-1.991584	30.62842

De estos resultados se observa que la variable  $Z2$  es significativa a un nivel de confianza del 95% ( $|t| > 2$  y  $p < 5\%$ ), por lo tanto se rechaza  $H_1$ .

Interpretación:

Al aceptar  $H_0$  y rechazar  $H_1$  se elige el modelo lineal (3.6.36), siendo éste el más apropiado. De acuerdo con los resultados del modelo lineal en (3.6.38) se dice que; ante el aumento de 1000 personas desempleadas, la cantidad de migrantes aumentará, en promedio, 268 personas.

### 3.7 Modelos de regresión con variables dicótomas

Los modelos de regresión con variables dicótomas son aquellos que incluyen entre sus regresoras a variables cualitativas. Estas variables indican la presencia o ausencia de una cualidad o atributo. Para cuantificar estos atributos se utilizan los valores 0 y 1, donde 1 significa presencia del atributo y 0 su ausencia. Las variables que toman los valores de 0 y 1 se las denomina variables dicótomas (Gujarati & Porter, 2010).

Dentro de los modelos que utilizan variables dicótomas inicialmente se hace hincapié a los modelos ANOVA, los cuales tienen como regresoras únicamente variables de naturaleza cualitativa, luego se analizan los modelos ANCOVA, que son aquellos que presentan una

mezcla de variables cualitativas y cuantitativas entre sus regresoras, posteriormente se aplica la variable dicótoma como alternativa para la prueba de Chow, se continúa explicando los efectos de interacción al utilizar variables dicótomas, el uso de variables dicótomas en el análisis estacional, se analiza una regresión lineal por segmentos, y se finaliza con la aplicación de variables dicótomas en regresiones semilogarítmicas. Cabe mencionar que para la resolución de la mayoría de los ejercicios (6 de 8 ejercicios) se han utilizado datos del salario mensual de trabajadores y las diversas variables regresoras que afectan su valor como sexo, estado civil, nivel de instrucción y escolaridad. Los 2 ejercicios restantes utilizan datos del pib, desempleo, y consumo de hogares. Además los datos utilizados están contenidos en siete tablas (Tabla 3.7.1 – Tabla 3.7.7), los mismos que son extraídos de fuentes oficiales como INEC (2010), el Banco Mundial (2015) y el BCE (2015).

### 3.7.1 Modelos ANOVA

#### Ejercicio 1:

Los modelos ANOVA son aquellos en que la variable dependiente,  $Y$ , es de naturaleza cuantitativa, y las variables independientes,  $X$ , son de naturaleza cualitativa. Es importante recordar que “si una variable cualitativa tiene  $m$  categorías, sólo hay que agregar  $(m - 1)$  variables dicótomas” (Gujarati & Porter, 2010, p.281). Por ejemplo, si una variable cualitativa tiene tres categorías, deben agregarse solamente dos variables dicotómicas, ya que si se añadieran tres (una para cada categoría) se caería en la *trampa de la variable dicótoma*. Por lo tanto este tipo de modelo se escribiría así:

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \mu_i \quad (3.7.1)$$

En (3.7.1) la categoría a la que no se añade una variable dicótoma sería la categoría base o de comparación, cuyo valor medio es igual a valor de intercepto  $\beta_1$ . La elección de la categoría base se deja al criterio del investigador. Los coeficientes asociados a las variables dicótomas  $\beta_2$  y  $\beta_3$  se conocen como coeficientes de intercepto diferencial, los cuáles indican en cuánto difieren del valor de la categoría base  $\beta_1$ .

Comandos a utilizar:

**i.**: Este comando permite desagrupar variables cualitativas de acuerdo a su categoría. Las categorías van en formato numérico.

Su estructura es:

**i.varcualitativa**

(para una variable)

**i.(varcualitativa1 varcualitativa2... varcualitativan)**

(para más de una variable)

Rutina utilizada:

**reg W SEC SUP POST**

**reg W i.NIVELESC**

Desarrollo:

Para ejemplificar un modelo ANOVA se van a utilizar los datos de la Tabla 3.7.1. Por consiguiente, se tiene como variable cuantitativa al salario mensual (Y) y como variable cualitativa al nivel de escolaridad, la misma que cuenta con cuatro categorías: primaria, secundaria, superior y post-grado.

Como la variable cualitativa tiene cuatro categorías se agregan tres variables dicótomas, en donde se ha elegido a la categoría “primaria” como categoría de comparación. Por lo tanto el modelo ANOVA a estimar sería el siguiente:

$$W_i = \beta_1 + \beta_2 SEC_i + \beta_3 SUP_i + \beta_4 POST_i + \mu_i \quad (3.7.2)$$

Donde, W = salario mensual (\$) de trabajadores de la ciudad de Quito

SEC= 1 secundaria y 0 los demás casos

SUP= 1 superior y 0 los demás casos

POST= 1 post-grado y 0 los demás casos

Para estimar el modelo (3.7.2) se ejecuta el comando: **reg W SEC SUP POST**, obteniendo los siguientes resultados:

Source	SS	df	MS	Number of obs = 1322		
Model	102891981	3	34297327.1	F( 3, 1318) = 131.05		
Residual	344937412	1318	261712.755	Prob > F = 0.0000		
Total	447829393	1321	339007.867	R-squared = 0.2298		
				Adj R-squared = 0.2280		
				Root MSE = 511.58		

W	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
SEC	109.8775	41.94346	2.62	0.009	27.59427	192.1607
SUP	506.5755	41.10314	12.32	0.000	425.9408	587.2102
POST	1144.831	75.22477	15.22	0.000	997.2576	1292.404
_cons	366.8131	34.9708	10.49	0.000	298.2086	435.4176

$$\hat{W}_i = 366.8131 + 109.8775 SEC_i + 506.5755 SUP_i + 1144.831 POST_i + \mu_i \quad (3.7.3)$$

$$R^2 = 0.2298$$

Por lo tanto se tiene que:

- Salario medio – primaria ( $\beta_1$ ) = \$366.81
- Salario medio – secundaria ( $\beta_1 + \beta_2$ ) = \$366.81 + \$109.88 = \$476.69
- Salario medio – superior: ( $\beta_1 + \beta_3$ ) = \$366.81 + \$506.58 = \$873.39
- Salario medio – post-grado: ( $\beta_1 + \beta_4$ ) = \$366.81 + \$1 144.83 = \$1 511.64

Una manera alterna de estimar la regresión (3.7.2) es mediante el uso del comando **i.** por lo tanto se ejecuta el comando: **reg W i. NIVELESC**. La variable cualitativa NIVELESC tiene cuatro categorías: 1=primaria, 2=secundaria, 3=superior, y 4=post-grado. Los resultados obtenidos son los siguientes:

Source	SS	df	MS	Number of obs = 1322		
Model	102891981	3	34297327.1	F( 3, 1318) = 131.05		
Residual	344937412	1318	261712.755	Prob > F = 0.0000		
				R-squared = 0.2298		
				Adj R-squared = 0.2280		
				Root MSE = 511.58		
Total	447829393	1321	339007.867			

W	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
NIVELESC						
2	109.8775	41.94346	2.62	0.009	27.59427	192.1607
3	506.5755	41.10314	12.32	0.000	425.9408	587.2102
4	1144.831	75.22477	15.22	0.000	997.2576	1292.404
_cons	366.8131	34.9708	10.49	0.000	298.2086	435.4176

Estos resultados son los mismos que los presentados en la regresión (3.7.3). Queda a elección del usuario utilizar el comando **i.** o correr la regresión de la forma (3.7.1). En los ejercicios posteriores de este apartado se estimarán las regresiones de la forma (3.7.1).

Interpretación:

Como muestran los resultados de la regresión (3.7.3) el salario medio de los trabajadores que tienen únicamente educación primaria es de casi \$366.81, el salario de los trabajadores que tienen educación secundaria es mayor con \$109.88 aproximadamente (igual a \$476.69), el salario de los trabajadores que tienen educación superior es mayor por cerca de \$506.58 (igual a \$873.39), y el salario de los trabajadores que tienen educación de post-grado es mayor por casi \$1 144.83 (igual a \$1 511.64). Además al observar la significancia de las tres variables dicótomas se afirma que, los salarios medios de los trabajadores con nivel de instrucción secundaria, superior y post-grado son estadísticamente diferentes del nivel de salario medio de los trabajadores con educación primaria.

### 3.7.2 Modelos ANOVA con dos variables.

#### Ejercicio 2:

En el ejemplo 1 se consideró un modelo ANOVA con solamente una variable cualitativa como regresora (nivel de instrucción), en este caso se desarrollará un modelo ANOVA con dos variables cualitativas. Por ejemplo, si se consideran dos variables cualitativas, cada una con dos categorías el modelo quedaría expresado de la siguiente manera:

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \mu_i \quad (3.7.4)$$

La diferencia del modelo (3.7.4) con (3.7.1) es que el modelo (3.7.4) presenta dos variables cualitativas con dos categorías cada uno, por ello se agrega solamente una variable dicótoma por cada variable cualitativa, mientras que el modelo (3.7.1) presenta una sola variable cualitativa pero con tres categorías, por ello presenta dos variables dicótomas.

Comandos a utilizar:

Rutina utilizada:  
**reg W SEXO ESTCIVIL**

Desarrollo:

Utilizando los datos de la Tabla 3.7.2 se va a estimar un modelo ANOVA con dos variables dicótomas. Se tiene como variable cuantitativa al salario mensual ( $Y$ ) y como variable cualitativa al sexo y al estado civil, cada una con dos categorías. Como cada variable cualitativa tiene dos categorías se agrega una sola variable dicótoma de cada una, así:

$$W_i = \beta_1 + \beta_2 SEXO_i + \beta_3 ESTCIVIL_i + \mu_i \quad (3.7.5)$$

Donde,  $W$  = salario mensual (\$) de trabajadores de la ciudad de Quito

SEXO= 1 hombre y 0 mujer

ESTCIVIL= 1 casado(a) y 0 los demás casos

En este caso la categoría en comparación serían las mujeres que no son casadas (aquellos con valor 0). Para estimar el modelo (3.7.5) se ejecuta el comando: **reg W SEXO ESTCIVIL**, obteniendo los siguientes resultados:

Source	SS	df	MS	Number of obs = 1369		
Model	15090221.6	2	7545110.8	F( 2, 1366) = 23.41		
Residual	440277745	1366	322311.673	Prob > F = 0.0000		
Total	455367967	1368	332871.321	R-squared = 0.0331		
				Adj R-squared = 0.0317		
				Root MSE = 567.72		

W	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
SEXO	87.09054	31.10515	2.80	0.005	26.0715	148.1096
ESTCIVIL	181.9194	31.02772	5.86	0.000	121.0523	242.7866
_cons	529.2436	26.00768	20.35	0.000	478.2243	580.2629

$$\hat{W}_i = 529.2436 + 87.09054 SEXO_i + 181.9194 ESTCIVIL_i + \mu_i \quad (3.7.6)$$

$$R^2 = 0.0317$$

Por lo tanto se tiene que:

- Salario medio – mujer y no casada ( $\beta_1$ ) = \$529.24
- Salario medio – hombre ( $\beta_1 + \beta_2$ ) = \$529.24 + \$87.09 = \$616.33
- Salario medio – casados(as): ( $\beta_1 + \beta_3$ ) = \$529.24 + \$181.92 = \$711.16

Interpretación:

Los resultados de la regresión (3.7.5) muestran que el salario medio de las mujeres no casadas es de aproximadamente \$529.24, el salario medio de los hombres es mayor con \$87.09 (igual a \$616.33), y el salario medio de los casados mayor con casi \$181.92 (igual a \$711.16). Como las variables: sexo y estado civil son significativas se puede concluir que el salario medio de los trabajadores hombres y trabajadores casados es estadísticamente diferente al salario medio de las trabajadoras mujeres no casadas.

### 3.7.3 Modelos ANCOVA.

#### Ejercicio 3:

Los modelos ANCOVA son aquellos que contienen una mezcla de variables cualitativas y cuantitativas entre sus regresoras, y la variable  $Y$  sigue siendo cuantitativa. Por ejemplo, si se tiene un modelo con una variable cualitativa (con tres categorías) y una variable cuantitativa, entre sus variables regresoras, el modelo se expresaría de la siguiente manera:

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 X_i + \mu_i \quad (3.7.7)$$

Comandos a utilizar:

Rutina utilizada:  
**reg VAB PRINC EMP**

Desarrollo:

Se va a estimar un modelo ANCOVA con los datos de la Tabla 3.7.3. Este modelo contiene como regresoras a dos variables cualitativas (sexo y estado civil) con dos categorías cada una y a una variable cuantitativa (escolaridad) como muestra la siguiente regresión:

$$W_i = \alpha_1 + \alpha_2 SEXO_i + \alpha_3 ESTCIVIL_i + \beta_1 ESC_i + \mu_i \quad (3.7.8)$$

Donde,  $W$  = salario mensual (\$) de trabajadores de la ciudad de Quito

SEXO = 1 para hombre y 0 para mujer

ESTCIVIL = 1 para casados(as) y 0 para los demás casos.

ESC = escolaridad (años de estudio aprobados).

En la regresión (3.7.8) se muestra que el salario mensual está en función de la variable cualitativa sexo y estado civil y de la variable cuantitativa escolaridad. Para estimar dicha regresión se ejecutan el comando: **reg W SEXO ESTCIVIL ESC**. Obteniendo los siguientes resultados:

Source	SS	df	MS	Number of obs = 1322		
Model	116682077	3	38894025.7	F( 3, 1318) = 154.80		
Residual	331147316	1318	251249.86	Prob > F = 0.0000		
				R-squared = 0.2606		
				Adj R-squared = 0.2589		
Total	447829393	1321	339007.867	Root MSE = 501.25		

W	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
SEXO	130.0432	28.02964	4.64	0.000	75.05566	185.0308
ESTCIVIL	184.9491	27.88675	6.63	0.000	130.2418	239.6564
ESC	63.29263	3.138927	20.16	0.000	57.13479	69.45047
_cons	-332.9409	49.19286	-6.77	0.000	-429.4457	-236.436

$$\hat{W}_i = -332.94 + 130.04SEXO_i + 184.95ESTCIVIL_i + 63.29ESC_i + \mu_t \quad (3.7.9)$$

$$R^2 = 0.26$$

Interpretación:

Manteniendo lo demás constante (*ceteris paribus*), si el nivel de escolaridad aumenta un año, el salario mensual de los trabajadores quiteños aumentará en promedio \$63.29. Nótese que los resultados difieren de los de (3.7.6) ya que no se tomó en cuenta la covariante que son las diferencias del grado de escolaridad de los trabajadores quiteños.

### 3.7.4 La variable dicótoma alternativa a la prueba de Chow.

#### Ejercicio 4:

En el ejercicio 12 del apartado 3.6.12 se vio la aplicación de la prueba de Chow utilizando la prueba  $F$  para averiguar si existía o no un cambio estructural en la relación entre el pib y el desempleo, es decir si existía un diferencia en la regresión del pib sobre el desempleo en los períodos: 1980-1999 y 2000-2014. Sin embargo, “no se pudo determinar si esta diferencia entre las dos regresiones se debía a las diferencias en los términos del intercepto o en los coeficientes de la pendiente, o a ambas situaciones” (Gujarati & Porter, 2010, p.285).

Según Gujarati & Porter (2010) hay cuatro posibilidades:



- Regresiones coincidentes: mismo intercepto y coeficientes de pendiente.
- Regresiones paralelas: distinto intercepto y misma pendiente.
- Regresiones concurrentes: mismo intercepto y distinta pendiente.
- Regresiones disímiles: distinto intercepto y pendiente.

Para averiguar qué tipo de regresión es, es necesario estimar un modelo de forma:

$$Y_t = \alpha_1 + \alpha_2(D_t) + \beta_1(X_t) + \beta_2(D_t)(X_t) + \mu_t \quad (3.7.10)$$

Decisión: Si el coeficiente del intercepto diferencial  $\alpha_2$  es estadísticamente significativo las dos regresiones tienen diferente intercepto. De modo semejante, si el coeficiente de la pendiente diferencial  $\beta_2$  es estadísticamente significativo las dos regresiones tienen diferente pendiente (Gujarati & Porter, 2010).

Comandos a utilizar:

Rutina utilizada:  
**gen DDES= D\*DES**  
**reg PIB D DES DDES**  
**twoway (lfit PIB DES in 1/20) (lfit PIB DES in 21/35)**

Desarrollo:

Se va a probar cómo actúa la técnica de la variable dicótoma como alternativa a la prueba de Chow del ejercicio 12 del apartado 3.6.12. Por lo tanto, utilizando los mismos datos del PIB y el desempleo (Tabla 3.7.4) se aplicará este método. Para ello es necesario agregar una variable dicótoma  $D$ , la cual será igual a 0 en los años anteriores al 2000 (1980-1999), e igual a 1 a partir del año 2000 (2000-2014).

Ahora bien, se procede a estimar la siguiente regresión:

$$PIB_t = \alpha_1 + \alpha_2 D_t + \beta_1 DES_t + \beta_2 (D_t DES_t) + \mu_t \quad (3.7.11)$$

Para estimar la regresión (3.7.11) es necesario primeramente que se genere la variable DDES, la cual es igual a la multiplicación de la variable desempleo (DES) con la variable dicótoma (D), y luego sí se estima (3.7.11), ejecutando los siguientes comandos:

**gen DDES= D\*DES**  
**reg PIB D DES DDES**

Los resultados de la regresión estimada son los siguientes:

Source	SS	df	MS	Number of obs =	35
Model	3.7491e+09	3	1.2497e+09	F( 3, 31) =	65.52
Residual	591248864	31	19072544	Prob > F =	0.0000
				R-squared =	0.8638
				Adj R-squared =	0.8506
Total	4.3403e+09	34	127656563	Root MSE =	4367.2

PIB	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
D	39675.21	4419.131	8.98	0.000	30662.34	48688.09
DES	562.6258	405.5915	1.39	0.175	-264.5835	1389.835
DDES	-2773.979	527.8216	-5.26	0.000	-3850.478	-1697.479
_cons	22307.31	3422.496	6.52	0.000	15327.08	29287.54

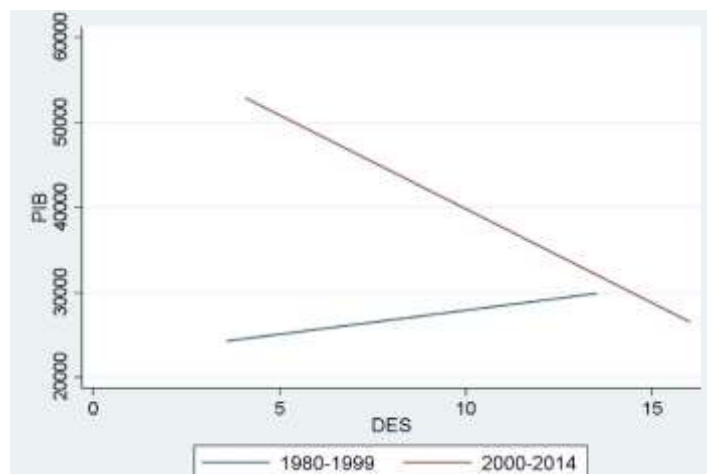
$$\widehat{PIB}_t = 22\,307.31 + 39\,375.21D_t + 562.6258DES_t - 2\,773.979(D_tDES_t) + \mu_t \quad (3.7.12)$$

$ee$  (3 422.49) (4 419.13) (405.59) (527.82)  
 $t$  (6.52) (8.98) (1.39) (-5.26)

$R^2 = 0.86$

Los resultados de (3.7.12) muestran que el coeficiente del intercepto diferencial  $\alpha_2 = 8.98$  es estadísticamente significativo, por lo tanto las dos regresiones (1980-1999 y 2000-2014) tienen diferente intercepto. Del mismo modo, el coeficiente de la pendiente diferencial  $\beta_2$  es estadísticamente significativo, por lo tanto las dos regresiones tienen diferente pendiente. Como consecuencia las dos líneas de regresiones son disímboles.

Para comprobar que las dos líneas de regresión son disímboles se van a graficar estas líneas con el comando: **twoway (lfit PIB DES in 1/20) (lfit PIB DES in 21/35)**, y se obtiene:



En el gráfico se puede observar claramente que las dos líneas no parten desde un mismo punto, por lo tanto las dos regresiones no tienen el mismo intercepto. En el mismo sentido, se puede observar que su pendiente tampoco es la misma, ya que las líneas tienden a cruzarse entre ellas.

Interpretación:

La técnica de la variable dicótoma ha permitido determinar que la diferencia entre las dos regresiones (1980-1999 y 2000-2014) o cambio estructural del pib sobre el desempleo se debe a la diferencia tanto del intercepto como de la pendiente, por lo tanto son regresiones disímboles.

### 3.7.5 Efectos de interacción al utilizar variables dicótomas.

#### Ejercicio 5:

El efecto de interacción de utilizar variables dicótomas puede ser aditivo o multiplicativo. Para analizar el efecto aditivo se puede considerar el siguiente modelo:

$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta X_i + \mu_i \quad (3.7.13)$$

El modelo (3.7.13) incluye a dos variables cualitativas con dos categorías cada una. Por consiguiente, el efecto diferencial de la variable dicótoma sexo  $D_2$  es constante en las dos categorías de  $D_3$ , y el efecto diferencial de la variable dicótoma  $D_3$  también es constante en las dos categorías de  $D_2$  (Gujarati & Porter, 2010).

Por otro lado, para analizar el efecto multiplicativo se puede considerar el siguiente modelo:

$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 (D_{2i} D_{3i}) + \beta X_i + \mu_i \quad (3.7.14)$$

La diferencia del modelo (3.7.13) con el (3.7.14) es que el segundo añade el efecto diferencial de las dos variables dicótomas al mismo tiempo, a esta variable se la denomina “variable dicótoma de interacción”.

Comandos a utilizar:

```
Rutina utilizada:  
reg W SEXO ESTCIVIL ESC  
gen SEXOESTCIVIL= SEXO*ESTCIVIL  
reg W SEXO ESTCIVIL SEXOESTCIVIL ESC
```

Desarrollo:

Se van a utilizar los datos de la Tabla 3.7.3 para estimar dos modelos: uno de forma (3.7.13) que muestra el efecto aditivo de las variables dicótomas y otro de forma (3.7.14) que muestra el efecto multiplicativo de las variables dicótomas. Por lo tanto, primeramente se considera el siguiente modelo:

$$W_i = \alpha_1 + \alpha_2 SEXO_i + \alpha_3 ESTCIVIL_i + \beta_1 ESC_i + \mu_i \quad (3.7.15)$$

Donde, W= salario mensual (\$) de trabajadores de la ciudad de Quito

SEXO= 1 para hombre y 0 para mujer

ESTCIVIL= 1 para casados(as) y 0 para los demás casos.

ESC= escolaridad (años de estudio aprobados).

Para estimar el modelo (3.7.15) se ejecuta el comando: **reg W SEXO ESTCIVIL ESC**.  
Obteniendo los resultados ya presentados en (3.7.9).

$$\hat{W}_i = -332.94 + 130.04SEXO_i + 184.95ESTCIVIL_i + 63.29ESC_i + \mu_t \quad (3.7.9) = (3.7.16)$$

$$R^2 = 0.26$$

Ahora, se considera el modelo que incluye la variable dicótoma de interacción, así:

$$W_i = \alpha_1 + \alpha_2 SEXO_i + \alpha_3 ESTCIVIL_i + \alpha_4 SEXOESTCIVIL_i + \beta_1 ESC_i + \mu_t \quad (3.7.17)$$

$\alpha_2$ = efecto diferencial de ser hombre

$\alpha_3$ = efecto diferencial de ser casado (a)

$\alpha_4$ = efecto diferencial de ser hombre y ser casado.

Para estimar el modelo (3.7.17) se ejecutan los comandos:

**gen SEXOESTCIVIL= SEXO\*ESTCIVIL**

**reg W SEXO ESTCIVIL SEXOESTCIVIL ESC**

Obteniendo los siguientes resultados:

Source	SS	df	MS	Number of obs = 1322		
Model	118685578	4	29671394.4	F( 4, 1317) = 118.72		
Residual	329143815	1317	249919.374	Prob > F = 0.0000		
				R-squared = 0.2650		
				Adj R-squared = 0.2628		
Total	447829393	1321	339007.867	Root MSE = 499.92		

W	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
SEXO	59.11297	37.5378	1.57	0.116	-14.52745	132.7534
ESTCIVIL	95.26205	42.1538	2.26	0.024	12.56612	177.958
SEXOESTCIVIL	158.9374	56.1347	2.83	0.005	48.81424	269.0606
ESC	63.64495	3.133077	20.31	0.000	57.49859	69.79132
_cons	-302.7316	50.20918	-6.03	0.000	-401.2303	-204.2329

$$W_i = -302.73 + 59.11SEXO_i + 95.26ESTCIVIL_i + 158.94SEXOESTCIVIL_i + 63.64ESC_i + \mu_i \quad (3.7.18)$$

$$R^2 = 0.26$$

En este caso las variables dicótomas aditivas son SEXO  $D_2$  y ESTCIVIL  $D_3$ , y la variable dicótoma interactiva o multiplicativa es SEXOESTCIVIL.

Interpretación:

Al tener a las variables dicótomas aditivas significativa y a la variable dicótoma de interacción SEXOESTCIVIL también significativa, la interpretación es la siguiente: si se mantiene constante la escolaridad ESC y se suman los tres coeficientes de las variables dicótomas, se obtiene 313.31 (59.11+95.26+158.94), lo cual significa que los salarios mensuales de los trabajadores hombres y casados es superior por casi \$313.31.

### 3.7.6 Uso de variables dicótomas en el análisis estacional.

#### Ejercicio 6:

Frecuentemente las series de tiempo económicas, que se basan en datos mensuales o trimestrales, presentan estacionalidad, por ejemplo, las ventas en navidad, el gasto de una familiar en temporada de vacaciones, los precios del cultivo después de la cosecha, etc. (Gujarati & Porter, 2010). Para desestacionalizar o eliminar el factor estacional de una serie de tiempo existen diversos métodos, sin embargo se va a hacer hincapié al método de las variables dicótomas.

Comandos a utilizar:

**mean:** Obtiene el valor promedio de los datos de una variable determinada.

Su estructura es:

**mean** (variable)

Rutina utilizada:

*/\*Para averiguar si existe estacionalidad\*/*

**reg** CONH TRIM1 TRIM2 TRIM3 TRIM4, **noconst**

**reg** CONH TRIM2 TRIM3 TRIM4

*/\*Para desestacionalizar la serie de tiempo\*/*

**reg** CONH TRIM1 TRIM2 TRIM3 TRIM4, **noconst**

**predict** r, **resid**

**mean**(CONH)

**gen** Ynueva= r+9248.718

Desarrollo:

Se pretende hacer un análisis estacional utilizando el método de las variables dicótomas para los datos trimestrales del consumo de hogares en el Ecuador en el período 2000-I – 2015-IV (Tabla 3.7.5). Para ello se ejecutarán dos regresiones, una que no incluya el término del intercepto o constante y otra que sí lo incluya, sin caer en ambos casos en la trampa de la variable dicótoma.

A priori se esperaría que el consumo de los hogares sea estacional en el cuarto trimestre de cada año, ya que por lo general en época navideña hay mayor circulación de dinero debido a un mayor consumo, sin embargo se va a averiguar si efectivamente así o no lo es.

Primeramente se estima la siguiente regresión:

$$CONH_t = \alpha_1 TRIM1_t + \alpha_2 TRIM2_t + \alpha_3 TRIM3_t + \alpha_4 TRIM4_t + \mu_t \quad (3.7.19)$$

Donde, cada trimestre es una variable dicótoma. Nótese además que se utiliza una variable dicótoma por cada trimestre (por cada categoría) pero para no caer en la trampa de la variable dicótoma se ha omitido el término del intercepto, lo cual es correcto. Por lo tanto, para estimar la regresión (3.7.19) se ejecuta el comando: **reg CONH TRIM1 TRIM2 TRIM3 TRIM4, noconst**

Se obtienen los siguientes resultados:

Source	SS	df	MS	Number of obs =	64
Model	5.4775e+09	4	1.3694e+09	F( 4, 60) =	90.31
Residual	909817467	60	15163624.5	Prob > F =	0.0000
				R-squared =	0.8576
				Adj R-squared =	0.8481
Total	6.3873e+09	64	99802043.4	Root MSE =	3894

CONH	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
TRIM1	8947.789	973.5125	9.19	0.000	7000.474	10895.1
TRIM2	9163.723	973.5125	9.41	0.000	7216.408	11111.04
TRIM3	9348.999	973.5125	9.60	0.000	7401.685	11296.31
TRIM4	9534.361	973.5125	9.79	0.000	7587.046	11481.68

$$\widehat{CONH}_t = 8947.79TRIM1_t + 9163.72TRIM2_t + 9348.99TRIM3_t + 9534.36TRIM4_t + \mu_t \quad (3.7.20)$$

$$t = \quad (9.19) \quad (9.41) \quad (9.60) \quad (9.79) \quad R^2 = 0.86$$

Los errores estándar de todos los coeficientes estimados son iguales a 973.51, ya que todas las variables dicótomas toman el valor de 1 o de 0.

Los coeficientes estimados  $\alpha$  de (3.7.20) representan el valor promedio o media del consumo de los hogares ecuatorianos. Por lo tanto, el promedio de consumo de los hogares en el primer trimestre es de casi 8947.79 millones de dólares, en el segundo trimestre es de casi 9163.72 millones de dólares, en el tercer trimestre es de casi 9348.99 millones de dólares, y en el cuarto trimestre es de casi 9534.36 millones de dólares.

Ahora, se considera la regresión:

$$CONH_t = \beta_1 + \beta_2 TRIM2_t + \beta_3 TRIM3_t + \beta_4 TRIM4_t + \mu_t \quad (3.7.21)$$

Nótese que la regresión (3.7.21) incluye el término de intercepto, y excluye un trimestre, que en este caso es el primer trimestre, el cual es la categoría base o de comparación. Por consiguiente para estimar la regresión (3.7.21) se ejecuta el comando:

**reg CONH TRIM2 TRIM3 TRIM4**, Y se obtiene los siguientes resultados:

Source	SS	df	MS	Number of obs =	64
Model	3030890.63	3	1010296.88	F( 3, 60) =	0.07
Residual	909817467	60	15163624.5	Prob > F =	0.9774
				R-squared =	0.0033
				Adj R-squared =	-0.0465
Total	912848358	63	14489656.5	Root MSE =	3894

CONH	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
TRIM2	215.9344	1376.755	0.16	0.876	-2537.985 2969.853
TRIM3	401.2106	1376.755	0.29	0.772	-2352.709 3155.13
TRIM4	586.572	1376.755	0.43	0.672	-2167.347 3340.491
_cons	8947.789	973.5125	9.19	0.000	7000.474 10895.1

$$\widehat{CONH}_t = 8947.79 + 215.93TRIM2_t + 401.21TRIM3_t + 586.57TRIM4_t + \mu_t \quad (3.7.22)$$

$$t = (9.19) \quad (0.16) \quad (0.29) \quad (0.43) \quad R^2 = 0.86$$

Al tener no significativos los coeficientes diferenciales de la regresión (3.7.22), se afirma que no existe estacionalidad en la serie de tiempo consumo de los hogares. Sin embargo es necesario saber cómo desestacionalizar una serie de tiempo que si presenta estacionalidad. Por tal razón se va a suponer que la variable consumo de hogares presenta estacionalidad y para desestacionalizarla se procede a estimar (3.7.19) [o (3.7.21)], predecir los residuos, sumarle a los residuos la media de la variable  $Y$  real, y generar la nueva variable  $Y$  desestacionalizada. Para ello se ejecutan los siguientes comandos:

**reg CONH TRIM1 TRIM2 TRIM3 TRIM4, noconst**

**predict r, resid**

**mean(CONH)**

**gen Ynueva= r+9248.718**

Interpretación:

Los resultados de la regresión (3.7.22) muestran que los coeficientes diferenciales de las variables dicótomas no son estadísticamente significativos, lo cual significa que el segundo, tercer y cuarto trimestre son estadísticamente iguales al primer trimestre, por lo tanto no se evidencia estacionalidad en el consumo de los hogares contrario a lo que se esperaba a priori.

### 3.7.7 Regresión lineal por segmentos.

#### Ejercicio 7:

En una regresión lineal por segmentos la variable  $Y$  aumenta linealmente con  $X$  hasta el nivel de umbral  $X^*$ , después del cual  $Y$  también aumenta linealmente con  $X$  pero a una tasa mayor. Con la técnica de las variables dicótomas se estiman las diferentes pendientes de los dos segmentos de la regresión lineal por secciones (Gujarati & Porter, 2010). La regresión lineal por segmentos se expresa de la siguiente manera:

$$Y_i = \alpha_1 + \beta_1 X_i + \beta_2 (X_i - X^*) D_i + \mu_i \quad (3.7.23)$$

Comandos a utilizar:

Rutina utilizada:  
**gen** VENTAS2= VENTAS-500  
**gen** VENTAS2D= VENTAS2\*D  
**reg** COMISION VENTAS VENTAS2D

Desarrollo:

Se consideran los datos hipotéticos de comisión de ventas y ventas (en dólares) de la Tabla 3.7.6 para estimar una regresión lineal por segmentos. Se dice que la comisión por ventas va a cambiar su pendiente al alcanzar un monto de ventas de 500 dólares. Por lo tanto se procederá a estimar la siguiente regresión:

$$COMISION_i = \alpha_1 + \beta_1 VENTAS_i + \beta_2 (VENTAS_i - 500) D_i + \mu_i \quad (3.7.24)$$

Donde, COMISIÓN= comisión por ventas (\$)

VENTAS= monto de ventas (\$)

500= valor del umbral de ventas (\$)

D= 1 si VENTAS>500; 0 si VENTAS<500

Para estimar la regresión (3.7.24) se ejecutan los siguientes comandos:

**gen** VENTAS2= VENTAS-500

**gen** VENTAS2D= VENTAS2\*D

**reg** COMISION VENTAS VENTAS2D



Y se obtienen los siguientes resultados:

Source	SS	df	MS	Number of obs = 49		
Model	1629195.51	2	814597.757	F( 2, 46) = 1128.17		
Residual	33214.2807	46	722.049581	Prob > F = 0.0000		
				R-squared = 0.9800		
				Adj R-squared = 0.9792		
Total	1662409.8	48	34633.5374	Root MSE = 26.871		

COMISION	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
VENTAS	.4673969	.0286835	16.29	0.000	.40966	.5251338
VENTAS2D	.1300731	.0449733	2.89	0.006	.0395466	.2205997
_cons	-27.68641	9.881628	-2.80	0.007	-47.5771	-7.795727

$$COMISION_i = -27.69 + 0.48VENTAS_i + 0.13(VENTAS_i - 500)D_i + \mu_i \quad (3.7.25)$$

$$t = \quad (-2.80) \quad (16.29) \quad (2.89)$$

$$R^2 = 0.98 \quad X^* = \$500$$

Interpretación:

Los resultados de la regresión lineal por segmentos (3.7.25) muestran que tanto la variable VENTAS como la variable dicótoma son significativas, por lo tanto, a medida que aumente 1 dólar en las ventas las comisiones por ventas aumentarán, en promedio, 61 centavos (48+13).

### 3.7.8 Variables dicótomas en regresiones semilogarítmicas.

#### Ejercicio 8:

En el apartado 3.5.5 se vio las regresiones semilogarítmicas cuando la variable  $X$  es cuantitativa. El presente ejercicio explica que sucede cuando la variable  $X$  de una regresión semilogarítmica es de naturaleza cualitativa. Por consiguiente, considere el siguiente modelo:

$$\ln Y_i = \beta_1 + \beta_2 D_i + \mu_i \quad (3.7.26)$$

Si  $D$  es igual a 1 entonces:

$$\ln Y_i = \beta_1 + \beta_2 \quad (3.7.27)$$

Si  $D$  es igual a 0 entonces:

$$\ln Y_i = \beta_1 \quad (3.7.28)$$

Comandos a utilizar:

Rutina utilizada:  
**gen** lnW= **log**(W)  
**reg** lnW SEXO

Desarrollo:

Para ilustrar (3.7.26) se utilizarán los datos del salario mensual y sexo de la Tabla 3.7.7. En este sentido, se considera la siguiente regresión:

$$\ln W_i = \beta_1 + \beta_2 SEXO_i + \mu_i \quad (3.7.29)$$

Donde,  $\ln W$  es el logaritmo del salario mensual de los trabajadores de Quito, y SEXO es igual a 1 para hombres y 0 para mujeres. Para estimar (3.7.29) se genera primeramente la variable logarítmica de  $W$  (salario) con el comando: **gen lnW=log(W)**, y se estima la regresión con el comando: **reg lnW SEXO**, obteniendo los siguientes resultados:

Source	SS	df	MS	Number of obs = 1369		
Model	9.29759077	1	9.29759077	F( 1, 1367) = 20.31		
Residual	625.756416	1367	.457758899	Prob > F = 0.0000		
Total	635.054006	1368	.464220765	R-squared = 0.0146		
				Adj R-squared = 0.0139		
				Root MSE = .67658		

lnW	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
SEXO	.1658604	.0368024	4.51	0.000	.0936652	.2380557
_cons	6.148558	.0274389	224.08	0.000	6.094732	6.202385

$$\ln W_i = 6.148558 + 0.1658604 SEXO_i + \mu_i \quad (3.7.30)$$

$$t = (224.08) \quad (4.51) \quad R^2 = 0.008$$

Con el antilogaritmo de 6.148558 se tiene 468.04 que es la mediana del salario mensual de las trabajadoras, y si se toma el antilogaritmo de (6.148558+0.1658604=6.3144184) se tiene 552.48 que es la mediana del salario mensual de los trabajadores.

Para obtener la semielasticidad para una regresora dicótoma se tome el antilogaritmo (base e) del coeficiente estimado de la variable dicótoma, a ello se le resta 1 y se multiplica la diferencia por 100, así:  $\text{antilog}(0.1658604)=1.1804$ ;  $1.1804-1=0.1804$ ; y  $0.1804 \times 100= 18.04$ .

Interpretación:

De acuerdo con los resultados de (3.7.30) el salario de los hombres es mayor que el de las mujeres por \$ 84.44 (resultado de 552.48 – 468.04). Además la semielasticidad es igual a 18.04.

### 3.8. Multicolinealidad

Las aportaciones de Gujarati & Porter (2010) definen a la multicolinealidad o colinealidad cuando alguna de las variables explicativas tiene una relación lineal exacta o aproximadamente exacta con otra variable explicativa dentro de la ecuación, es decir estas variables explicativas no son linealmente independientes. Para aplicar estos aportes dentro del caso ecuatoriano se utilizan tres tablas de datos provenientes de fuentes estadísticas nacionales e internacionales, el área de la economía al que se aplican se describe a continuación: 1) Modelo macroeconómico de crecimiento en función de capital humano y recaudación de impuestos, 2) Producción hotelera total determinada por mano de obra, participación de utilidades y medidas publicitarias, 3) Modelo de crecimiento económico, ambiente y comercio.

#### Ejercicio 1:

El modelo que se presenta a continuación analiza la relación entre el Valor Agregado Bruto cantonal, recuento de personas estudiando desde preescolar hasta postrado y los impuestos para el año 2010, utilizando la tabla 3.8.1. Esta regresión permite analizar el desarrollo y crecimiento económico vía capital humano y políticas impositivas mediante la ecuación 3.8.1:

$$VAB = \beta_1 + \beta_2 EST + \beta_3 IMP + u_i \quad (3.8.1)$$

Source	SS	df	MS	Number of obs = 219		
Model	4.0889e+14	2	2.0444e+14	F( 2, 216) = 6865.45		
Residual	6.4322e+12	216	2.9779e+10	Prob > F = 0.0000		
Total	4.1532e+14	218	1.9051e+12	R-squared = 0.9845		
				Adj R-squared = 0.9844		
				Root MSE = 1.7e+05		

VAB	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
EST	6.073999	.1831442	33.17	0.000	5.71302	6.434977
IMP	1.23e-08	1.24e-09	9.89	0.000	9.82e-09	1.47e-08
_cons	-29968.38	13048.93	-2.30	0.023	-55687.92	-4248.835

#### 3.8.1 Detección de Multicolinealidad.

- **Factor Inflacionario de la Varianza.**

Calcula los valores inflacionarios de varianza de las variables explicativas, generalmente se lo usa para detectar colinealidad entre estas.

- **Tolerancia.**

La tolerancia (TOL) es la inversa del factor inflacionario de la varianza, cuanto más se acerque su valor a cero mayor será el grado de correlación entre las variables.

Comandos a utilizar:

**estat vif:** Muestra un tabla de valores de factores inflacionarios de varianza, su media y la inversa de cada uno de estos.

Su estructura es:

**estat vif**

Rutina Utilizada:

**regress VAB EST IMP**

**estat vif**

Desarrollo:

Luego de realizar una regresión se aplica la estructura descrita y se obtiene la siguiente tabla:

Variable	VIF	1/VIF
EST	7.59	0.131759
IMP	7.59	0.131759
Mean VIF	7.59	

Interpretación:

Si el valor de VIF es superior a la unidad y la TOL es cercana a cero se confirma la correlación entre estas variables, por lo que para el presente ejercicio las variables recuento de personas estudiando y los impuestos muestran un alto grado de correlación.<sup>14</sup>

### **Ejercicio 2:**

Este ejercicio utiliza los datos de la tabla 3.8.2 y muestra la relación entre producción total, total de empleados, total de horas trabajadas en el mes de noviembre, sueldos y salarios de empleados mujeres, sueldos y salarios de empleados hombres, participación en las utilidades y publicidad de los hoteles del Ecuador en el año 2013. Cabe recalcar que se tomó en cuenta exclusivamente a las microempresas hoteleras que invirtieron en publicidad para dicho periodo obteniendo la siguiente ecuación:

$$PT = \beta_1 + \beta_2 TE + \beta_3 HT + \beta_4 SEM + \beta_5 SEH + \beta_6 U + \beta_7 P + u_i \quad (3.8.2)$$

<sup>14</sup> Véase <http://support.minitab.com/es-mx/minitab/17/topic-library/modeling-statistics/regression-and-correlation/model-assumptions/what-is-a-variance-inflation-factor-vif/> para criterios de directrices y su grado de correlación.

Source	SS	df	MS	Number of obs =	205
Model	5.4545e+12	6	9.0909e+11	F( 6, 198) =	42.77
Residual	4.2089e+12	198	2.1257e+10	Prob > F =	0.0000
				R-squared =	0.5644
				Adj R-squared =	0.5513
Total	9.6635e+12	204	4.7370e+10	Root MSE =	1.5e+05

PT	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
TE	-13760.74	5080.342	-2.71	0.007	-23779.26	-3742.217
HT	53.56482	20.55557	2.61	0.010	13.02889	94.10075
SEM	48.86181	11.44301	4.27	0.000	26.296	71.42763
SEH	49.33041	10.29743	4.79	0.000	29.02369	69.63713
U	3.991629	.8621849	4.63	0.000	2.291386	5.691873
P	10.46109	2.558723	4.09	0.000	5.415241	15.50693
_cons	95513.14	18518.18	5.16	0.000	58994.97	132031.3

- **Tabla de correlación.**

### Ejercicio 3:

Genera los coeficientes de correlación ( $r$ ) de las variables incluidas en el modelo econométrico, de esta manera se analiza su valor entre parejas de regresoras para detectar la presencia de multicolinealidad.

Comandos a utilizar:

**correlate:** Muestra un tabla de coeficientes de correlación de las variables seleccionadas.

Su estructura es:

**correlate** Y X1 X2...Xn

Rutina Utilizada:

**reg** PT TE HT SEM SEH U P

**correlate**

Desarrollo:

Se procede a aplicar el comando de correlación para obtener una tabla de coeficientes como la que se muestra a continuación:

	PT	TE	HT	SEM	SEH	U	P
PT	1.0000						
TE	0.6130	1.0000					
HT	0.6313	0.9518	1.0000				
SEM	0.5352	0.7061	0.6434	1.0000			
SEH	0.5672	0.8486	0.8301	0.3634	1.0000		
U	0.3224	0.1135	0.1093	0.2786	-0.0162	1.0000	
P	0.3512	0.1828	0.2130	0.1849	0.1451	0.0932	1.0000

Interpretación:

Gran parte de los coeficientes de correlación poseen valores relativamente bajos a excepción de los que se muestran entre: total de empleados y horas trabajadas; total de empleados y salarios de las empleadas mujeres y finalmente horas trabajadas y salarios de empleados hombres.

#### **Ejercicio 4:**

Se presenta la relación entre crecimiento económico, ambiente y comercio del periodo 1970-2014 mediante las variables PIB a precios constantes del Ecuador en función de las emisiones de CO2, las exportaciones de bienes y servicios y el comercio de mercaderías en base a datos de la tabla 3.8.3, la ecuación utilizada es la siguiente:

$$PIB = \beta_1 + \beta_2 CO2 - \beta_3 EXP + \beta_4 CM + u_t \quad (3.8.3)$$

Source	SS	df	MS	Number of obs = 42		
Model	6.0565e+21	3	2.0188e+21	F( 3, 38) =	214.10	
Residual	3.5832e+20	38	9.4294e+18	Prob > F =	0.0000	
				R-squared =	0.9441	
				Adj R-squared =	0.9397	
Total	6.4148e+21	41	1.5646e+20	Root MSE =	3.1e+09	

PIB	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
CO2	1.85e+10	3.37e+09	5.48	0.000	1.17e+10	2.53e+10
EXP	-4.52e+08	2.11e+08	-2.14	0.039	-8.80e+08	-2.40e+07
CM	1.18e+09	1.08e+08	10.89	0.000	9.61e+08	1.40e+09
_cons	-9.54e+09	2.44e+09	-3.90	0.000	-1.45e+10	-4.59e+09

- **Diagrama de dispersión.**

Crea una matriz de dispersión de las variables seleccionadas y señala su relación de manera gráfica, este método muestra correlación si la línea de puntos se muestra de manera diagonal dentro del recuadro de las ordenadas y abscisas.

Comandos a utilizar:

**graph matrix:** Genera una matriz de gráficos de dispersión de las variables seleccionadas.

Su estructura es:

**graph matrix** Y X1 X2...Xn

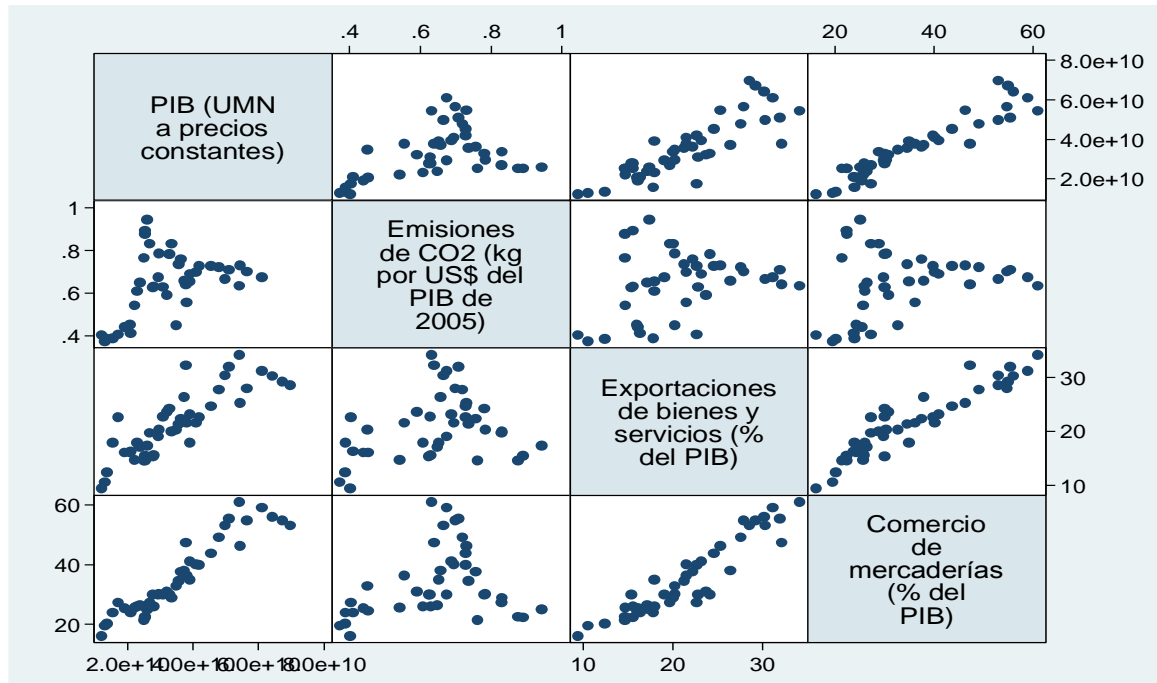
Rutina Utilizada:

**reg** PIB CO2 EXP CM

**graph matrix** PIB CO2 EXP CM

Desarrollo:

Se ejecuta el comando a través de la barra de comandos en la parte inferior de la ventana y se genera la matriz de gráficos que se muestra a continuación:



Interpretación:

Si se trazara una línea de tendencia entre los cuadros de la matriz se mostraría una correlación casi perfecta entre X2 y X3 (exportaciones y el comercio de bienes) dado que los puntos de dispersión se alinean de manera lineal positiva, este no sería el caso para X1 cuyos puntos se muestran demasiado dispersión en relación a las demás regresoras, de esta manera se puede aseverar la presencia de multicolinealidad en el modelo.

- **Regresiones auxiliares.**

### Ejercicio 5:

Son combinaciones lineales entre regresoras de la tabla 3.8.3, es decir un conjunto de regresiones de X en función de las X restantes del modelo.

Comandos a utilizar:

Rutina Utilizada:  
**regress** CO2 EXP CM  
**regress** EXP CO2 CM  
**regress** CM CO2 EXP

Desarrollo:

Se realizan tres regresiones con las variables independientes del modelo resultando en:

$$CO_2 = \beta_1 + \beta_2 EXP - \beta_3 CM + v_t \quad (3.8.4)$$

Source	SS	df	MS	Number of obs = 42		
Model	.073649995	2	.036824998	F( 2, 39) = 1.73		
Residual	.829362573	39	.021265707	Prob > F = 0.1903		
Total	.903012568	41	.022024697	R-squared = 0.0816		
				Adj R-squared = 0.0345		
				Root MSE = .14583		

CO2	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
EXP	.0073589	.0099732	0.74	0.465	-.0128139	.0275316
CM	-.0001311	.0051449	-0.03	0.980	-.0105376	.0102755
_cons	.4942569	.0849452	5.82	0.000	.3224391	.6660747

$$EXP = \beta_1 + \beta_2 CO_2 + \beta_3 CM + w_t \quad (3.8.5)$$

Source	SS	df	MS	Number of obs = 42		
Model	1239.95314	2	619.97657	F( 2, 39) = 114.67		
Residual	210.857084	39	5.40659189	Prob > F = 0.0000		
Total	1450.81022	41	35.3856152	R-squared = 0.8547		
				Adj R-squared = 0.8472		
				Root MSE = 2.3252		

EXP	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
CO2	1.870918	2.535591	0.74	0.465	-3.257799	6.999634
CM	.4700321	.0326333	14.40	0.000	.404025	.5360393
_cons	3.776151	1.749688	2.16	0.037	.2370724	7.31523

$$CM = -\beta_1 - \beta_2 CO_2 + \beta_3 EXP + z_t \quad (3.8.6)$$

Source	SS	df	MS	Number of obs = 42		
Model	4648.2901	2	2324.14505	F( 2, 39) = 112.83		
Residual	803.379458	39	20.5994733	Prob > F = 0.0000		
Total	5451.66955	41	132.96755	R-squared = 0.8526		
				Adj R-squared = 0.8451		
				Root MSE = 4.5387		

CM	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
CO2	-.1269574	4.983704	-0.03	0.980	-10.20745	9.953535
EXP	1.790854	.1243351	14.40	0.000	1.539362	2.042345
_cons	-3.491801	3.569956	-0.98	0.334	-10.71272	3.729118



Interpretación:

Siguiendo la regla práctica de Klein, en la cual se dice que existe multicolinealidad si el  $R^2$  de las regresiones auxiliares es mayor al  $R^2$  de la regresión global, se determina que si bien hay valores muy altos de  $R^2$  en las regresiones de EXP y CM no se puede asegurar la presencia de multicolinealidad.

### 3.8.2 Corrección de Multicolinealidad.

- **Especificación correcta del modelo.**

#### **Ejercicio 6:**

Cuando dentro de un modelo se presentan variables cuantificadas con distintas formas de medición se debe analizar que las regresiones estén correctamente especificadas, ya que en ocasiones los modelos lineales de MCO no se ajustan a la naturaleza de los datos provistos.

Comandos a utilizar:

```
Rutina Utilizada:
gen LVAB= log(VAB)
gen LEST= log(EST)
gen LIMP= log(IMP)
reg LVAB LEST LIMP
estat vif
```

Desarrollo:

Para el caso del ejercicio 1 visto en este capítulo se obtienen los logaritmos de las variables regresoras y regresadas, obteniendo una regresión doble-log o log-lineal.

$$LVAB = \beta_1 + \beta_2 LEST + \beta_3 LIMP + u_i \quad (3.8.7)$$

Source	SS	df	MS	Number of obs = 219		
Model	376.229305	2	188.114652	F( 2, 216) = 565.98		
Residual	71.7920909	216	.332370791	Prob > F = 0.0000		
				R-squared = 0.8398		
				Adj R-squared = 0.8383		
Total	448.021396	218	2.05514402	Root MSE = .57652		

LVAB	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
LEST	.8501656	.0509463	16.69	0.000	.7497501	.9505811
LIMP	.1099233	.0217803	5.05	0.000	.0669942	.1528525
_cons	1.3136	.3240766	4.05	0.000	.6748425	1.952357

Seguido de esto, se compara los valores de VIF y TOL de la regresión logarítmica con la original.

Variable	VIF	1/VIF
LEST	2.60	0.385190
LIMP	2.60	0.385190
Mean VIF	2.60	

Interpretación:

El valor de la inflación de la varianza es menor a la del original evidenciando una disminución en la correlación de las variables, de igual forma la tolerancia es comparativamente mayor y se reduce la posibilidad de multicolinealidad en el modelo logarítmico.

- **Transformación de una variable.**

### **Ejercicio 7:**

Existen distintas de manera de transformar una variable, a través de la estandarización, operaciones algebraicas, aplicando logaritmos, derivadas, etc. Con el fin de corregir la multicolinealidad presentada en el ejercicio 2 de este capítulo se realiza el siguiente proceso.

Comandos a utilizar:

Rutina Utilizada:

```

gen PTP=(PT/TE)
gen SE= SEM+ SEH
gen HTP= (HT/TE)
gen SEP=(SE/TE)
gen UP=(U/TE)
gen PP=(P/TE)
reg PTP HTP SEP UP PP
correlate PTP HTP SEP UP PP

```

Desarrollo:

Para este ejercicio se realizan distintos procesos de transformación con el fin de eliminar la multicolinealidad del modelo. Primeramente se suman los sueldos de empleados hombres y mujeres y se consolidan como una sola variable de sueldos de empleados (SE). Luego de esto se generan nuevas variables por trabajador, es decir, así como existen variables per cápita dentro de la macroeconomía que resultan de la división de la población para cada variable, en este modelo se presentarán valores por trabajador al dividir cada variable por el total de empleados obteniendo así: producción total per cápita (PTP) horas trabajadas per cápita (HTP), Salario de empleados per cápita (SEP), utilidades per cápita (UP) y finalmente publicidad per cápita (PP) para realizar la nueva regresión.

$$PTP = \beta_1 + \beta_2 HTP + \beta_3 SEP + \beta_5 UP + \beta_6 PP + u_i \quad (3.8.8)$$

Source	SS	df	MS	Number of obs = 203		
Model	7.3093e+10	4	1.8273e+10	F( 4, 198) = 53.36		
Residual	6.7809e+10	198	342468481	Prob > F = 0.0000		
				R-squared = 0.5188		
				Adj R-squared = 0.5090		
Total	1.4090e+11	202	697532113	Root MSE = 18506		

PTP	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
HTP	243.5451	18.93193	12.86	0.000	206.211	280.8792
SEP	34.94534	14.48965	2.41	0.017	6.371491	63.5192
UP	4.268609	1.233739	3.46	0.001	1.835653	6.701565
PP	5.89523	1.689711	3.49	0.001	2.56309	9.22737
_cons	-36886.24	6616.764	-5.57	0.000	-49934.61	-23837.86

El modelo corregido permanece robusto y con sentido económico en el signo de sus coeficientes, por lo que ahora se prueba si el problema de multicolinealidad fue corregido mostrando una tabla de correlación.

	PTP	HTP	SEP	UP	PP
PTP	1.0000				
HTP	0.6651	1.0000			
SEP	0.1902	0.0897	1.0000		
UP	0.2084	0.0607	-0.0228	1.0000	
PP	0.2071	0.0332	0.0975	0.0081	1.0000

Interpretación:

Al aplicar la transformación per cápita a las variables se aprecian coeficientes de correlación menores a uno, por lo que se afirma que el presente modelo no presenta multicolinealidad.

- **Eliminación de una variable.**

### **Ejercicio 8:**

Una de las maneras más comunes de eliminar la multicolinealidad es con la eliminación de la variable que genera este fenómeno, utilizando los datos de la tabla 3.8.3 realizamos la siguiente corrección.

Comandos a utilizar:

Rutina Utilizada:  
**drop** CM  
**reg** PIB CO2 EXP  
**graph matrix** PIB CO2 EXP

Desarrollo:

Se elimina la variable problemática del modelo y se procede a estimar la regresión de manera que se explique el crecimiento económico (cuantificado por el valor agregado bruto) en función al capital (recuento de personas estudiando desde preescolar hasta postgrado) con la siguiente ecuación:

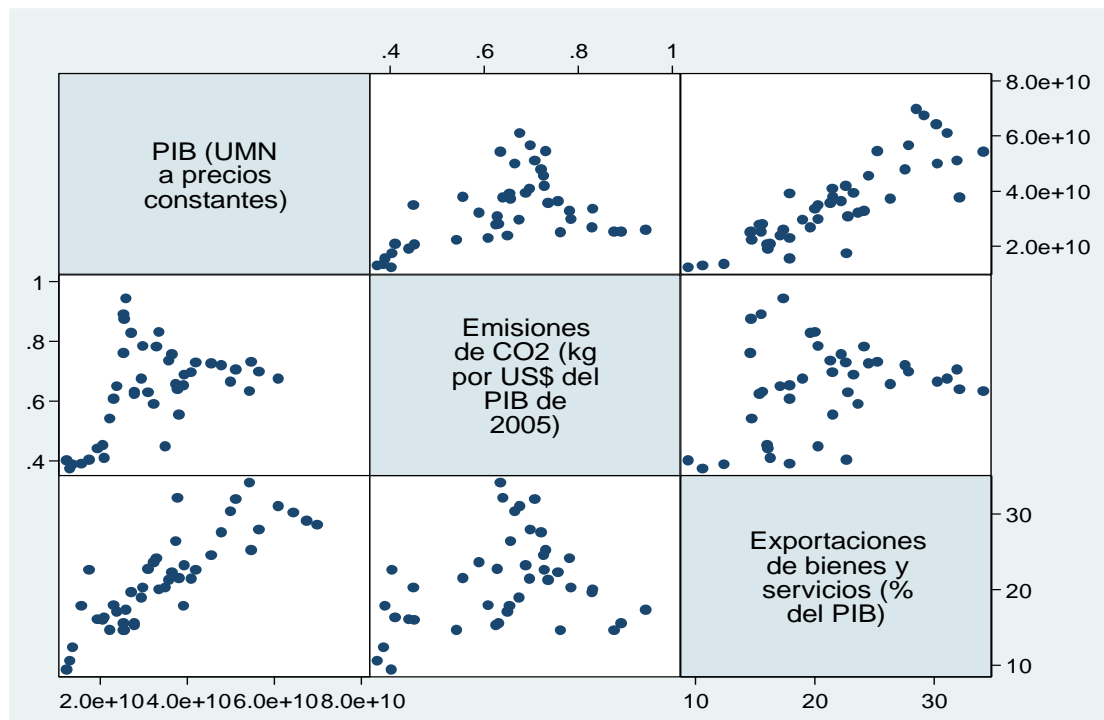
$$PIB = \beta_1 + \beta_2 CO_2 + \beta_3 EXP + u_t \quad (3.8.9)$$

Source	SS	df	MS	Number of obs =	42
Model	4.9376e+21	2	2.4688e+21	F( 2, 39) =	65.18
Residual	1.4772e+21	39	3.7877e+19	Prob > F =	0.0000
				R-squared =	0.7697
				Adj R-squared =	0.7579
Total	6.4148e+21	41	1.5646e+20	Root MSE =	6.2e+09

PIB	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
CO2	1.83e+10	6.76e+09	2.71	0.010	4.66e+09	3.20e+10
EXP	1.66e+09	1.69e+08	9.85	0.000	1.32e+09	2.00e+09
_cons	-1.37e+10	4.84e+09	-2.82	0.007	-2.34e+10	-3.87e+09

Para comprobar si permanece la multicolinealidad se genera la matriz de correlación de las variables:



Interpretación:

Al eliminar la variable comercio de mercancías del modelo, los signos de los coeficientes se ajustan de manera positiva y su significancia permanece. La multicolinealidad desaparece entre las regresoras dada la baja correlación entre las emisiones de  $CO_2$  y las exportaciones de bienes.

### 3.9. Heteroscedasticidad

Cuando el supuesto de homoscedasticidad en una regresión lineal no se cumple, se presenta el problema de la heteroscedasticidad en el modelo, el cual no es más que varianzas desiguales del término de perturbación  $u$  ante variaciones en los valores de  $X$ . Para que se mantenga el supuesto de homoscedasticidad, los valores de las perturbaciones no necesitan ser los mismos (ante mayores valores de  $X$ , mayor el valor del término de error), sino que las perturbaciones se muestren igualmente dispersas ante variaciones en  $X$  (Arce & Mahía, 2008). En este apartado se estudia la presencia de heteroscedasticidad en modelos de desempeño académico, ingreso nacional neto ajustado y determinantes de valor de las edificaciones.

#### Ejercicio 1:

La tabla 3.9.1 presenta información recabada de un grupo de estudiantes de la Universidad Técnica Particular de Loja sobre su desempeño académico y situación socioeconómica con el siguiente modelo de regresión:

$$NCA = \beta_1 + \beta_2 Edad + \beta_3 NEA + \beta_4 NCM + \beta_5 PROM + u_i \quad (3.9.1)$$

Donde, NCA representa el número de créditos aprobados en el periodo abril-agosto 2013, NEA muestra la nota del examen de admisión, NCM es el número de componentes por matrícula del periodo abril-agosto 2013 y finalmente PROM es el promedio sobre 40 puntos obtenido en el periodo académico respectivo.

Source	SS	df	MS	Number of obs = 209		
Model	7436.81228	4	1859.20307	F( 4, 204) = 49.28		
Residual	7695.92456	204	37.7251204	Prob > F = 0.0000		
				R-squared = 0.4914		
				Adj R-squared = 0.4815		
Total	15132.7368	208	72.7535425	Root MSE = 6.1421		

NCA	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
edad	.3698733	.2172191	1.70	0.090	-.0584091	.7981557
NEA	.0476808	.0251105	1.90	0.059	-.0018286	.0971902
NCM	2.620421	.2470769	10.61	0.000	2.13327	3.107573
PROM	.7798526	.0969103	8.05	0.000	.5887782	.9709269
_cons	-27.67653	6.287733	-4.40	0.000	-40.07381	-15.27926

### 3.9.1 Detección de Heteroscedasticidad.

Arce (2001) menciona que algunas de las causas frecuentes de heteroscedasticidad se presentan cuando los datos de un modelo son de corte transversal, cuando sus variables muestran grandes variaciones en su recorrido ya sea porque están ordenados bajo ciertos parámetros (alfabético, sectorial, geográfico, etc.) y su varianza tiende a ser muy grande, además de la omisión de variables importantes y especificación incorrecta del modelo como otras causantes del problema.

#### 3.9.1.1 Métodos informales.

- **Método Gráfico.**

#### Ejercicio 2:

Continuando con los datos de ejercicio 1, se procede a realizar un diagrama de dispersión los residuos al cuadrado de la regresión frente a los valores de Y estimados.

Comandos a utilizar:

Rutina Utilizada:

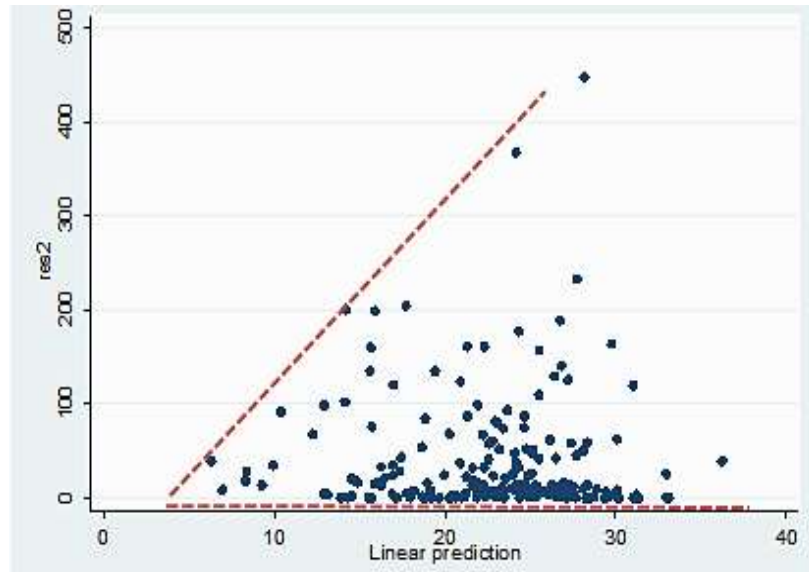
```

reg NCA edad NEA NCM PROM
predict YF, xb
predict res, residual
gen res2= res*res
scatter res2 YF

```

Desarrollo:

Al generar la regresión planteada se procede a predecir los valores estimados (YF) y los residuos al cuadrado (res2), los cuales se muestran el siguiente gráfico de dispersión:



Interpretación:

La dispersión muestra un esquema similar al de la figura 11.8 b) del texto de Gujarati & Porter (2010), confirmando la presencia de heteroscedasticidad en el modelo.

### **Ejercicio 3:**

Adicionalmente, un método no encontrado en el texto básico de Econometría y que es muy usado por los investigadores para detectar heteroscedasticidad es la representación de los residuos versus los valores ajustados.

Comandos a utilizar:

**rvfplot:** Genera un gráfico de dispersión de los residuos ante los valores ajustados de la regresión.

Su estructura es:

**rvfppplot** . Para generar un gráfico simple

**rvfppplot**, yline(0) . Para generar el gráfico con una línea en el origen.

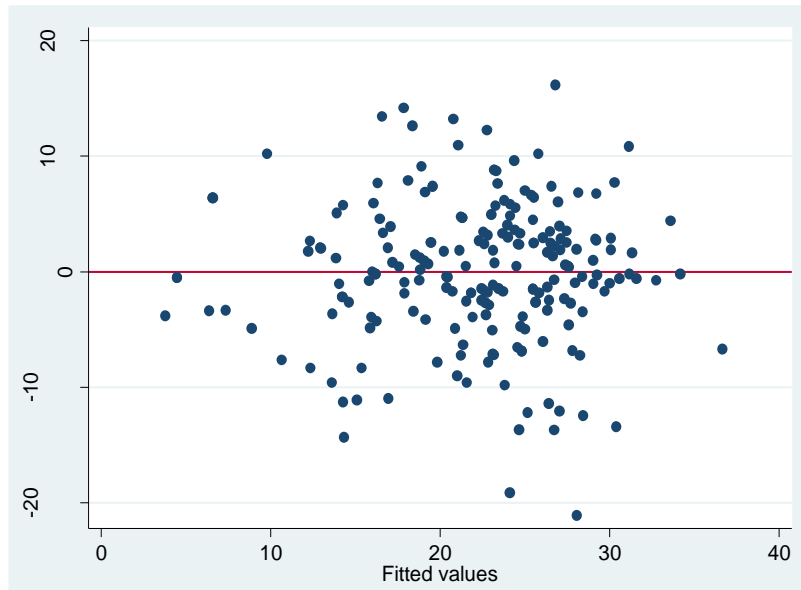
Rutina Utilizada:

**reg** NCA edad NEA NCM PROM

**rvfppplot**, yline(0)

Desarrollo:

Se efectúa la regresión del número de créditos en función de la nota del examen de admisión, el número de componentes y el promedio obtenido, luego de esto se aplica el comando anteriormente descrito y se obtiene la siguiente gráfica.



Interpretación:

Se aprecia una distribución sistemática en ciertos valores que tienden a centrarse en un área específica por lo que, de manera informal se confirma la presencia de heteroscedasticidad en el modelo.

### 3.9.1.2 Métodos formales.

- **Test de White.**

#### Ejercicio 4:

Esta prueba determina si las variables explicativas del modelo presentan heteroscedasticidad bajo el supuesto de hipótesis. Según Gujarati & Porter (2010) esta prueba no se apoya en el supuesto de normalidad y es relativamente fácil de aplicar.

Comandos a utilizar:

**estat imtest, white:** Genera una matriz de información donde se presenta heteroscedasticidad irrestricta, simetría y curtosis.

Su estructura es:

**estat imtest, white**

Rutina Utilizada:

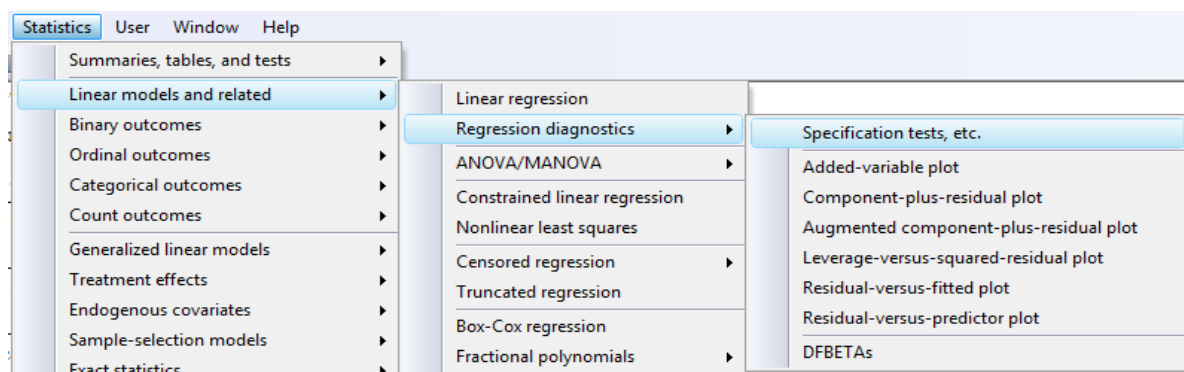
**reg NCA edad NEA NCM PROM**

**estat imtest, white**

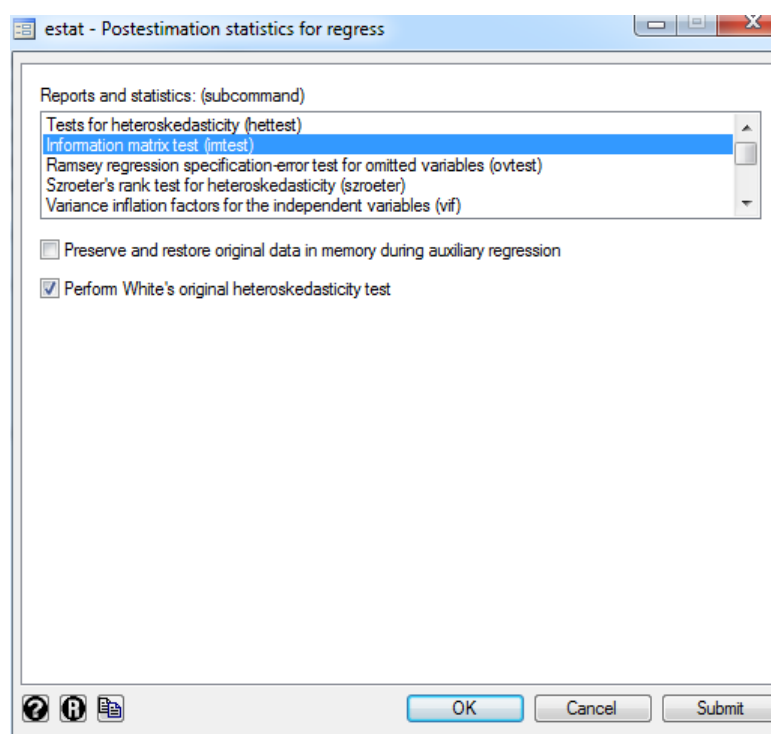


Desarrollo:

Con los datos del ejercicio 1 se aplica la rutina descrita anteriormente para obtener el recuadro de resultados. Este test también se lo puede obtener a través del cuadro de herramientas de Stata bajo el siguiente procedimiento:



Seleccionando *Statistics*, *Linear models and related*, *Regression diagnostics* y finalmente *Specification, test, etc.* Se despliega la siguiente tabla:



En la que se selecciona la opción de *Information matrix test (imtest)*, se marca la pestaña de *Perform White's original heteroskedasticity test* y finalmente se hace clic en OK.

```

White's test for Ho: homoskedasticity
against Ha: unrestricted heteroskedasticity

chi2(14)      =      21.80
Prob > chi2   =      0.0828

Cameron & Trivedi's decomposition of IM-test

```

Source	chi2	df	p
Heteroskedasticity	21.80	14	0.0828
Skewness	5.46	4	0.2431
Kurtosis	2.26	1	0.1328
Total	29.52	19	0.0582

Como se aprecia, se genera una matriz de resultados donde se tienen los valores probabilísticos y de chi2 para comprobar la presencia de heteroscedasticidad, simetría y curtosis. Si se desea analizar de manera exclusiva a la heteroscedasticidad se puede usar el comando **whitetst** luego de haber efectuado la regresión, sin embargo este comando debe instalarse en el sistema siguiendo los pasos mencionados en el apartado 3.1 de este capítulo.

Interpretación:

El test de White asevera que existe heteroscedasticidad si su probabilidad es menor a 0.05 y el valor de chi2 es alto, por lo que en este ejercicio particular se aprecia la posible presencia de heteroscedasticidad en las variables ya que si bien no se acepta la hipótesis nula al 5% de confianza se la aceptaría en el 10% de confianza.

- **Breusch-Pagan / Cook-Weisberg Test de Heteroscedasticidad.**

Williams (2015) menciona que este test está diseñado para detectar heteroscedasticidad lineal dentro del modelo bajo la hipótesis nula de que existe homoscedasticidad en las variables o si se acepta la hipótesis alterna de heteroscedasticidad.

### **Ejercicio 5:**

Con los datos del World Development Indicators (WDI) presentados en la tabla 3.9.2, se analiza el modelo de ingreso nacional neto ajustado en función del gasto de consumo de los hogares, el porcentaje de trabajadores asalariados y el porcentaje de inscripción escolar para el caso ecuatoriano en los años de 1988 al 2013.

$$IN = \beta_1 + \beta_2 GC - \beta_3 TA + \beta_4 IES + u_t \quad (3.9.2)$$

Source	SS	df	MS	Number of obs =	18
Model	9.0829e+20	3	3.0276e+20	F( 3, 14) =	115.77
Residual	3.6612e+19	14	2.6151e+18	Prob > F =	0.0000
				R-squared =	0.9613
				Adj R-squared =	0.9530
Total	9.4490e+20	17	5.5582e+19	Root MSE =	1.6e+09

IN	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
GC	1.90e+07	4099277	4.62	0.000	1.02e+07	2.77e+07
TA	-2.79e+08	1.54e+08	-1.80	0.093	-6.10e+08	5.26e+07
IES	2.66e+08	8.54e+07	3.11	0.008	8.24e+07	4.49e+08
_cons	-1.20e+10	1.09e+10	-1.10	0.290	-3.53e+10	1.14e+10

El modelo muestra que el gasto de consumo de hogares y la tasa de inscripción escolar generan un efecto positivo sobre los ingresos netos de un país, más no así el porcentaje de trabajadores asalariados el cual no es significativo al 5% de nivel de confianza. Generalmente los modelos de ingresos, al presentar desigualdades o inequidades en la sociedad, presentan características heteroscedásticas por lo que se comprueba esta hipótesis bajo el test de Breusch-Pagan / Cook-Weisberg.

Comandos a utilizar:

**estat hettest:** Desarrolla el test de Breusch-Pagan / Cook-Weisberg para detectar heteroscedasticidad

Su estructura es:

**estat hettest**

Rutina Utilizada:

**reg IN GC TA IES**

**estat hettest**

Desarrollo:

Una vez aplicada la regresión y el comando de detección de heteroscedasticidad se analiza los siguientes datos.

```

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance
Variables: fitted values of IN

chi2(1)          =      3.92
Prob > chi2      =     0.0476

```

Interpretación:

Donde se aprecia que la hipótesis nula ( $H_0$ ) asume varianzas constantes, es decir homoscedasticidad. Si la probabilidad mayor a chi cuadrado ( $\text{Prob} > \chi^2$ ) es estadísticamente significativa al 5% se rechaza  $H_0$  y se acepta la alternativa de heteroscedasticidad, como en este caso la probabilidad es de 0.04 se rechaza  $H_0$  y se asevera que el modelo presenta características heteroscedásticas.

- **White-Koenker Test.**

El test desarrollado en 1981 por White & Koenker asume la distribución normal de los residuos bajo la hipótesis nula de homoscedasticidad, en caso de rechazarse  $H_0$  se asume la presencia de heteroscedasticidad en la regresión estimada.

### **Ejercicio 6:**

Con la información brindada por la tabla 3.9.3 referente a la Encuesta de Edificaciones 2013, se realiza el siguiente modelo econométrico:

$$VTE = \beta_1 + \beta_2 VTT + \beta_3 TP + \beta_4 UE + u_i \quad (3.9.3)$$

Donde el valor total de las edificaciones está en función del valor del terreno, el total de pisos construidos y el número de unidades de edificación. Para limitar el número de variables utilizadas en el modelo, se tomó en cuenta exclusivamente a las nuevas construcciones de la ciudad y provincia de Loja, resultando en:

Source	SS	df	MS	Number of obs =	77
Model	1.5784e+12	3	5.2613e+11	F( 3, 73) =	48.77
Residual	7.8759e+11	73	1.0789e+10	Prob > F =	0.0000
				R-squared =	0.6671
				Adj R-squared =	0.6534
Total	2.3660e+12	76	3.1131e+10	Root MSE =	1.0e+05

VTE	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
VTT	2.642256	.4800665	5.50	0.000	1.685485	3.599028
TP	44158.56	15315.44	2.88	0.005	13634.94	74682.18
UE	71987.66	24842.05	2.90	0.005	22477.52	121497.8
_cons	-147715.8	37498.19	-3.94	0.000	-222449.6	-72982.04

Una relación positiva y estadísticamente significativa de los determinantes del valor total de las edificaciones en la ciudad de Loja. Si aplicamos el test de White-Koenker se obtiene:

Comandos a utilizar:

**ivhettest:** Despliega un test de heteroscedasticidad bajo la verificación de hipótesis.

Su estructura es:

**ivhettest**

Rutina Utilizada:

**reg** VTE VTT TP UE

**ivhettest**

Desarrollo:

Luego de establecer la regresión y aplicarla, se ejecuta el comando para detectar heteroscedasticidad y se obtiene el siguiente recuadro:

```
OLS heteroskedasticity test(s) using levels of IVs only
Ho: Disturbance is homoskedastic
White/Koenker nR2 test statistic      :   9.240  Chi-sq(3) P-value = 0.0263
```

Al igual que las pruebas anteriores, la hipótesis nula se rechaza si el valor de la probabilidad es menor al 5%, en consiguiente si esta probabilidad es mayor a cinco se acepta la hipótesis de homoscedasticidad.

Interpretación:

Para el caso de las edificaciones en la ciudad de Loja, se evidencia la presencia de heteroscedasticidad en sus variables.

Otras medidas de detección de heteroscedasticidad como la Prueba de Park o la Goldfeld-Quandt se pueden realizar a través de varios procesos en stata al generar las variables residuales y correr distintas regresiones, por lo que se omiten en este caso dada la extensión de sus procesos y la aplicación de comandos ya analizados y que seguramente el estudiante domina para aplicar sin problema estas pruebas.

### **3.9.2 Corrección de Heteroscedasticidad.**

- **Robustecer los errores estándar y los coeficientes.**

#### **Ejercicio 7:**

Al existir limitantes como el desconocimiento de la varianza para realizar Mínimos Cuadrados Generalizados (MCG) o Mínimos Cuadrados Ponderados (MCP) dentro de las regresiones, se emplea la táctica más común para solucionar este problema al robustecer los errores estándar de los modelos al igual que sus coeficientes.

Comandos a utilizar:

**robust:** Es un comando que estima varianzas robustas de los estimadores  
**rreg:** Genera una versión de regresión robusta de la variable dependiente sobre las independientes

Su estructura es:

**reg** Y X1 X2..Xn, robust

**rreg** Y X1 X2..Xn

Rutina Utilizada:

**reg** VTE VTT TP UE, robust

**rreg** VTE VTT TP UE

Desarrollo:

Aplicando en primer comando al final de la regresión establecida se genera la siguiente regresión:

```
Linear regression                               Number of obs =          77
                                                F(   3,    73) =    19.65
                                                Prob > F       =    0.0000
                                                R-squared      =    0.6671
                                                Root MSE      =    1.0e+05
```

		Robust				
VTE	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
VTT	2.642256	.4826733	5.47	0.000	1.68029	3.604223
TP	44158.56	13689.76	3.23	0.002	16874.91	71442.21
UE	71987.66	57030.89	1.26	0.211	-41674.75	185650.1
_cons	-147715.8	70164.04	-2.11	0.039	-287552.6	-7879.098

Se observa mayor robustez en los errores estándar de los coeficientes, sin embargo en el caso de la variable unidades de edificación se perdió significancia.

Ahora empleando la regresión robusta:

```
Robust regression                               Number of obs =          75
                                                F(   3,    71) =   405.74
                                                Prob > F       =    0.0000
```

VTE	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
VTT	2.491172	.1418144	17.57	0.000	2.208403	2.773942
TP	34647.06	4501.044	7.70	0.000	25672.23	43621.88
UE	35507.02	30751.64	1.15	0.252	-25810.01	96824.06
_cons	-99711.05	32073.07	-3.11	0.003	-163662.9	-35759.18

Los resultados muestran una disminución significativa de los errores estándar, en comparación al modelo estimado por MCO pero con la misma característica del ajuste por robustecimiento anterior, que la variable de unidades edificadas pierde significancia.

De manera resumida se presentan estas correcciones para los ejercicios tratados en el apartado de heteroscedasticidad:

		<b>MCO</b>		<b>Robustecimiento de errores</b>		<b>Regresión robusta</b>	
Estadísticos		Std. Err.	P>[ t ]	Std. Err.	P>[ t ]	Std. Err.	P>[ t ]
	Edad	0.2172191	0.090	0.2209411	0.0960	0.3299159	0.1220
	NEA	0.0251105	0.059	0.0261124	0.0690	0.0466357	0.0590
	NCM	0.2470769	0.000	0.2846638	0.0000	2.709782	0.0000
	PROM	0.0969103	0.000	0.0923247	0.0000	0.8254369	0.0000
	GC	4099277	0.0000	3173259	0.0000	2042586	0.0000
	TA	1.54E+08	0.0930	1.71E+08	0.1250	7.70E+07	0.0000
	IES	8.54E+07	0.0080	6.18E+07	0.0010	4.26E+07	0.0000
	VTT	0.480067	0.0000	0.4826733	0.0000	0.141814	0.0000
	TP	15315.44	0.0050	13689.76	0.0020	4501.044	0.0000
	UE	71987.66	0.0050	57030.89	0.2110	30751.64	0.2520

Interpretación:

El mejor método de corrección de heteroscedasticidad, basándose en la robustez de los resultados y en la significancia de los coeficientes, es el de regresión robusta ya que esta opción mantiene los estimadores MELI mejorando los errores estándar y no altera en gran medida la significancia del modelo.

### 3.10. Autocorrelación

Se la define como la vinculación de variables con el espacio (autocorrelación espacial) o tiempo (autocorrelación serial) que invalidan el uso de la econometría clásica al romper el supuesto de no haber autocorrelación en las perturbaciones del modelo (Pineda, 2006) alega que si existe correlación entre dos miembros de una serie de observaciones ordenadas en el tiempo o espacio se presenta el problema de autocorrelación. Los ejercicios usados en este apartado hacen referencia al crecimiento económico vía producción petrolera, gasto de consumo final en función a la formación bruta de capital fijo y exportaciones, y diversos determinantes de la inversión extranjera directa en Ecuador.

### 3.10.1 Detección de Autocorrelación.

- **Métodos Gráficos.**

Existen varios recursos gráficos que ofrece el paquete Stata 13 para analizar la distribución de los residuos a través del tiempo, lo cual proporciona una perspectiva útil para detectar problemas de autocorrelación o incluso heteroscedasticidad.

#### Ejercicio 1:

Una regresión bastante simple que se suele realizar dentro del contexto ecuatoriano es la del producto interno bruto en función de los barriles de petróleo producidos, para este caso se analiza los datos desde el año 1972 hasta el 2011, tal como se aprecia en la tabla 3.10.1.

$$PIB = \beta_1 + \beta_2 Petroleo + u_t \quad (3.10.1)$$

Source	SS	df	MS	Number of obs = 40		
Model	3.7785e+21	1	3.7785e+21	F( 1, 38) =	346.43	
Residual	4.1447e+20	38	1.0907e+19	Prob > F =	0.0000	
				R-squared =	0.9012	
				Adj R-squared =	0.8986	
Total	4.1930e+21	39	1.0751e+20	Root MSE =	3.3e+09	

pib	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
petroleo	217854.5	11704.7	18.61	0.000	194159.6	241549.4
_cons	3.30e+09	1.49e+09	2.21	0.033	2.82e+08	6.33e+09

Tal como se esperaba el efecto que tiene la producción de petróleo sobre el Producto Interno Bruto es positivo y altamente significativo, además el modelo planteado posee un coeficiente de determinación superior al 90% y sus valores probabilísticos en t y F son menores al 5%.

Comandos a utilizar:

**rstandard y rstudent** : Son opciones del comando predict, que permiten estimar los errores estandarizados y estudentizados respectivamente.

Su estructura es:

**predict** Nombre de la variable, **rstandard**

**predict** Nombre de la variable, **rstudent**

Rutina Utilizada:

**tsset** year, yearly

**reg** pib petroleo

**predict resid**

**predict** rstand,rstandard

**predict** restudent, rstudent

**twoway** (line restudent rstand year)

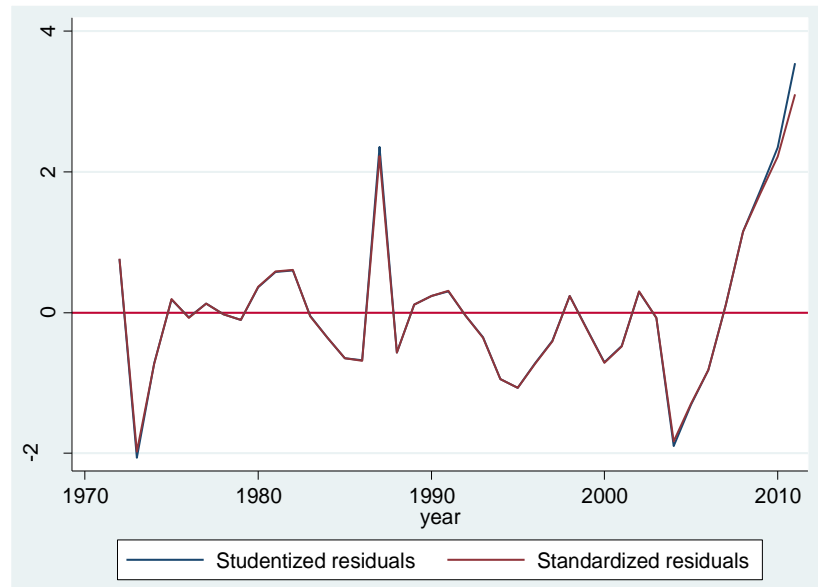
**gen** laags= rstand[\_n-1]

**twowav** (**scatter** rstand laags). **vline**(0) **xline**(0)

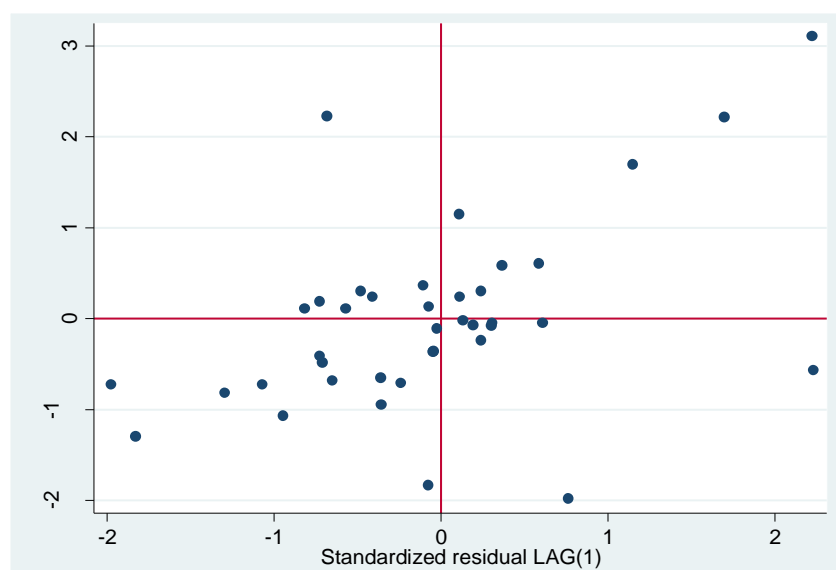


Desarrollo:

Para analizar la posible presencia de autocorrelación en este ejercicio se realiza una gráfica de residuos estudentizados y estandarizados en el tiempo, obteniendo:



A simple vista no se puede afirmar si la autocorrelación es positiva o negativa entre estos residuos, sin embargo una característica general de las series de tiempo es que suelen presentar autocorrelación positiva ya que sus valores tienden a fluctuar positiva o negativamente en el tiempo, por lo que se aplica un gráfico de dispersión entre los residuos estandarizados y los residuos estandarizados rezagados en un periodo de tiempo tal como se muestra a continuación:



Interpretación:

El primer gráfico no revela un patrón definido, por lo que es posible que los residuos no sean aleatorios. Sin embargo, la segunda gráfica comprueba la presencia de autocorrelación serial positiva en el modelo ya que los residuos se concentran en el primer y tercer cuadrante del plano cartesiano.

## **Ejercicio 2**

Comandos a utilizar:

**ac:** Genera un gráfico de autocorrelación con intervalos de confianza.

Su estructura es:

**ac** variable, lags(10)

Rutina Utilizada:

**tsset** year, yearly

**reg** pib petroleo

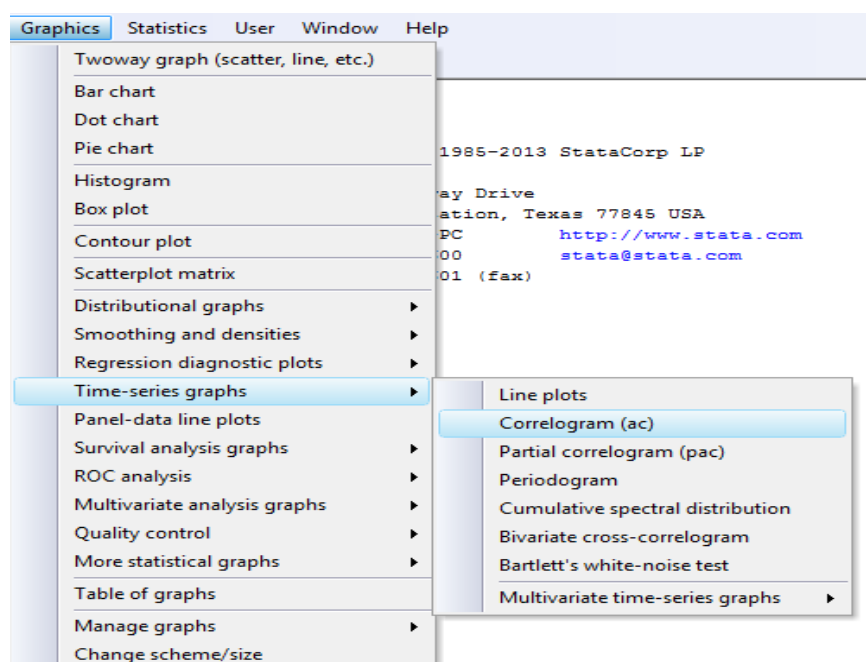
**predict** resid

**ac** resid, lags(10)

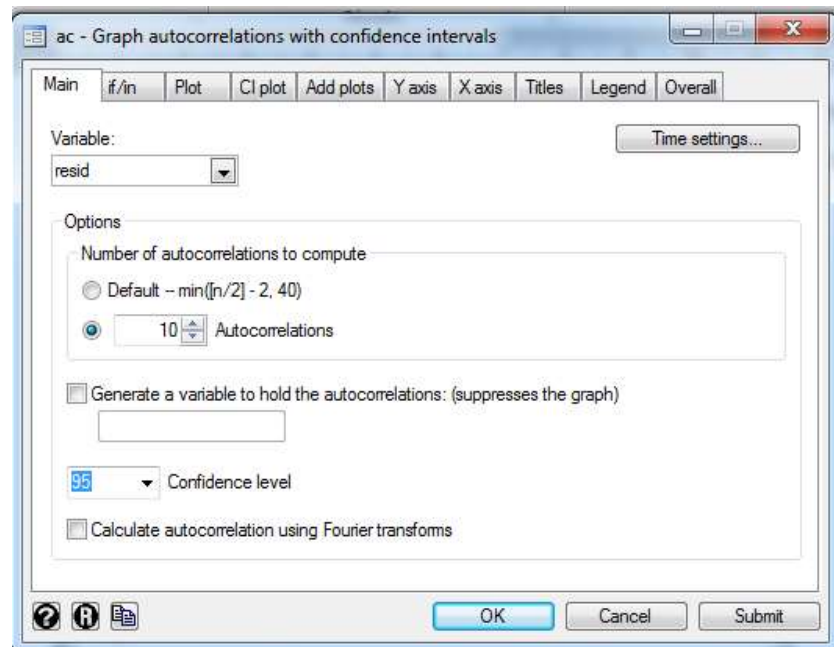
Desarrollo:

Otro método empleado en la detección de autocorrelación es el gráfico de correlograma con intervalos de confianza, el cual se genera bajo el comando descrito en el procedimiento utilizado bajo el siguiente procedimiento:

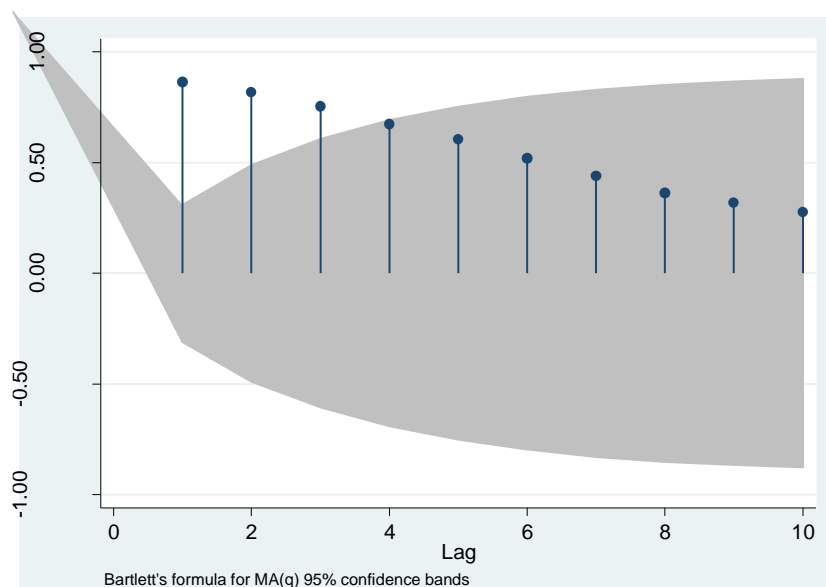
Desplegando las funciones de la barra de *Graphics*, se selecciona *Time-series graphs* y finalmente la opción de *Correlogram (ac)*.



A continuación se despliega un cuadro de opciones en donde se selecciona la variable a analizar, el número de rezagos a calcular, la elección de crear una variable de autocorrelación en lugar de un gráfico, el nivel de confianza y la opción de calcular la autocorrelación usando las transformaciones de Fourier.



Haciendo un clic en *OK* se obtiene el siguiente gráfico:



Interpretación:

Con el 95% de confianza se podría decir que en los tres primeros rezagos de los residuos estimados de la regresión existe autocorrelación, además se aprecia un patrón definido a lo largo del tiempo.

- **Durbin Watson.**

Arranz, J. M., & Zamora, M. M (2002) menciona que el contraste  $d$  de Durbin Watson desarrollado en 1951 es uno de los métodos más utilizados para detectar la presencia de autocorrelación en los errores de un modelo (Arranz & Zamora, 2002) mencionan que este test contrasta las hipótesis nula de no autocorrelación cuando  $d \approx 2$  y la hipótesis alterna de autocorrelación cuando  $d \approx 0/d \approx 4$ .

### **Ejercicio 3:**

El gasto de consumo final del gobierno y hogares ecuatorianos es analizado en función a la formación bruta de capital fijo (FBKF) y exportaciones de bienes y servicios (EXP) dentro del periodo 2002-2015, como lo refleja la tabla 3.10.2. La relación pretende analizar el efecto de la variable dependiente bajo una perspectiva keynesiana, cuya función de consumo es determinada por los ingresos disponibles. Para el caso ecuatoriano se tiene que los ingresos son principalmente las variables  $X_1$  y  $X_2$  como se muestra en la siguiente ecuación:

$$GCF = \beta_1 + \beta_2 FBKF + \beta_3 EXP + u_t \quad (3.10.2)$$

Source	SS	df	MS	Number of obs =	14
Model	4.1553e+15	2	2.0777e+15	F( 2, 11) =	602.54
Residual	3.7930e+13	11	3.4482e+12	Prob > F =	0.0000
Total	4.1933e+15	13	3.2256e+14	R-squared =	0.9910
				Adj R-squared =	0.9893
				Root MSE =	1.9e+06

GCF	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
FBKF	2.196312	.167714	13.10	0.000	1.827176	2.565448
EXP	.0037235	.1794565	0.02	0.984	-.3912575	.3987045
_cons	1.39e+07	1350007	10.31	0.000	1.09e+07	1.69e+07

Los resultados son acorde a la teoría macroeconómica establecida, dados los coeficientes positivos se tiene que ante una mayor inversión en formación bruta de capital fijo y exportaciones el consumo aumentará. Para detectar la presencia de autocorrelación se procede a aplicar el test de Durbin Watson.

Comandos a utilizar:

**estat dwatson**: Es un test estadístico que sirve para probar correlación serial de primer orden.

Su estructura es:

**estat dwatson**

Rutina Utilizada:

**tsset year, yearly**

**reg GCF FBKF EXP**

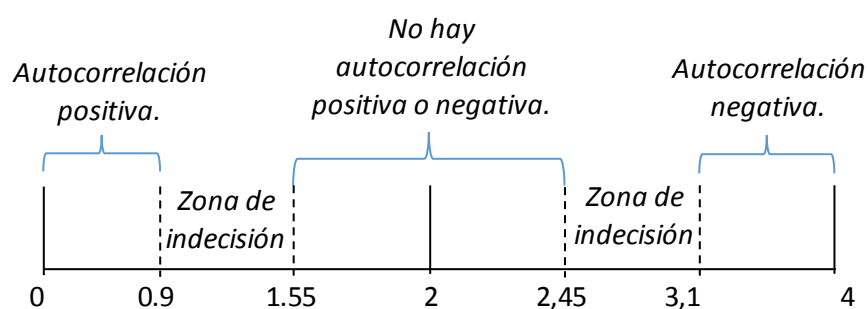
**estat dwatson**

Desarrollo:

En primera instancia se declara la base de datos como una serie de tiempo para poder efectuar el procedimiento, luego se realiza la regresión de la variable dependiente en función de las independientes y seguido de esto se aplica el comando que genera el valor  $d$  de Durbin Watson como se muestra a continuación:

```
Durbin-Watson d-statistic( 3, 14) = .8457392
```

Los criterios de decisión mencionados señalan que si el valor  $d$  es cercano a cero hay evidencia de autocorrelación positiva y si es cercano a 4 la autocorrelación es negativa, sin embargo es necesario saber los límites superiores e inferiores para poder tomar una decisión certera en cuanto a este valor, tal como se muestra en el siguiente cuadro de decisión.



Interpretación:

Los valores de los límites fueron obtenidos de la tabla D.5A del texto de Gujarati & Porter (2010), con  $n=14$ ,  $k=2$  y un 0.05 de significancia. Con estos límites establecidos se puede concluir la presencia de autocorrelación positiva en el ejercicio 3.

- **Prueba de Breusch-Godfrey.**

#### **Ejercicio 4:**

Esta prueba desarrollada en 1978 permite analizar procesos de autocorrelación en órdenes superiores, procesos de media móviles y procesos mixtos. En este apartado se utiliza el test para probar la existencia de autocorrelación en orden superior a uno AR(1) con los siguientes comandos:

Comandos a utilizar:

**estat bgodfrey:** Es un test de autocorrelación para distintos órdenes autorregresivos.

Su estructura es:

**estat bgodfrey, lags(n)**

Rutina Utilizada:

**tsset year, yearly**

**reg GCF FBKF EXP**

**estat bgodfrey, lags(1)**

**estat bgodfrey, lags(2)**

Desarrollo:

Se realiza el test de autocorrelación para los órdenes autorregresivos AR(1) y AR(2) y se obtiene el siguiente cuadro de resultados.

Test de Autocorrelación de primer orden:

Breusch-Godfrey LM test for autocorrelation

lags (p)	chi2	df	Prob > chi2
1	2.972	1	0.0847

H0: no serial correlation

Test de Autocorrelación de segundo orden:

Breusch-Godfrey LM test for autocorrelation

lags (p)	chi2	df	Prob > chi2
2	3.560	2	0.1687

H0: no serial correlation

Interpretación:

La prueba de autocorrelación de primer orden de Breusch-Godfrey para el modelo propuesto evidencia que al 5% de significancia se acepta la hipótesis nula de no correlación serial entre los residuos, sin embargo al 10% de significancia se puede rechazar  $H_0$  y por ende aceptar la presencia de autocorrelación de primer orden en el modelo. En el caso de la autocorrelación de segundo orden, se acepta la hipótesis nula y se confirma que no existe autocorrelación AR(2) en el ejercicio 3.

- **Otros métodos de detección.**

Existen muchas otras maneras de detectar la presencia de autocorrelación usando el software estadístico Stata 13, a continuación se mencionan los procesos más relevantes y de mayor uso por los investigadores para detectar el problema de autocorrelación.

### **Ejercicio 5:**

Graham (1992) menciona que uno de los determinantes fundamentales de la inversión extranjera directa es el control de los precios, basándose en este supuesto y utilizando los datos de la tabla 3.10.3 se aplica un modelo de inversión para el caso ecuatoriano del periodo 1976-2013 en donde se agregan variables adicionales que son características de la economía ecuatoriana, como son la educación, petróleo e industria con la siguiente ecuación:

$$IED = \beta_1 + \beta_2 IPC + \beta_3 GE + \beta_4 IND + \beta_5 RPET + u_t \quad (3.10.3)$$

Dónde:

IED= Inversión extranjera directa.

IPC= Índice de precios al consumidor.

GE= Gasto en educación

IND= Industrialización

RPET= Rentas del petróleo.

Comandos a utilizar:

Rutina Utilizada <b>tsset</b> year, <b>yearly</b> <b>reg</b> IED ipc GE IND RPET
--

Desarrollo:

La regresión efectuada muestra los siguientes resultados:

Source	SS	df	MS	Number of obs =	38
Model	9.9172783	4	2.47931957	F( 4, 33) =	3.14
Residual	26.0279498	33	.788725751	Prob > F =	0.0270
				R-squared =	0.2759
				Adj R-squared =	0.1881
Total	35.9452281	37	.971492651	Root MSE =	.8881

IED	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
ipc	.0048871	.0072502	0.67	0.505	-.0098635 .0196376
GE	-.175612	.0764318	-2.30	0.028	-.3311137 -.0201103
IND	-.0475504	.0282346	-1.68	0.102	-.1049941 .0098933
RPET	-.0558077	.0306921	-1.82	0.078	-.1182512 .0066358
_cons	2.551332	.5536325	4.61	0.000	1.424958 3.677706

Interpretación:

La estabilidad de precios, cuantificada por el IPC, no es un factor significativo que determine la inversión extranjera directa en el caso ecuatoriano; sin embargo un factor determinante es el gasto en educación cuyo efecto es negativo para la inversión, al igual que las rentas de petróleo y la industrialización cuya significancia es menor y su relación se mantiene inversa ante la variable dependiente.

- **Correlograma de los residuos.**

### **Ejercicio 6:**

Este método tabula y grafica la autocorrelación y autocorrelación parcial de los residuos para distintos órdenes regresivos, brindando estadísticos Q y sus respectivas probabilidades.

Comandos a utilizar:

**corrgram:** Genera una tabla de autocorrelación y correlación parcial para los órdenes regresivos establecidos.

Su estructura es:

**corrgram** resid, lags(n)

Rutina Utilizada

**tsset** year, **yearly**

**reg** IED ipc GE IND RPET

**predict** resid

**corrgram** resid, lags(15)



## Desarrollo:

Una vez efectuada la regresión, se predicen los valores residuales de la misma y se genera un correlograma de los residuos para analizar la posible presencia de autocorrelación, los resultados obtenidos son los siguientes:

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]			[Partial Autocor]		
1	0.4520	0.4548	8.3943	0.0038						
2	0.4624	0.3498	17.42	0.0002						
3	0.3511	0.1102	22.774	0.0000						
4	0.2692	-0.0178	26.014	0.0000						
5	0.0457	-0.2958	26.11	0.0001						
6	0.0604	-0.0242	26.283	0.0002						
7	-0.1577	-0.1821	27.502	0.0003						
8	-0.2719	-0.2672	31.248	0.0001						
9	-0.4045	-0.3097	39.822	0.0000						
10	-0.3998	-0.2717	48.5	0.0000						
11	-0.4001	-0.0555	57.51	0.0000						
12	-0.4016	-0.1998	66.938	0.0000						
13	-0.4005	-0.3419	76.689	0.0000						
14	-0.3208	-0.5360	83.206	0.0000						
15	-0.1150	0.2223	84.08	0.0000						

La tabla muestra los coeficientes de autocorrelación y autocorrelación parcial de los 15 rezagos, además del estadístico Q que es acumulativo a medida que aumentan los rezagos y valor probabilístico respectivo. Una gráfica de las autocorrelaciones se genera para analizar la tendencia que estos tienen a medida que se aumentan los rezagos y ver si su correlación es positiva o negativa.

## Interpretación:

La gráfica de la autocorrelación muestra un patrón cíclico, donde los primeros 6 rezagos son positivos y el resto son negativos, por su parte la autocorrelación parcial no muestra un patrón definido y sus valores fluctúan entre medidas de autocorrelación que incrementan y disminuyen, en este caso solo los tres primeros rezagos son positivos lo que puede brindar una idea de los órdenes regresivos en el que el modelo debe ser especificado en futuras regresiones. Finalmente se concluye la presencia de autocorrelación en los primeros rezagos del ejercicio 5 ya que sus coeficientes de autocorrelación y autocorrelación son elevados y sus probabilidades altamente significativas.

- **Matriz de correlación de los residuos.**

### **Ejercicio 7:**

Este método es bastante simple de aplicar, consiste en el análisis de los coeficientes de correlación entre los residuos y los rezagos que se generan de los mismos.

Comandos a utilizar:

```
Rutina Utilizada
reg IED ipc GE IND RPET
predict resid
gen R1=L1.resid
gen R2=L2.resid
gen R3=L3.resid
gen R4=L4.resid
corrgram resid
corr resid R1
corr resid R2
corr resid R3
corr resid R4
```

Desarrollo:

Una vez efectuada la regresión, se predicen los valores residuales de la misma y se generan los rezagos de los residuos en los órdenes deseados para comprobación, para este caso se establecen en: AR(1), AR(2), AR(3) y AR(4).

Correlación entre  $u$  y  $u_{t-1}$

	resid	R1
resid	1.0000	
R1	0.4543	1.0000

Correlación entre  $u$  y  $u_{t-2}$

	resid	R2
resid	1.0000	
R2	0.4939	1.0000

Correlación entre  $u$  y  $u_{t-3}$

	resid	R3
resid	1.0000	
R3	0.3885	1.0000

Correlación entre  $u$  y  $u_{t-4}$

	resid	R4
resid	1.0000	
R4	0.3018	1.0000

Interpretación:

Las matrices muestran valores de autocorrelación cercanos al 40% en los tres primeros rezagos residuales y una autocorrelación del 30% en los residuos rezagados en cuatro periodos, por tanto los valores más contundentes de autocorrelación se encuentran en AR(1), AR(2) y AR(3).

### 3.10.2 Corrección de Autocorrelación.

- Autocorrelación pura o mala especificación del modelo.

#### Ejercicio 8:

Para corregir la autocorrelación presente dentro del ejercicio 1 se comprueba la existencia de autocorrelación pura agregando la variable de tendencia o tiempo  $t$  dentro del modelo:

$$PIB = \beta_1 + \beta_2 Petroleo + \beta_3 t + u_t \quad (3.10.4)$$

Comandos a utilizar:

Rutina Utilizada:  
**gen** t= [\_n]  
**reg** pib petroleo t  
**estat** bgodfrey, lags(1)

Desarrollo:

Al generar la variable tiempo e introducirla en la ecuación, se genera la regresión siguiente:

Source	SS	df	MS	Number of obs = 40		
Model	4.0186e+21	2	2.0093e+21	F( 2, 37) = 426.21		
Residual	1.7443e+20	37	4.7143e+18	Prob > F = 0.0000		
Total	4.1930e+21	39	1.0751e+20	R-squared = 0.9584		
				Adj R-squared = 0.9562		
				Root MSE = 2.2e+09		

pib	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
petroleo	27461.33	27769.44	0.99	0.329	-28804.9	83727.56
t	7.66e+08	1.07e+08	7.14	0.000	5.48e+08	9.83e+08
_cons	1.04e+10	1.39e+09	7.44	0.000	7.53e+09	1.32e+10

Con el paso del tiempo, el PIB aumentó en 7.66e+06 unidades al año y el efecto del petróleo sobre la economía continuará siendo positivo pero insignificante. Para determinar si la autocorrelación fue eliminada con este ajuste, se utiliza la prueba de Breusch-Godfrey y se obtiene el siguiente resultado.

Breusch-Godfrey LM test for autocorrelation

lags (p)	chi2	df	Prob > chi2
1	34.417	1	0.0000

H0: no serial correlation

Interpretación:

Se rechaza la hipótesis nula de correlación serial inexistente y se confirma que el problema de autocorrelación sigue presente y se debe aplicar otros métodos correctivos.

### **Ejercicio 9:**

Si la autocorrelación pura, no es la causa de correlación entre los residuos del ejercicio 1 es posible que exista una mala especificación del modelo, por lo que se procede a generar los logaritmos de las variables y utilizar una ecuación doble-log.

$$LPIB = \beta_1 + \beta_2 LPET + u_t \quad (3.10.5)$$

Comandos a utilizar:

Rutina Utilizada:  
**gen** LPIB=log( pib)  
**gen** LPET =log( petroleo )  
**reg** LPIB LPET  
**estat** bgodfrey, lags(1)

Desarrollo:

Obteniendo los logaritmos de la variable dependiente e independiente, se genera la siguiente regresión:

Source	SS	df	MS	Number of obs = 40		
Model	4.59922369	1	4.59922369	F( 1, 38) =	296.59	
Residual	.589271906	38	.015507155	Prob > F =	0.0000	
Total	5.1884956	39	.133038349	R-squared =	0.8864	
				Adj R-squared =	0.8834	
				Root MSE =	.12453	

LPIB	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
LPET	.8028925	.046621	17.22	0.000	.7085133	.8972717
_cons	14.71734	.541652	27.17	0.000	13.62082	15.81385

Se aprecia que los coeficientes son robustos y altamente significativos, por lo que a *priori* es posible que el problema de autocorrelación esté solucionado, por lo que se aplica el test de Breusch-Godfrey y se obtiene:

Breusch-Godfrey LM test for autocorrelation

lags (p)	chi2	df	Prob > chi2
1	0.576	1	0.4480

H0: no serial correlation

Interpretación:

La hipótesis nula se acepta ya que la probabilidad del 44.8% es mayor al 5% establecido y se concluye que la mala especificación del modelo fue la causa de autocorrelación.<sup>15</sup>

- **Método de mínimos cuadrados generalizados.**

### **Ejercicio 10:**

Los estimadores obtenidos por este método son MELI a diferencia de los obtenidos por MCO, la falta de eficiencia de estos últimos estimadores suele ser la causa de autocorrelación en el modelo, por lo que aplicamos los mínimos cuadrados generalizados de Prais–Winsten y Cochrane–Orcutt para dar solución al ejercicio 3 del presente apartado.

Comandos a utilizar:

**prais:** Genera una regresión por MCG calculando automáticamente el valor  $\rho(p)$  y bajo el orden regresivo AR(1).

Su estructura es:

**prais** Y X1 X2..Xn

Rutina Utilizada

**tsset** year, **yearly**

**prais** GCF FBKF EXP

Desarrollo:

Utilizando los comandos anteriormente descritos y después de 14 iteraciones para obtener el valor  $\rho$ , se obtiene la siguiente regresión:

<sup>15</sup> Otros test de autocorrelación como el de Durbin Watson, correlogramas y demás pruebas gráficas fueron empleadas en este ejercicio y se concluyeron los mismos resultados.

Prais-Winsten AR(1) regression -- iterated estimates

Source	SS	df	MS	Number of obs =	14
Model	4.6916e+14	2	2.3458e+14	F( 2, 11) =	113.21
Residual	2.2793e+13	11	2.0721e+12	Prob > F =	0.0000
				R-squared =	0.9537
				Adj R-squared =	0.9452
Total	4.9195e+14	13	3.7842e+13	Root MSE =	1.4e+06

GCF	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
FBKF	2.457839	.169037	14.54	0.000	2.085791	2.829887
EXP	-.2631332	.1514773	-1.74	0.110	-.5965325	.070266
_cons	1.42e+07	1969139	7.22	0.000	9875762	1.85e+07
rho	.651565					

Durbin-Watson statistic (original) 0.845739

Durbin-Watson statistic (transformed) 1.142547

La significancia de los coeficientes es mucho mejor a la de la regresión original, además como se aprecia en la parte inferior de la tabla, los valores  $d$  de Durbin Watson son estimados automáticamente para compararlos dentro del cuadro de decisión y comprobar si los MCG dieron solución al problema de autocorrelación.

Interpretación:

El valor  $d$  equivalente a 1.14 se encuentra en la zona de indecisión de autocorrelación positiva, por lo que no se puede asegurar la completa solución del problema de autocorrelación a pesar de que este valor está muy cercano a la zona de rechazo.

### **Ejercicio 11:**

Continuando con los datos de la tabla 3.10.1, tenemos que cuando no se conoce el valor de  $\rho$ , un método muy utilizado es el de las primeras diferencias, donde se recalca la ausencia del intercepto en la ecuación como se muestra a continuación:

$$\Delta GCF = \Delta \beta_2 FBKF + \Delta \beta_3 EXP + \Delta u_t \quad (3.10.6)$$

Comandos a utilizar:

Rutina Utilizada:  
**tsset** year, **yearly**  
**gen** D1GFC=D1.GCF  
**gen** D1FBKF =D1.FBKF  
**gen** D1EXP =D1.EXP  
**reg** D1GFC D1FBKF D1EXP, **noconstant**

Desarrollo:

El resultado del método de primeras diferencias es el siguiente:

Source	SS	df	MS	Number of obs = 13		
Model	2.1992e+14	2	1.0996e+14	F( 2, 11) = 61.92		
Residual	1.9534e+13	11	1.7759e+12	Prob > F = 0.0000		
Total	2.3945e+14	13	1.8420e+13	R-squared = 0.9184		
				Adj R-squared = 0.9036		
				Root MSE = 1.3e+06		

D1GFC	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
D1FBKF	2.32894	.2436236	9.56	0.000	1.792728	2.865152
D1EXP	-.3060161	.1355081	-2.26	0.045	-.6042674	-.0077649

Los coeficientes se muestran significativos y con un cambio de signo en su constante. Para saber si este método es adecuado, Maddala propuso una rudimentaria regla para aplicar las primeras diferencias como solución al problema de autocorrelación, este establece que el presente método se debería aplicar cuando el valor  $d$  es alto (superior a 0.8). El resultado del test de Durbin Watson con la regresión de primeras diferencias es el siguiente:

$$\text{Durbin-Watson } d\text{-statistic}(2, 13) = .50302$$

Interpretación:

El método de primeras diferencias no resulta conveniente en este ejercicio, la autocorrelación positiva en los residuos permanece.

- **Método Newey West.**

### Ejercicio 12:

Este método genera una regresión con errores estándar consistentes con heteroscedasticidad y autocorrelación (errores CHA), además este método asume la presencia de autocorrelación y heteroscedasticidad en el modelo.

Con el fin de corregir la autocorrelación presente en el ejercicio 5 se procede a aplicar la regresión con errores estándar de Newey-West con la siguiente rutina de comandos:

Comandos a utilizar:

**newey**: Genera una regresión con errores estándar ajustados a la heteroscedasticidad y autocorrelación.

Su estructura es:  
**newey** Y X1 X2...Xn, lag(n)

Rutina Utilizada:  
**tsset** year, **yearly**  
**newey** IED ipc GE IND RPET, lag(1)  
**predict** NR1, **resid**  
**corrgram** NR1, lags(15)

Desarrollo:

Esta prueba es sugerida para muestras grandes, por lo que las 38 observaciones del ejercicio son un número considerable de datos y se procede a aplicar este método de corrección de autocorrelación resultando en:

Regression with Newey-West standard errors	Number of obs	=	38
maximum lag: 1	F( 4, 33)	=	3.43
	Prob > F	=	0.0189

IED	Newey-West					[95% Conf. Interval]
	Coef.	Std. Err.	t	P> t		
ipc	.0048871	.0070199	0.70	0.491	-.009395	.0191692
GE	-.175612	.0697096	-2.52	0.017	-.3174373	-.0337867
IND	-.0475504	.0249052	-1.91	0.065	-.0982204	.0031196
RPET	-.0558077	.0390492	-1.43	0.162	-.1352538	.0236384
cons	2.551332	.8761392	2.91	0.006	.7688133	4.33385

Los valores de F siguen siendo significativos al 5% al igual que el coeficiente de gasto en educación, la industrialización ganó significancia en el modelo al contrario de las rentas de petróleo que se muestran insignificantes. Los errores estándar en general son más robustos por lo que se procede a comprobar la presencia de autocorrelación en este nuevo modelo.



LAG	AC	PAC	Q	Prob>Q	-1 [Autocorrelation]	0 [Partial Autocor]	1
1	0.4511	0.4512	8.3601	0.0038			
2	0.2271	0.0728	10.538	0.0051			
3	0.3562	0.3233	16.048	0.0011			
4	0.3135	0.1198	20.442	0.0004			
5	0.3305	0.2626	25.472	0.0001			
6	0.3008	0.1255	29.77	0.0000			
7	0.1451	-0.0474	30.802	0.0001			
8	0.0652	-0.0936	31.018	0.0001			
9	0.1987	0.3276	33.088	0.0001			
10	0.0699	-0.1281	33.353	0.0002			
11	-0.1144	-0.1319	34.09	0.0003			
12	-0.1107	-0.4095	34.806	0.0005			
13	-0.1874	-0.5492	36.941	0.0004			
14	-0.0622	-0.1709	37.186	0.0007			
15	-0.0198	-0.0474	37.212	0.0012			

Interpretación:

La tendencia cíclica que mostraba el modelo original se desvanece en cierta medida al tener coeficientes que fluctúan en valores aleatorios, tanto positivos como negativos. La autocorrelación del primer rezago permanece con un coeficiente elevado y estadísticamente significativo por lo que se concluye que persiste el problema de autocorrelación de orden regresivo AR (1).

### 3.11. Creación de modelos econométricos.

La economía es una amplia ciencia que combina el ámbito social, demográfico, matemático, estadístico, monetario entre otros, para dar explicación a los comportamientos y fenómenos de la sociedad, dentro de este campo la econometría surge como una herramienta de suma importancia para cuantificar, analizar y predecir los acontecimientos que se presenten en el tiempo y espacio. La creación de los modelos econométricos se ven sujetos a los supuestos del modelo clásico de regresión lineal, con el fin de obtener los mejores estimadores que sean confiables para la toma de decisiones.

Gujarati & Porter (2010) mencionan algunos de los errores más comunes cuando se desarrollan modelos econométricos, entre los cuales se tiene:

- Omisión de una variable relevante.
- Inclusión de una variable innecesaria.
- Adopción de la forma funcional incorrecta.
- Errores de medición.
- Suposición de que el término de error está normalmente distribuido.

A continuación se analizan algunas de las pruebas para detectar errores de especificación dentro del modelo y las medidas correctivas pertinentes para cada caso. Las regresiones a utilizar abordan temas como el comercio de las empresas, un enfoque ampliado de los determinantes del valor agregado bruto cantonal y la producción de arroz en el Ecuador.

### 3.11.1 Pruebas de errores de especificación.

- **Sobreajuste de un modelo.**

Este problema surge cuando se agregan dentro del modelo variables innecesarias o superfluas, los estimadores obtenidos serán insesgados pero ineficientes, por lo que es necesario ajustar el modelo a estimadores más consistentes.

#### **Ejercicio 1:**

En base a la Encuesta de Comercio 2013 proporcionada por el INEC, se tomaron los datos de la provincia de Zamora para analizar la producción de sus empresas en función al total de empleadas mujeres, total de empleados hombres y el total de remuneraciones como se aprecia en la tabla 3.11.1, la ecuación generada es la siguiente:

$$PT = \beta_1 + \beta_2 THE + \beta_3 TME + \beta_4 TR + u_i \quad (3.11.1)$$

Los resultados obtenidos se presentan a continuación:

Source	SS	df	MS	Number of obs = 42		
Model	2.1152e+13	3	7.0507e+12	F( 3, 38) = 55.57		
Residual	4.8210e+12	38	1.2687e+11	Prob > F = 0.0000		
Total	2.5973e+13	41	6.3349e+11	R-squared = 0.8144		
				Adj R-squared = 0.7997		
				Root MSE = 3.6e+05		

PT	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
TEM	114.8517	7155.257	0.02	0.987	-14370.21	14599.91
THE	14228.73	6401.043	2.22	0.032	1270.496	27186.97
TR	2.192871	.4361341	5.03	0.000	1.309964	3.075778
_cons	-90810.75	97563.18	-0.93	0.358	-288317.1	106695.6

Los coeficientes de THE y TR son significativos al nivel del 5%, sin embargo TEM presenta un valor en *t* cercano a cero y una probabilidad sumamente alta. Para analizar la inserción de variables innecesarias en el modelo se procede a aplicar la prueba F.

Comandos a utilizar:

```
Rutina Utilizada:  
reg PT TEM THE TR  
test TEM  
test THE  
test TR
```

Desarrollo:

Se aplica la prueba F a cada una de las variables explicativas del modelo, sus valores probabilísticos demostrarán sus pertinencia dentro de la regresión establecida.

Total de empleadas Mujeres

( 1) TEM = 0

F( 1, 38) = 0.00  
Prob > F = 0.9873

Total de empleados hombres

( 1) THE = 0

F( 1, 38) = 4.94  
Prob > F = 0.0322

Total de remuneraciones

( 1) TR = 0

F( 1, 38) = 25.28  
Prob > F = 0.0000

Interpretación:

En el caso de la variable TME el valor probabilístico de la regresión y del test F es sumamente alto, por lo que se considera que su presencia es irrelevante en el modelo, el resto de variables explicativas son significativas al 0.05 de probabilidad dentro del test F por lo que se mantienen dentro de la ecuación.

### **Ejercicio 2:**

En consecuencia el nuevo modelo de producción para la provincia de Zamora es el siguiente:

$$TR = \beta_1 + \beta_2 THE + \beta_3 TME + \beta_4 PT + u_i \quad (3.11.2)$$

Comandos a utilizar:

```
Rutina Utilizada:  
reg PT THE TR
```

Desarrollo:

Ejecutando el comando descrito en el cuadro anterior la regresión obtenida es:

Source	SS	df	MS	Number of obs =	42
Model	2.1152e+13	2	1.0576e+13	F( 2, 39) =	85.56
Residual	4.8211e+12	39	1.2362e+11	Prob > F =	0.0000
				R-squared =	0.8144
				Adj R-squared =	0.8049
Total	2.5973e+13	41	6.3349e+11	Root MSE =	3.5e+05

PT	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
THE	14238.93	6287.247	2.26	0.029	1521.774	26956.09
TR	2.194304	.4213886	5.21	0.000	1.341965	3.046643
_cons	-90309.15	91230.8	-0.99	0.328	-274840.9	94222.57

Interpretación:

Pese a la eliminación de una variable el coeficiente de determinación  $R^2$  mantiene su valor de 81.4%, la significancia de las variables explicativas se mantiene y los coeficientes son robustos y acordes a los esperado.

- **Subajuste de un modelo.**

En base a la teoría y evidencia empírica se pueden construir un sin número de modelos que expliquen el funcionamiento de la economía dentro de una sociedad, sin embargo las teorías y enfoques para un mismo tema son diversas y todas toman validez en distintos casos. Bajo esta premisa se realizó el ejercicio 1 del aparatado de multicolinealidad utilizando algunas de las variables de la tabla 3.8.1 bajo enfoques de crecimiento económico como los de Thirlwall (1979) que sostiene que si en el largo plazo el equilibrio de la balanza de pagos de la cuenta corriente es necesario, y el tipo de cambio real se mantiene relativamente constante, el crecimiento a largo plazo de una nación se puede explicar por la relación entre el crecimiento de las exportaciones a la elasticidad ingreso de la demanda de importaciones y los aportes de Summer, Heston, Barro y Lee (1990) sobre desarrollo y capital humano.

**Ejercicio 3:**

$$LVAB = \beta_1 + \beta_2 LEST + \beta_3 LIMP + u_i \quad (3.11.3)$$

El modelo presentado anteriormente mostraba los siguientes coeficientes:

Source	SS	df	MS	Number of obs =	219
Model	376.229305	2	188.114652	F( 2, 216) =	565.98
Residual	71.7920909	216	.332370791	Prob > F =	0.0000
				R-squared =	0.8398
				Adj R-squared =	0.8383
Total	448.021396	218	2.05514402	Root MSE =	.57652

LVAB	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
LEST	.8501656	.0509463	16.69	0.000	.7497501	.9505811
LIMP	.1099233	.0217803	5.05	0.000	.0669942	.1528525
_cons	1.3136	.3240766	4.05	0.000	.6748425	1.952357

A simple vista el modelo muestra robustez en sus coeficientes al ser altamente significativos y con errores estándar pequeños, además su coeficiente de determinación es superior al 80% por lo que parecería que este modelo explica al crecimiento económico cantonal del Ecuador. Sin embargo, cuando se emplea el test de Ramsey desarrollada en 1969 para regresiones con errores de especificación (RESET) para variables omitidas, los resultados muestran una realidad diferente.

Comandos a utilizar:

**ovtest:** Desarrolla un test para variables omitidas.

Su estructura es:

**estat ovtest**

Rutina Utilizada:

**reg LVAB LEST LIMP**

**estat ovtest**

Desarrollo:

Luego de realizar la regresión logarítmica y aplicar el test de Ramsey, se tiene el siguiente resultado:

```
Ramsey RESET test using powers of the fitted values of LVAB
Ho: model has no omitted variables
      F(3, 213) =      6.22
      Prob > F =      0.0005
```

Los criterios de aceptación de esta prueba de variables omitidas estable que: si  $Prob > F < 0.05$  se rechaza  $H_0$ , por tanto existen variables omitidas en el modelo y si  $Prob > F > 0.05$  se acepta  $H_0$  concluyendo que el modelo no tiene variables omitidas.

Interpretación:

El ejercicio analizado tiene variables omitidas dentro de su modelo de crecimiento económico, por tanto es necesario añadir variables explicativas que sustenten la teoría económica y den mayor solidez al modelo econométrico.

### 3.11.2 Corrección del subajuste de un modelo.

En función a la disponibilidad de datos y a la teoría económica pertinente se insertan dos nuevas variables al ejercicio anteriormente propuesto, con esta modificación se genera la siguiente ecuación de crecimiento para los cantones del Ecuador.

#### Ejercicio 4:

$$LVAB = \beta_1 + \beta_2 LEST + \beta_3 LIMP + \beta_4 LEXP + \beta_5 DP + u_i \quad (3.11.4)$$

Dónde:

LEXP: Son los valores logarítmicos del recuento de empresas exportadores en cada cantón.

DP: Es una variable dicotómica de polos cantonales, donde Cuenca, Guayaquil y Quito tienen valores de 1 y el resto de cantones son asignados con 0.

Comandos a utilizar:

```
Rutina Utilizada:  
gen LEXP= log(1+EXP)*  
reg LVAB LEST LIMP LEXP DP  
estat ovtest
```

Desarrollo:

Agregando las variables de exportación de empresas y la dicótoma de polos cantonales se obtiene los siguientes resultados:

---

\* Con el fin de no perder datos dentro de la variable EXP que registra información de cero para ciertos cantones (no existe logaritmo de 0), se añade una unidad a cada valor de esta variable de manera que no se perderán datos ni se alterarán los valores originales, puesto que el logaritmo de 1 es cero.

Source	SS	df	MS	Number of obs =	219
Model	384.803707	4	96.2009268	F( 4, 214) =	325.65
Residual	63.2176886	214	.29540976	Prob > F =	0.0000
				R-squared =	0.8589
				Adj R-squared =	0.8563
Total	448.021396	218	2.05514402	Root MSE =	.54352

LVAB	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
LEST	.6839234	.0579773	11.80	0.000	.5696437	.7982031
LIMP	.2148865	.0402312	5.34	0.000	.1355864	.2941867
LEXP	.1053844	.0438628	2.40	0.017	.0189258	.1918429
DP	-2.632818	.628177	-4.19	0.000	-3.871025	-1.394612
_cons	1.47148	.3893279	3.78	0.000	.7040711	2.238888

El valor de  $R^2$  incrementa ligeramente dentro del nuevo modelo, los coeficientes mantienen la coherencia económica y se muestran altamente significativos. La aplicación del test de Ramsey se muestra a continuación:

```
Ramsey RESET test using powers of the fitted values of LVAB
Ho: model has no omitted variables
F(3, 211) = 0.43
Prob > F = 0.7286
```

Interpretación:

El modelo planteado no presenta variables omitidas, por tanto se puede asegurar que los coeficientes de la regresión reflejan la situación económica de los cantones del Ecuador, de manera que el incremento en las variables de escolaridad, impuestos y exportaciones de empresas genera mayor valor agregado bruto mientras que la concentración de recursos en los cantones de Quito, Guayaquil y Cuenca crean un efecto contrario al crecimiento.

### 3.11.3 Test de especificación.

Con el fin de validar los resultados obtenidos en una regresión se realizan varias pruebas de especificación dentro de las variables regresoras como regresadas, de esta manera se espera mantener el supuesto de correcta especificación del modelo.

- **Prueba de Enlace.**

Este test fue implementado bajo las ideas de Turkey(1949) y Pregibon(1979) con el fin de determinar un enlace de prueba en la especificación de la variable dependiente en ecuaciones simples.

### Ejercicio 5:

Dentro de este apartado se utilizó la información brindada en la tabla 3.11.2 referente a la producción de arroz en el Ecuador en los años 2000 al 2012, con la superficie sembrada y el rendimiento como variables explicativas. Su ecuación es la siguiente:

$$PA = \beta_1 + \beta_2 SS + \beta_3 R + u_t \quad (3.11.5)$$

Comandos a utilizar:

**linktest:** Genera una prueba para especificación del modelo luego de cada estimación.

Su estructura es:

**linktest**

Rutina Utilizada

**tsset** YEAR, yearly

**reg** PA SS R

**linktest**

Desarrollo:

Ejecutando la regresión anteriormente descrita se obtienen los siguientes resultados:

Source	SS	df	MS	Number of obs = 13		
Model	7.4147e+11	2	3.7073e+11	F( 2, 10) = 215.63		
Residual	1.7193e+10	10	1.7193e+09	Prob > F = 0.0000		
Total	7.5866e+11	12	6.3222e+10	R-squared = 0.9773		
				Adj R-squared = 0.9728		
				Root MSE = 41464		

PA	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
SS	3.523789	.7737661	4.55	0.001	1.79973	5.247847
R	294868.1	29239.66	10.08	0.000	229718.1	360018.2
_cons	-1122739	227929	-4.93	0.001	-1630596	-614881.1

Seguido de esto, se ejecuta el test de enlace con el comando **linktest** obteniendo:

Source	SS	df	MS	Number of obs = 13		
Model	7.4810e+11	2	3.7405e+11	F( 2, 10) = 354.35		
Residual	1.0556e+10	10	1.0556e+09	Prob > F = 0.0000		
Total	7.5866e+11	12	6.3222e+10	R-squared = 0.9861		
				Adj R-squared = 0.9833		
				Root MSE = 32490		

PA	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
_hat	-.3212113	.5282522	-0.61	0.557	-1.498231	.8558081
_hatsq	5.21e-07	2.08e-07	2.51	0.031	5.81e-08	9.84e-07
_cons	803575.4	323661.4	2.48	0.032	82412.92	1524738



Si el valor probabilístico de las variables presentadas por este test es menor a 0.05, su inserción dentro del modelo debería ser considerada, caso contrario las variables serían irrelevantes.

Interpretación:

Se muestra que la predicción al cuadrado de la variable independiente tiene un poder explicativo dentro del modelo, por lo que la especificación del modelo no es del todo correcta.

### **Ejercicio 6**

Para corregir la mala especificación del modelo evidenciado en el ejercicio anterior, se procede transformar la variable dependiente bajo el siguiente razonamiento:

El test de enlaces demostró que el cuadrado de la producción de arroz explica a la producción de arroz, por lo que para esta demostración omitimos temporalmente las demás variables explicativas y expresamos lo siguiente:

$$PA = PA^2 \quad (3.11.6)$$

Entonces

$$\frac{PA}{PA^2} \quad (3.11.7)$$

Simplificando obtenemos:

$$\frac{1}{PA} \quad (3.11.8)$$

De manera que la variable explicativa es transformada a 1/PA y la nueva ecuación es la siguiente:

$$PA_{\text{Arroz}} = \beta_1 + \beta_2 SS + \beta_3 R + u_t \quad (3.11.9)$$

Comandos a utilizar:

<pre>Rutina Utilizada <b>tsset</b> YEAR, yearly <b>gen</b> PArroz=1/PA <b>reg</b> PArroz SS R</pre>
---

Desarrollo:

Aplicando la rutina especificada se obtiene el siguiente resultado:

Source	SS	df	MS	Number of obs = 13		
Model	3.0964e-13	2	1.5482e-13	F( 2, 10) = 397.87		
Residual	3.8913e-15	10	3.8913e-16	Prob > F = 0.0000		
				R-squared = 0.9876		
				Adj R-squared = 0.9851		
Total	3.1353e-13	12	2.6128e-14	Root MSE = 2.0e-08		

PArroz	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
SS	-2.88e-12	3.68e-13	-7.81	0.000	-3.70e-12	-2.06e-12
R	-1.70e-07	1.39e-08	-12.21	0.000	-2.01e-07	-1.39e-07
_cons	2.52e-06	1.08e-07	23.27	0.000	2.28e-06	2.76e-06

Interpretación:

Se aprecia una leve mejora en la robustez de los coeficientes, por lo que el ajuste en la especificación del modelo se muestra pertinente.

### 3.11.4 Normalidad en las variables y los residuos.

Continuando con los datos de la tabla 3.11.2 se procede a analizar la normalidad de las variables explicativas como de los residuos obtenidos en el ejercicio anterior, a través de métodos gráficos y cuantitativos.

- **Densidad de Kernel.**

Especifica la función de kernel para calcular la densidad estimada mediante una representación gráfica, entre las principales ventajas de este método se encuentra la eliminación de la discontinuidad y el hecho de que su estimación no depende del punto de origen.

### Ejercicio 7

Comandos a utilizar:

**kdensity:** Genera un gráfico de densidad de las variables .

**normal:** Es una opción del comando **kdensity** que superpone la densidad normal de la variable seleccionada.

Su estructura es:

**Kdensity, normal**

Rutina Utilizada

**tsset** YEAR, yearly

**reg** PArroz SS R

**predict** res1, resid

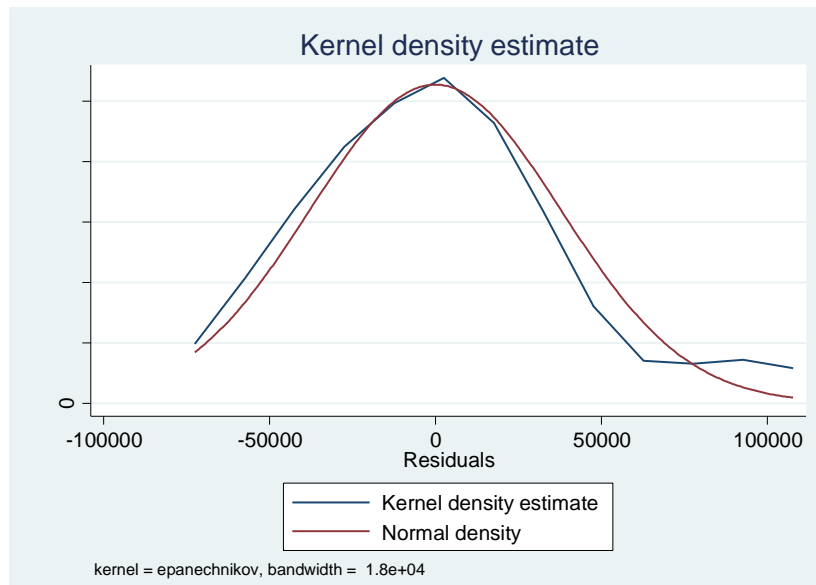
**kdensity** res1, normal

**kdensity** SS, normal

**kdensity** R, normal

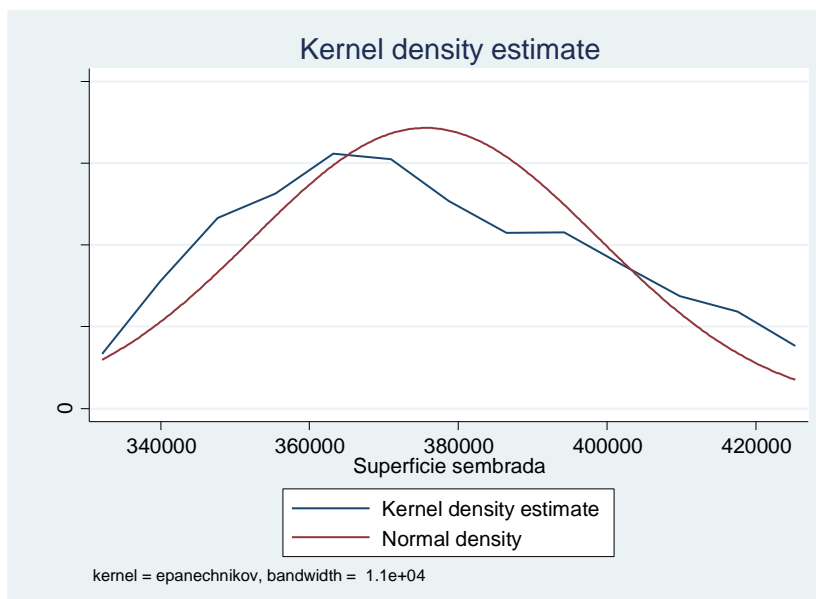
Desarrollo:

Al obtener los residuos de la regresión se aplica el comando **kdensity** con la opción de normalidad para tener un gráfico superpuesto de la densidad normal de los residuos y la densidad estimada de los mismos.

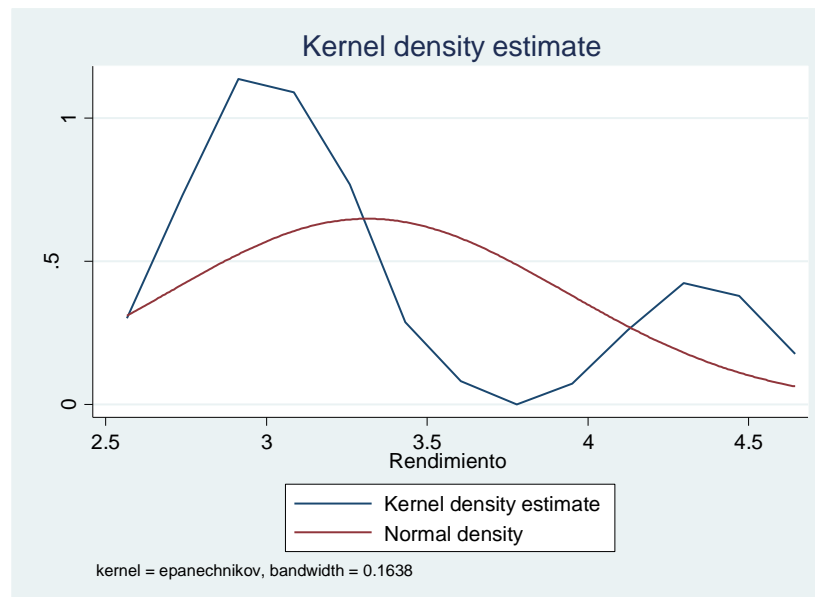


Esta opción no se limita al contraste de normalidad de los residuos ya que puede ser aplicada a las demás variables del modelo, en este caso es usado en las variables explicativas obteniendo los siguientes gráficos:

Superficie Sembrada :



Rendimiento:



Interpretación:

Para el caso de los residuos (*res1*) y la variable superficie sembrada (*SS*) se aprecia una distribución normal de sus datos, ya que se ajustan en gran medida a la densidad normal de kernel. De manera contraria, la variable rendimiento (*R*) dista mucho del ajuste propuesto ya que su estimación (representada por la línea azul) forma un patrón anormal.

- **Test de Curtosis y Asimetría.**

Es un test de normalidad que presenta la curtosis y la asimetría de los datos seleccionados para luego combinarlos es un test estadístico *t*.

### **Ejercicio 8**

Comandos a utilizar:

**sktest:** Presenta una tabla de normalidad para las variables especificadas.

Su estructura es:

**sktest** Variable1 Variable2... Variable *n*

Rutina Utilizada

**tsset** YEAR, yearly

**reg** PArroz SS R

**predict** res1, **resid**

**sktest** res1 SS R

Desarrollo:

La tabla de asimetría y curtosis se muestra a continuación:

Skewness/Kurtosis tests for Normality					
Variable	Obs	Pr (Skewness)	Pr (Kurtosis)	adj chi2 (2)	joint Prob>chi2
res1	13	0.1603	0.1944	4.00	0.1350
SS	13	0.4860	0.3924	1.37	0.5034
R	13	0.0570	0.9213	3.99	0.1362

Interpretación:

El valor probabilístico de la asimetría refleja que *res1* y *SS* tiene una distribución normal, mientras que *R* muestra con un 5.7% de confianza que existe una distribución asimétrica de sus datos. La curtosis de estas tres variables posee valores superiores al 5% o 10% de confianza por lo que no se rechaza la hipótesis de que los residuos, la superficie sembrada y el rendimiento estén normalmente distribuidos.

### 3.12. Modelos de regresión no lineales

Dentro de la teoría económica las relaciones que surgen entre variables no son siempre lineales, es por ello que en el presente apartado se abordan algunos ejercicios que ejemplifican la cuantificación de este tipo de modelos econométricos. Se utilizan datos de crecimiento económico, poblacional y medio ambiente para mostrar relaciones polinomiales y logarítmicas que brinden una perspectivas más amplia de los modelos no lineales.

#### 3.12.1 Estimación de modelos de regresión lineal y no lineal.

Para estimar los supuestos del ganador del premio Nobel, Simon Kuznets sobre la relación entre crecimiento económico y deterioro ambiental y su forma de U invertida en el tiempo se utilizan los siguientes modelos:

Modelo Lineal:

$$CO_2 = \beta_1 + \beta_2 PIBPC + u_t \quad (3.12.1)$$

Modelo Polinomial:

$$CO_2 = \beta_1 + \beta_2 PIBPC - \beta_3 PIBPC^2 + u_t \quad (3.12.2)$$

### Ejercicio 1

Utilizando los datos de la tabla 3.3.4 en la función lineal anteriormente propuesta se tiene:

Comandos a utilizar:

Rutina Utilizada

```
tsset year, yearly
reg CO2 PIBPC
scatter CO2 PIBPC, connect(1)
```

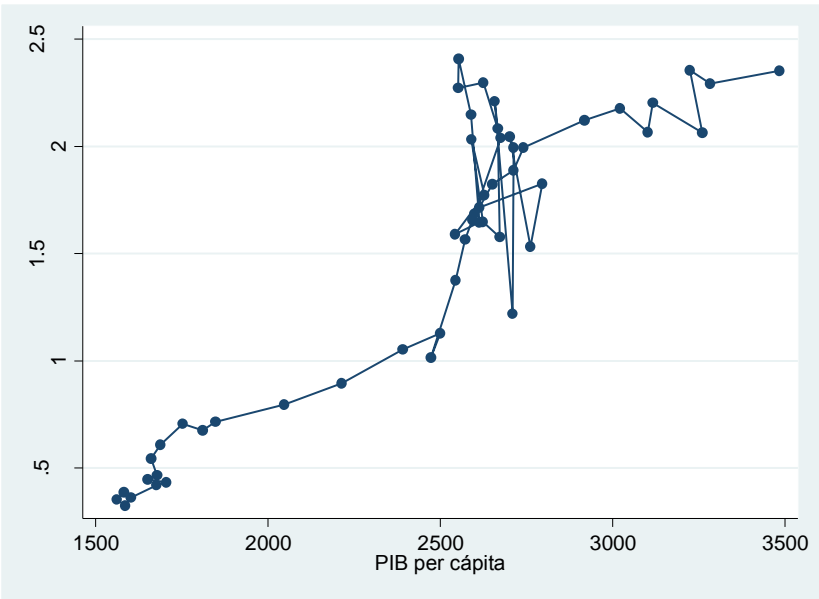
Desarrollo:

Se ejecuta la regresión lineal y se grafica la relación entre  $CO_2$  y  $PIBPC$  y comprobar si su ajusta a la curva de Kuznets Ambiental.

Source	SS	df	MS	Number of obs = 52		
Model	20.0985015	1	20.0985015	F( 1, 50) = 244.44		
Residual	4.11104914	50	.082220983	Prob > F = 0.0000		
				R-squared = 0.8302		
				Adj R-squared = 0.8268		
Total	24.2095506	51	.474697071	Root MSE = .28674		

co2	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
pibp	.0012156	.0000778	15.63	0.000	.0010595	.0013718
_cons	-1.500522	.1948153	-7.70	0.000	-1.89182	-1.109224

Los coeficientes son significativos y los signos de la relación con coherentes, además el valor  $R^2$  es superior al 80% y la probabilidad de F es cero. Procediendo con el gráfico el resultado es el siguiente:



Interpretación:

Pese a que los coeficientes eran robustos, la representación gráfica de los mismos no muestra una tendencia similar a la propuesta por Kuznets, los datos se muestran dispersos y con ciertas secciones crecientes.

## Ejercicio 2

Aplicando el modelo polinomial al conjunto de datos mencionado anteriormente se ejecuta un modelo de regresión no lineal

Comandos a utilizar:

```
Rutina Utilizada
tsset year, yearly
gen PIBPC2=PIBPC*PIBPC
reg CO2 PIBPC PIBPC2
predict YF
scatter YF PIBPC, connect(1)
```

Desarrollo:

Se genera la variable al cuadrado del Pib per cápita y se la incliye dentro del modelo de regresión, seguido de esto se predice el valor de la variable dependiente y se genera un gráfico de este último en función del Pib per cápita.

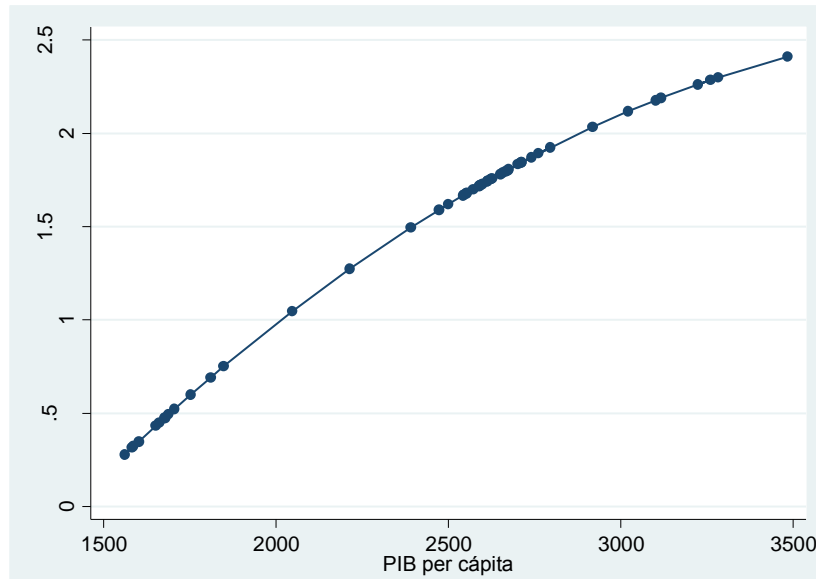
La regresión obtenida es la siguiente:

Source	SS	df	MS	Number of obs = 52		
Model	20.521655	2	10.2608275	F( 2, 49) = 136.33		
Residual	3.68789562	49	.075263176	Prob > F = 0.0000		
Total	24.2095506	51	.474697071	R-squared = 0.8477		
				Adj R-squared = 0.8415		
				Root MSE = .27434		

co2	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
pibp	.0027534	.0006528	4.22	0.000	.0014416	.0040652
pi2	-3.26e-07	1.38e-07	-2.37	0.022	-6.03e-07	-4.97e-08
_cons	-3.22489	.7507367	-4.30	0.000	-4.733551	-1.716228

Los coeficientes obtenidos son significativos al 5%, sus errores estándar son bajos y el valor de  $R^2$  es aceptable al 84%. La representación obtenida de esta relación es:



Interpretación:

La gráfica concuerda con la etapa pre industrial de la curva de Kuznet, siendo que Ecuador es un país en vías de desarrollo y se analizó una muestra del año 1960 al 2011 se concluye que a mayor crecimiento (*PIBPC*) la degradación ambiental será mayor y que es probable que en años futuros se llegue a un punto de inflexión en donde esta relación se invierta.

### 3.12.2 Métodos para estimar modelos de regresión no lineales.

En este apartado se analizan distintas formas funcionales para los datos de la tabla 3.12.1 referente a la evolución poblacional del Ecuador en los años de 1960 al 2014.

Las formas funcionales a analizar son:

Caso	Función de regresión poblacional.
Lineal	$PT = \beta_1 + \beta_2 year + u_t$
Semi-log	$LogPT = \beta_1 + \beta_2 year + u_t$

(3.12.3)

(3.12.4)

Dónde:

PT= población Total

year= Años

- **Regresión lineal.**

Para tener una base de referencia se estima un modelo lineal para la regresión de función poblacional



### Ejercicio 3

Comandos a utilizar:

```
Rutina Utilizada  
reg PT year  
predict YFL  
scatter YFL year, connect(1)
```

Desarrollo:

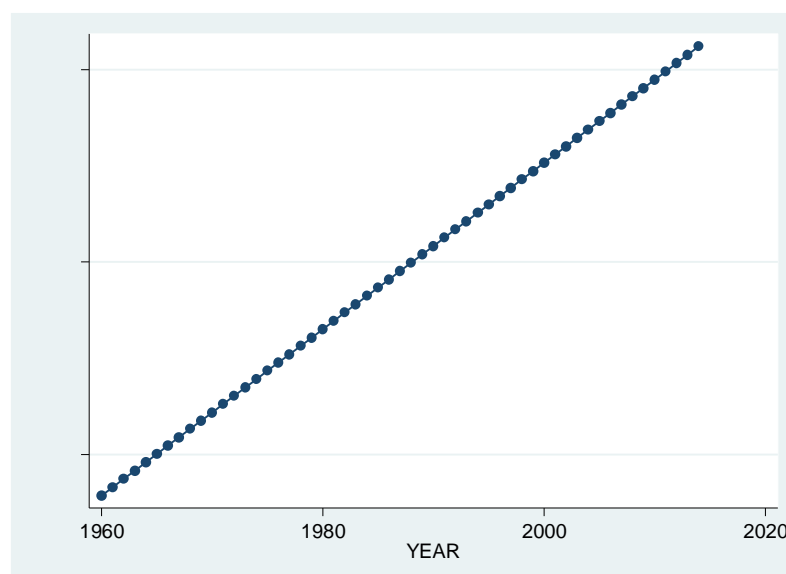
Utilizando los comandos descritos anteriormente se obtiene la siguiente regresión:

Source	SS	df	MS	Number of obs = 55		
Model	6.4721e+14	1	6.4721e+14	F( 1, 53) =12230.38		
Residual	2.8047e+12	53	5.2918e+10	Prob > F = 0.0000		
				R-squared = 0.9957		
				Adj R-squared = 0.9956		
Total	6.5002e+14	54	1.2037e+13	Root MSE = 2.3e+05		

PT	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
year	216093.4	1953.987	110.59	0.000	212174.2	220012.6
_cons	-4.20e+08	3882696	-108.07	0.000	-4.27e+08	-4.12e+08

Cada año en el Ecuador la población aumenta en 216.093 personas, esto coeficiente se acepta a una significancia del 1% y con un  $R^2$  de 99%. El gráfico representativo de esta función es el siguiente:



Interpretación.

Los datos población se ajustan a una tendencia lineal, siendo que su relación es positiva y creciente.

- **Regresión semilogarítmica.**

Permite analizar la variable dependiente en términos de porcentaje o elasticidades en vez de linealmente, de esta manera la pendiente obtenida de esta función multiplicada por 100 es igual a la tasa de crecimiento poblacional del Ecuador.

#### **Ejercicio 4**

Comandos a utilizar:

```
Rutina Utilizada  
gen LPT= log(PT)  
reg LPT year  
predict YFSL  
scatter YFSL year, connect(1)
```

Desarrollo:

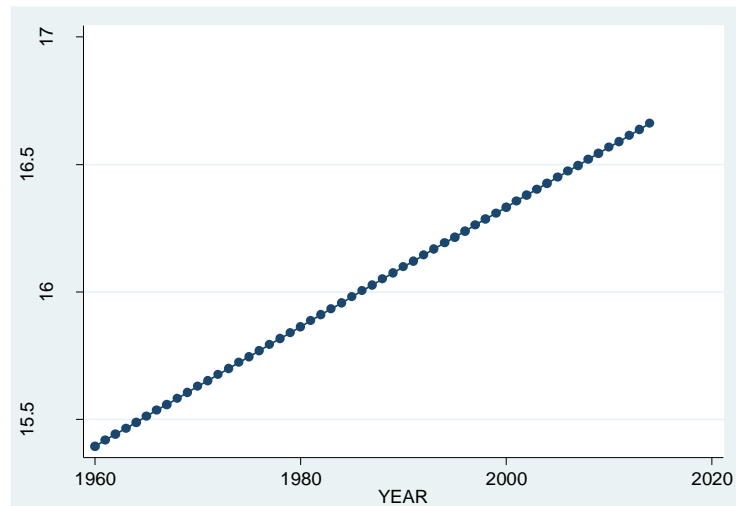
Se genera el logaritmo de la variable de población total (*PT*) y se estima en función de los años disponibles, los resultados son los siguientes:

Source	SS	df	MS	Number of obs = 55		
Model	7.62308859	1	7.62308859	F( 1, 53) = 5533.98		
Residual	.073007768	53	.001377505	Prob > F = 0.0000		
Total	7.69609636	54	.142520303	R-squared = 0.9905		
				Adj R-squared = 0.9903		
				Root MSE = .03711		

LPT	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
year	.0234522	.0003153	74.39	0.000	.0228199	.0240845
_cons	-30.57201	.626436	-48.80	0.000	-31.82848	-29.31553

La representación gráfica de los valores ajustados de Y en función de X es la siguiente:



Interpretación:

Se muestra que para los años de 1960 al 2014, la población de Ecuador creció a una tasa del 2.3% anual. Con los coeficientes altamente estadísticos y el  $R^2$  de 99% se tiene un modelo robusto con estimadores eficientes.

### 3.13 Modelos de regresión de respuesta cualitativa

(Fernández & Rodríguez, 2008) definen a los modelos de respuesta cualitativa o modelos de elección discreta, como aquellos modelos que predicen o explican la decisión de respuesta de un individuo frente a un rango de alternativas presentadas. Esta decisión de respuesta es cuantificada en términos porcentuales para medir de manera probabilística la decisión del individuo dentro de un rango finito de posibilidades. Se utilizaron cuatro tablas de datos para ejemplificar los contenidos del presente apartado, abordando temas de crecimiento económico, seguridad alimentaria, equidad de género y rendimiento académico.

#### 3.13.1 Modelo lineal de probabilidad MLP.

Cuando la variable dependiente es binaria o dicótima, es decir de respuesta cualitativa con valores de 1 y 0, se presenta un modelo lineal de probabilidades (MLP) cuya interpretación es denominada “probabilidad condicional”. La interpretación de este modelo es: A medida que se incrementa una unidad en los valores de X, la probabilidad de que el suceso asignado con valor de 1 en Y suceda es el obtenido en  $\beta_2$ .

#### Ejercicio 1

Utilizando los datos de la tabla 3.13.1 se analiza la relación entre crecimiento económico cantonal y escolaridad, como se muestra en la siguiente ecuación:

$$DE = \beta_1 + \beta_2 VAB + u_i \quad (3.13.1)$$

Donde

DE= Dicótoma de escolaridad; donde 1= personas con más de 7 años de escolaridad y 0= personas con menos de 7 años de escolaridad.

VAB= Valor Agregado Bruto Cantonal.

Comandos a utilizar:

Rutina Utilizada  
**reg DE VAB**

Desarrollo:

Realizando la regresión descrita anteriormente se obtiene:

Source	SS	df	MS	Number of obs = 221		
Model	.177719984	1	.177719984	F( 1, 219) = 0.88		
Residual	44.4286149	219	.202870387	Prob > F = 0.3503		
				R-squared = 0.0040		
				Adj R-squared = -0.0006		
Total	44.6063348	220	.202756067	Root MSE = .45041		

DE	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
VAB	.0204586	.0218583	0.94	0.350	-.022621	.0635382
_cons	.713301	.0310036	23.01	0.000	.6521973	.7744047

Interpretación:

El intercepto 0.71 representa la probabilidad de que el suceso asignado con valor de 1 suceda si X=0, es decir que existe un 71% de probabilidad de que la población del cantón tenga más de 7 años de escolaridad frente a un valor agregado bruto de cero. El intercepto de 0.20 representa que para un cambio unitario en VAB (\$1.000.000), la probabilidad de que la población tenga más de 7 años de escolaridad aumenta en un 20%.

De esta manera, utilizando la ecuación obtenida se puede calcular las probabilidades de escolaridad en base a supuestos de la variable X. Si VAB=1 (es decir \$1.000.000) entonces la probabilidad estimada de tener más de 7 años de escolaridad es:

$$(\hat{Y}_i | X = 1) = 0.71 + 0.20 (1) \quad (3.13.2)$$

$$=0.91$$

### 3.13.2 Estimación del modelo logit.

Lema (2009) menciona que estos modelos parten explícitamente en términos de probabilidades (  $P_t = \beta_1 + \beta_2 X$  ), donde luego se obtiene el logaritmo natural de esta razón de probabilidades para obtener la siguiente ecuación:

$$L_i = \ln\left(\frac{P_i}{1-P_i}\right) = Z_i = \alpha + \beta X_i \quad (3.13.3)$$

El logaritmo obtenido no es sólo lineal en  $X$ , también es lineal en los parámetros. Por tanto  $L_i$  se denomina **logit**, al igual que todos los modelos que lo incluyen.

#### Ejercicio 2

Continuando con los datos de la tabla 3.13.1 se transforma el ejercicio 1 de este capítulo en logit, como se muestra en la siguiente ecuación:

$$LDE = \beta_1 + \beta_2 VAB + u_i \quad (3.13.4)$$

En esta ocasión se recurre al método de máxima verosimilitud (MV) para la estimación de los parámetros, dado que la variable  $Y$  es de respuesta binaria.<sup>17</sup>

Comandos a utilizar:

**logit:** Estima un modelo lógico de respuesta binaria por máxima verosimilitud.

Su estructura es:

**logit** Y X1...Xn

Rutina Utilizada

**logit** DE VAB

Desarrollo:

Empleando el comando **logit** a las variables de escolaridad y valor agregado bruto se obtiene el siguiente resultado:

Logistic regression	Number of obs	=	221
	LR chi2(1)	=	1.31
	Prob > chi2	=	0.2527
Log likelihood = -130.50166	Pseudo R2	=	0.0050

DE	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
VAB	.2199127	.2880844	0.76	0.445	-.3447223	.7845477
_cons	.8933263	.1578307	5.66	0.000	.5839837	1.202669

<sup>17</sup> Los comandos **logit** y **logistic** son ambos estimadores de máxima verosimilitud, la diferencia reside en que este último genera radios de probabilidad y queda a criterio del investigador elegir que comando usar.

Interpretación:

La relación entre escolaridad y crecimiento económico se mantiene positiva, la probabilidad del intercepto es bastante alta en gran parte porque el VAB no es la única variable que explica la escolaridad cantonal o no es la que genera más impacto en Y.

Ante el incremento de un millón de dólares en el VAB, la probabilidad de que la escolaridad cantonal sea mayor a 7 años incrementa en 22% aproximadamente.

### 3.13.3 Logit para datos agrupados.

Medina (2003) argumenta que la estimación con datos agrupados se puede efectuar mediante el método de MCO utilizado para regresiones lineales, dado que ya no se encuentran variables dicótomas dentro del modelo sino por rangos. Para esto se debe linealizar el modelo tal como se mencionó en el apartado 3.13.2.

Con la  $\hat{P}_t$  estimada igual a

$$\hat{P}_t = \frac{n_t}{N_t} \quad (3.13.5)$$

El logit estimado es:

$$\hat{L}_t = \ln \left( \frac{\hat{P}_t}{1-\hat{P}_t} \right) = \hat{\beta}_1 + \hat{\beta}_2 X_t \quad (3.13.6)$$

Con la cual se podrá producir estimaciones ponderadas de mínimos cuadrados MCO.

### Ejercicio 3

Para este ejercicio usaremos los datos de la tabla 3.13.2, los cuales muestran un conjunto de observaciones correspondientes a una encuesta de seguridad alimentaria realizada en el cantón Pangui de la provincia de Zamora Chinchipe. La información está agrupada por grupos de edad y se analiza cuantas de estas han sufrido de alguna enfermedad en el último año.

**Tabla 3.13.2.** Personas que aquejan de alguna enfermedad por grupos de edad.

EDAD	Personas incluidas dentro de EDAD	Personas enfermas el último año
0.5	16	3
2.2	16	9
5.3	28	18
9	32	21
13.5	32	15
16.7	28	12
20	16	9
23	20	12
28	20	12
58.5	12	6

**Fuente y Elaboración:** Propia a partir de encuestas realizadas.

La información está agrupada por grupos de edad y se analiza cuantas de estas han sufrido de alguna enfermedad en el último año con la siguiente ecuación:

$$\widehat{LEDAD}_t = \widehat{\beta}_1 + \widehat{\beta}_2 ENF^*_t \quad (3.13.7)$$

Comandos a utilizar:

**glogit:** Genera estimaciones ponderadas de mínimos cuadrados.

Su estructura es:

**glogit** Y XN Xn: Donde (XN) es igual al conteo de observaciones dentro del rango Y y (Xn) es igual a la variable analizada que se desglosa de (XN)

Rutina Utilizada

**glogit** EDAD N ENF

Desarrollo:

Aplicando el comando **glogit** se obtiene la siguiente estimación:

Weighted LS logistic regression for grouped data

Source	SS	df	MS	Number of obs = 6		
Model	.081622368	1	.081622368	F( 1, 4) =	0.08	
Residual	4.30887452	4	1.07721863	Prob > F =	0.7967	
				R-squared =	0.0186	
				Adj R-squared =	-0.2268	
Total	4.39049689	5	.878099378	Root MSE =	1.0379	

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ENF	.0254631	.0925036	0.28	0.797	-.231368	.2822942
_cons	-1.27457	1.536503	-0.83	0.453	-5.540587	2.991447

Para facilitar la interpretación de los resultados obtenidos, se procede a calcular el antilogaritmo de los coeficientes de la regresión para luego restar la unidad de esos valores obteniendo:

$$\widehat{LEDAD}_t = -0.7205 + 0.0258 ENF^*_t \quad (3.13.8)$$

Interpretación:

Se presenta una relación positiva entre las variables analizadas, siendo que ante un incremento ponderado de un año de edad de la población, las posibilidades ponderadas de sufrir una enfermedad incrementan en 2,58%.

### 3.13.4 Modelo probit.

Este modelo es presentado por Gujarati & Porter (2010) en base a la teoría de la utilidad, o de la perspectiva de selección racional con base en el comportamiento, según el modelo desarrollado en 1974 por McFadden donde se estudia el " Análisis Logit Condicional de elección cualitativa de comportamiento". Basándose en la distribución normal acumulada, la decisión de cumplir el atributo señalado (ya sea tener más de siete años de escolaridad, o cualquier otro atributo que indique la variable dicótoma) es dependiente de un índice de conveniencia determinado por diversas variables explicativas.

#### Ejercicio 4

El presente ejercicio es tomado de la tabla 3.13.3, en la cual se relaciona el género de los empleados en función al ingreso mensual que estos perciben

$$GEN(Hombre = 1) = I(\beta_1 + \beta_2 SALARIO + u_i) \quad (3.13.9)$$

Donde,

GEN= Género de las personas, siendo que los hombres tienen la asignación de 1 y las mujeres de 0.

SALARIO= Es la remuneración mensual de los trabajadores expresada en cientos de dólares.

I= Índice de conveniencia no observable.

Estos datos son obtenidos de la Encuesta Nacional de Desempleo y Subempleo del 2014, de la cual se tomaron únicamente los datos de la parroquia Cumbe de la ciudad de Cuenca con salarios menores a \$600.00 dólares, el proceso de refinación de datos y la estimación del modelo probit se muestran a continuación.

Comandos a utilizar:

**probit:** Genera estimaciones por máxima verosimilitud.

Su estructura es:

**probit** Y X... Xn

Rutina Utilizada

**gen** GEN=p02

**replace** GEN =0 if GEN==2

**gen** SALARIO=p66/100

**drop if** SALARIO ==.

**drop if** SALARIO ==0

**drop if** SALARIO >=6

**keep if** ciudad==10152

**probit** GEN SALARIO



Desarrollo:

Luego de haber realizado las filtraciones correspondientes, se estima el modelo probit para obtener los siguientes resultados:

Probit regression	Number of obs	=	44
	LR chi2(1)	=	14.18
	Prob > chi2	=	0.0002
Log likelihood = -22.259658	Pseudo R2	=	0.2416

GEN	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
SALARIO	.6761827	.2080685	3.25	0.001	.2683759	1.083989
_cons	-1.635501	.6284497	-2.60	0.009	-2.86724	-.4037622

Interpretación:

La regresión aplicada genera la siguiente ecuación:

$$GEN(Hombre = 1) = I(-1.6355 + 0.676SALARIO + u_i) \quad (3.13.10)$$

La cual no puede interpretarse de manera directa, por lo que se procede con a obtener las derivadas de las medias respectivas de cada variable.

### Ejercicio 5

Con la aplicación del comando **mfx** se calculan los efectos marginales sobre el modelo probit.

Comandos a utilizar:

**mfx**: Calcula efectos marginales o elasticidades después de una estimación.

Su estructura es:

**mfx**

Rutina Utilizada

**probit** GEN SALARIO

**mfx**

Desarrollo:

Una vez realizada la estimación del modelo probit, se estiman los efectos marginales y se obtienen los siguientes resultados:

Marginal effects after probit  
 $y = \text{Pr}(\text{GEN}) \text{ (predict)}$   
 $= .63701012$

variable	dy/dx	Std. Err.	z	P> z	[ 95% C.I. ]	X
SALARIO	.2536885	.07858	3.23	0.001	.099674 .407703	2.93705

Interpretación:

El primer valor de Y es igual a 63.7% y determina la probabilidad de que los hombres estén en el punto medio de la muestra. Por su parte, ante el incremento de una unidad en el salario (\$100) la probabilidad de que este pertenezca a un hombre aumenta en un 25%.

### 3.13.5 Pruebas de ajuste para modelos de respuesta cualitativa.<sup>18</sup>

Las pruebas que se muestran a continuación pueden ser empleadas tanto para los modelos logit como los probit, con el fin de medir su bondad de ajuste y su correcta clasificación.

En los dos ejercicios subsecuentes se utilizarán los datos de la tabla 3.13.3 bajo la estructura de un modelo probit.

#### Ejercicio 6

- **Test de bondad de ajuste de Pearson.**

También conocido como test de Chi cuadrado de Pearson, es un test de bondad de ajuste que presenta estadísticas de los modelos logísticos, logit y probit.

Comandos a utilizar:

**estat gof:** Calcula efectos marginales o elasticidades después de una estimación.

Su estructura es:

**estat gof**

Rutina Utilizada

**probit GEN SALARIO**

**estat gof**

Desarrollo:

Empleando los comandos mencionados se obtiene:

<sup>18</sup> Para mayor información sobre pruebas a modelos de respuesta cualitativa visite el siguiente enlace [https://www.youtube.com/watch?v=2i0geT\\_xliw](https://www.youtube.com/watch?v=2i0geT_xliw)

Probit model for GEN, goodness-of-fit test

```
number of observations =      44
number of covariate patterns =    22
Pearson chi2(20) =      16.79
Prob > chi2 =      0.6667
```

Interpretación:

Los resultados muestran que el modelo se ajusta bastante bien, los patrones de covarianza son menores al número de observaciones y la probabilidad de chi2 es mayor al 5%, por tanto se rechaza la hipótesis nla de que le modelo no se encuentra bien especificado.

**Ejercicio 7**

- **Tabla de estadísticas y clasificación.**

Muestra un grupo de estadísticas para los datos del modelo luego de haberse realizado una estimación logística, logit, probit o ivprobit.

Comandos a utilizar:

**estat class:** Genera una tabla de clasificación de estadísticas de un modelo.

Su estructura es:

**estat class**

Rutina Utilizada

**probit GEN SALARIO**

**estat class**

Desarrollo:

Empleando la rutina anteriormente mencionada se obtiene:

Classified	True		Total
	D	~D	
+	24	9	33
-	3	8	11
Total	27	17	44

Classified + if predicted Pr(D) >= .5  
True D defined as GEN != 0

Sensitivity	Pr( +  D)	88.89%
Specificity	Pr( -  ~D)	47.06%
Positive predictive value	Pr( D  +)	72.73%
Negative predictive value	Pr( ~D  -)	72.73%

False + rate for true ~D	Pr( +  ~D)	52.94%
False - rate for true D	Pr( -  D)	11.11%
False + rate for classified +	Pr( ~D  +)	27.27%
False - rate for classified -	Pr( D  -)	27.27%

Correctly classified	72.73%
----------------------	--------

Interpretación:

La clasificación se muestra por grupos positivos y negativos de distribución con cada una de los índices probabilísticos respectivos, aunque el estadístico más relevante es el que se muestra al final de la tabla, cuyo valor es de 72.73% lo cual es bastante bueno y se determina que el presente modelo se encuentra correctamente clasificado.

### 3.13.6 Modelo tobit.

Una extensión del modelo probit es el modelo tobit, desarrollado por James Tobin, economista laureado con el Nobel en 1981. El modelo tobit también se conoce como modelo de regresión censurada. Algunos autores los llaman modelos de regresión con variable dependiente limitada debido a la restricción impuesta sobre los valores tomados por la variable regresada.

### Ejercicio 8

Utilizando los datos de la tala tabla 3.9.1 sobre la relación entre promedio obtenido, número de componentes matriculado y la nota de examen de admisión se presenta la siguiente ecuación:

$$PROM = \beta_1 + \beta_2 NCA + \beta_3 NEA + u_i \quad (3.13.11)$$

Con la condición de:

Límite inferior igual a 0.28, es decir se excluye de la muestra a todos los estudiantes que reprobaron durante el periodo de estudio.

Comandos a utilizar:

**tobit:** Ajusta un modelo de variables dependientes en variables independientes donde se fijan los valores de censura.

Su estructura es:

**tobit** Y X... Xn, ll(n) ul(n): Donde ll es igual límite inferior y ul es el límite superior de la muestra.

Rutina Utilizada

**tobit** PROM NCM NEA, ll(28)

Desarrollo:

La regresión obtenida al ejecutar el comando **tobit** es:

```

Tobit regression
Log likelihood = -493.13155
Number of obs   =      212
LR chi2(2)      =       3.25
Prob > chi2     =     0.1969
Pseudo R2      =     0.0033

```

PROM	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
NCM	.0619178	.1343782	0.46	0.645	-.2029852	.3268208
NEA	.0247444	.0141995	1.74	0.083	-.0032476	.0527363
_cons	29.6796	1.211067	24.51	0.000	27.29219	32.067
/sigma	3.466832	.2017068			3.069203	3.864462

```

Obs. summary:      45 left-censored observations at PROM<=28
                   167 uncensored observations
                   0 right-censored observations

```

#### Interpretación:

Todos los coeficientes se muestran positivos, siendo que ante el aumento de un componente más en el periodo de matriculación, el promedio aumentaría en 0.06 y en 0.02 si se incrementa un punto dentro la nota de examen de admisión, recalcando que esto aplica solo a los estudiantes cuyo promedio es superior a 28 puntos.

### 3.14 Modelos regresión con datos de panel

Como se mencionó en apartados iniciales de este tercer capítulo, los datos de panel son aquellos datos de corte transversal que se estudian a lo largo del tiempo, por tanto, estos datos se encuentran distribuidos en el espacio y en el tiempo (Gujarati & Porter, 2010).

Se utilizaron 4 tablas de datos referentes a la producción de banano en el Ecuador dentro del periodo 2000-2012, estas se dividen en una tabla general que agrupa todas las provincias a nivel nacional y las tres restantes son las provincias segregadas por las regiones costa, sierra y Amazonía.

#### 3.14.1 Modelo de regresión con MCO.

La ejecución de un modelo panel por Mínimos Cuadrados Ordinarios es bastante simple, ya que se realiza el mismo procedimiento de cualquier otra basa de datos con la única excepción de que se debe declarar la naturaleza de la base de datos como panel antes de efectuar cualquier regresión o procedimiento.

## Ejercicio 1

En base a la tabla 3.14.1 se efectúa una regresión sobre la producción de banano en las provincias de la región costa para los años 2000 a 2012. La ecuación que expresa esta relación es la siguiente:

$$P_{\text{banano}} = \beta_1 + \beta_2 SS + \beta_3 R + u_{it} \quad (3.14.1)$$

La producción de banano ( $P_{\text{banano}}$ ) está en función de la superficie sembrada ( $SS$ ) y del rendimiento ( $R$ ).

Comandos a utilizar:

Rutina Utilizada

**tsset** id Year, yearly  
**reg** Pbanano SS R

Desarrollo:

Primeramente se declara la naturaleza de los datos con el comando **tsset**, seguido de la variable de identificación (id), la variable de tiempo (Year) y la periodicidad de los datos, en esta ocasión estos son anuales. Los resultados de la regresión son los siguientes:

Source	SS	df	MS	Number of obs = 78		
Model	5.6913e+13	2	2.8456e+13	F( 2, 75) = 1600.30		
Residual	1.3336e+12	75	1.7782e+10	Prob > F = 0.0000		
Total	5.8246e+13	77	7.5645e+11	R-squared = 0.9771		
				Adj R-squared = 0.9765		
				Root MSE = 1.3e+05		

Pbanano	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
SS	22.86928	1.164336	19.64	0.000	20.5498	25.18876
R	27397.1	1819.04	15.06	0.000	23773.38	31020.81
_cons	-412863.6	27823.22	-14.84	0.000	-468290.3	-357436.9

Interpretación:

La muestra de 78 datos para las 6 provincias de la región costa resulta en una regresión bastante robusta, con valores de  $R$  mayores al 90% y coeficientes altamente significativos. Por tanto se puede asegurar que los efectos que tiene el rendimiento y la superficie sembrada sobre la producción de banano en estas 6 localidades son positivas y de gran impacto, siendo que ante el incremento de una hectárea de superficie sembrada de banano, su producción aumenta en 22.8 toneladas métricas.

### 3.14.2 Modelo de mínimos cuadrados con variable dicótoma (MCVD) de efectos fijos.

Este modelo permite que cada variable explicativa tenga su propio valor de intercepto, generando n-1 categorías para explicar cada una de las posibilidades presentes dentro de dicha variable, disminuyendo así la heterogeneidad dentro del modelo. Además, los efectos fijos permiten que el intercepto no varíe en el tiempo, es decir manteniéndolo fijo (Gujarati & Porter, 2010).

#### **Ejercicio 2**

Para demostrar el modelo MCVD utilizaremos la tabla 3.14.2 sobre producción del banano aplicada a la región amazónica del Ecuador. La ecuación a utilizar es la siguiente:

$$P_{\text{Abanano}} = \alpha_1 + \alpha_2 D1 + \alpha_3 D2 + \alpha_4 D3 + \alpha_5 D4 + \beta_2 SS + \beta_3 R + u_{it} \quad (3.14.2)$$

Dónde:

D1= 1 si la observación es igual a la provincia de Napo y cero si es otra.

D2= 1 si la observación es igual a la provincia de Orellana y cero si es otra.

D3= 1 si la observación es igual a la provincia de Sucumbíos y cero si es otra.

D4= 1 si la observación es igual a la provincia de Zamora y cero si es otra.

P<sub>Abanano</sub>= Producción de banano en la región amazónica.

SS= Superficie sembrada

R= Rendimiento

Comandos a utilizar:

Rutina Utilizada

**tsset** id Year, yearly

**reg** P<sub>Abanano</sub> D1 D2 D3 D4 SS R

Desarrollo:

De igual forma que el ejercicio anterior, se declara la base como datos de panel y se efectúa la regresión de Mínimos cuadrados con Variable Dicótoma y se obtiene:

Source	SS	df	MS	Number of obs = 64		
Model	5.5579e+09	6	926323664	F( 6, 57) = 144.86		
Residual	364487192	57	6394512.14	Prob > F = 0.0000		
				R-squared = 0.9385		
				Adj R-squared = 0.9320		
Total	5.9224e+09	63	94006812.4	Root MSE = 2528.7		

PAbanano	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
D1	7040.105	1870.402	3.76	0.000	3294.689	10785.52
D2	4204.603	1839.931	2.29	0.026	520.2055	7889.001
D3	4365.041	1680.763	2.60	0.012	999.3711	7730.712
D4	1508.189	1327.612	1.14	0.261	-1150.308	4166.686
SS	6.603186	.4211328	15.68	0.000	5.759882	7.44649
R	723.86	174.9027	4.14	0.000	373.6233	1074.097
_cons	-9071.37	1875.569	-4.84	0.000	-12827.13	-5315.608

Interpretación:

Gran parte de los coeficientes se muestran altamente significativos a excepción de la dicótoma 4 de Zamora Chinchipe, de manera comparativa se aprecia un mayor impacto por parte de la provincia de Napo sobre la producción de banano, seguido por Sucumbíos, Orellana y finalmente Zamora. Los efectos de *SS* y *R* son positivos pero con un pequeño impacto sobre la producción.

### 3.14.3 Modelos de Efectos Fijos (MEF).

Baronio (2014) considera a los modelos de efectos fijos como aquellos que poseen un término constante diferente para cada individuo ( $\alpha_{it}$ ) y se asume que los efectos individuales son independientes entre sí.

### Ejercicio 3

A continuación se utilizará la tabla 3.14.3 la cual contiene 20 provincias productoras de banano de la región Sierra, Costa y Amazonía. Partiendo de la ecuación 3.14.1 se presenta la siguiente ecuación:

$$PT_{\text{banano}} = \alpha_{it} + \beta_2 SS + \beta_3 R + u_{it} \quad (3.14.3)$$



Comandos a utilizar:

**xtreg:** Ajusta modelos de regresiones a datos de panel.  
**fe:** Es una opción del comando **xtreg** para efectos fijos dentro del modelo.

Su estructura es:  
**xtreg** Y X... Xn, **fe**

Rutina Utilizada  
**tsset** id Year, yearly  
**xtreg** PTbanano SS R, **fe**

Desarrollo:

Aplicando la rutina de comando anteriormente descrita se tiene:

```
Fixed-effects (within) regression      Number of obs      =      255
Group variable: id                   Number of groups   =      20

R-sq:  within  = 0.4822                Obs per group: min =      12
      between = 0.9415                      avg  =     12.8
      overall  = 0.9213                      max  =     13

                                         F(2,233)           =    108.47
corr(u_i, Xb)  = 0.0485                 Prob > F            =    0.0000
```

PTbanano	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
SS	27.00655	2.636401	10.24	0.000	21.81232	32.20078
R	17426.52	1779.126	9.79	0.000	13921.29	20931.75
_cons	-207518.3	37644.14	-5.51	0.000	-281684.7	-133351.9
sigma_u	176703.13					
sigma_e	117493.91					
rho	.69342255	(fraction of variance due to u_i)				

```
F test that all u_i=0:      F(19, 233) =    28.67          Prob > F = 0.0000
```

Interpretación:

En este tipo de modelos existen tres tipos de R cuadrado, como se evidencia en la parte superior izquierda los coeficientes varían desde 0.48 a 0.94, lo cual es bastante aceptable. El número de grupos es igual al número de provincias incluidas dentro del modelo, además de que los coeficientes son altamente significativos y explican con un considerable efecto la producción de banano en el Ecuador.

### 3.14.4 Modelos de Efectos aleatorios (MEFA).

A diferencia de los modelos de efectos fijos, la variable ( $\alpha_{it}$ ) es considerada aleatoria, además se suele prescindir de los efectos temporales de dicha variable para centrarse en sus efectos individuales.

## Ejercicio 4

Continuando con los datos de la tabla 3.14.3, realizamos el modelo de efectos aleatorios para la muestra nacional de producción de banano, con la siguiente ecuación:

$$PTbanano = \alpha + \beta_2 SS + \beta_3 R + u_{it} \quad (3.14.4)$$

Comandos a utilizar:

**xtreg**: Ajusta modelos de regresiones a datos de panel.  
**re**: Es una opción del comando **xtreg** para efectos aleatorios dentro del modelo.

Su estructura es:  
**xtreg** Y X... Xn, **re**

Rutina Utilizada  
**tsset** id Year, yearly  
**xtreg** PTbanano SS R, **re**

Desarrollo:

Aplicando el comando **xtreg** para efectos aleatorios se obtiene la siguiente regresión:

```

Random-effects GLS regression              Number of obs   =       255
Group variable: id                        Number of groups  =        20

R-sq:  within = 0.4821                    Obs per group: min =        12
       between = 0.9417                      avg =       12.8
       overall = 0.9215                      max =        13

                                           Wald chi2(2)      =       498.46
corr(u_i, X)  = 0 (assumed)                Prob > chi2       =       0.0000

```

PTbanano	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
SS	27.43619	1.849839	14.83	0.000	23.81057	31.0618
R	17263.8	1634.569	10.56	0.000	14060.11	20467.5
_cons	-210297.2	47739.65	-4.41	0.000	-303865.2	-116729.2
sigma_u	182271.14					
sigma_e	117493.91					
rho	.70645275	(fraction of variance due to u_i)				

Interpretación:

Los resultados son muy similares al modelo de efectos fijos, tanto en nivel de robustez, pruebas de hipótesis y coeficientes. La relación entre superficie sembrada, rendimiento y producción permanece positiva y altamente significativa.

### 3.14.5 Test de Hausman.

Como se aprecia en el ejercicio 3 y 4 de este apartado, los resultados de las estimaciones son apropiados para el conjunto de datos, basándose en el sentido económico de las variables y sus significancia, por lo que resulta pertinente explorar un test que permita asegurar cuál de estas dos metodologías es más adecuada para el estudio de la productividad de banano en Ecuador. Este test fue desarrollado por el matemático Hausman en 1978 para comparar estimadores y ver cuál de los dos es más eficiente.

### Ejercicio 5

En base a las ecuaciones 3.14.3 y 3.14.4 se procede a aplicar el test de Hausman y comprobar cuál de estas es más eficiente.

Comandos a utilizar:

**hausman**: Genera un test de eficiencia basado en hipótesis.

Su estructura es:

**hausman** estimación fija estimación aleatoria

Rutina Utilizada

**xtreg** PTbanano SS R, **fe**

**estimates store** fixed

**xtreg** PTbanano SS R, **re**

**estimates store** random

**hausman** fixed random

Desarrollo:

A medida de que se estiman las ecuaciones, estas deben ser guardadas con el comando **estimates store** bajo cualquier nombre deseado, en este caso se denominó a cada modelo como *fixed* para efectos fijos y *random* para efectos aleatorios. Una vez realizado esto, se aplica el test de Hausman y se obtiene el siguiente resultado:

—— Coefficients ——				
	(b) fixed	(B) random	(b-B) Difference	$\sqrt{\text{diag}(V_b - V_B)}$ S.E.
SS	27.00655	27.43619	-.4296333	1.878484
R	17426.52	17263.8	162.7196	702.4764

b = consistent under  $H_0$  and  $H_a$ ; obtained from xtreg  
 B = inconsistent under  $H_a$ , efficient under  $H_0$ ; obtained from xtreg

Test:  $H_0$ : difference in coefficients not systematic

$\chi^2(2) = (b-B)' [(V_b - V_B)^{-1}] (b-B)$   
 = 0.22  
 Prob> $\chi^2$  = 0.8944

Interpretación:

Como se mencionó anteriormente, esta prueba se rige bajo dos hipótesis:

$H_0$ : El modelo de efectos fijos es el más eficiente

$H_1$ : El modelo de efectos fijos no es el más eficiente.

Por tanto, dado que Prob> $\chi^2$  es mayor al 5% de aceptación, se rechaza la hipótesis alterna y se acepta la hipótesis nula de que el mejor modelo es el de efecto fijos.

## CONCLUSIONES

El presente trabajo de titulación permitió concluir con lo siguiente:

- Mediante el estudio exploratorio realizado a estudiantes de econometría se evidenció que existen dificultades tanto en el uso del software estadístico Stata como en la creación de modelos econométricos con datos del Ecuador. Aproximadamente el 87% de los estudiantes presenta una dificultad intermedia difícil y muy difícil del manejo del programa Stata, principalmente por la carencia de una guía instructiva que seguir y por la dificultad de entendimiento de los manuales de Stata en línea. El 47.27% de los estudiantes presenta dificultad en la creación de modelos econométrico con datos del Ecuador debido a la falta de aplicabilidad de la teoría a la realidad ecuatoriana y por la falta de disponibilidad de datos en las fuentes oficiales de información del país. Además el 98.18% de los estudiantes tienen la necesidad de una guía pedagógica de econometría con ejercicios resueltos mediante bases de datos ecuatorianas y utilizando el software estadístico Stata.
- Se aplicó de manera exitosa los contenidos de “Econometría” de Gujarati & Porter (2010) a 108 ejercicios, de los cuales 104 fueron resueltos con datos económicos, sociodemográficos y ambientales del Ecuador, tan solo 4 ejercicios fueron resueltos con datos hipotéticos con el único fin de demostrar aspectos teóricos de la econometría.
- Las bases de datos del Ecuador presentan limitantes en el aspecto temporal de ciertas variables, debido a que existe información con diferente año base en distintas fuentes, ciertos datos históricos no se encuentran actualizados a dólares americanos por lo tanto siguen constando en sucres, existe escasez de datos ya que algunas variables solamente cuentan con información de pocos años o de los años más actuales, y además la información existente es poco segregada a nivel cantonal y zonal.
- La realidad ecuatoriana no siempre va acorde a la teoría económica utilizada en el libro “Econometría” de Gujarati & Porter (2010), ya que la realidad de una economía desarrollada como Estados Unidos no siempre es concordante con la realidad de un país en vías de desarrollo como es el caso de Ecuador. La teoría económica no siempre se ajusta a la realidad de todos los países, pues cada sociedad presenta comportamientos económicos, sociales, culturales, demográficos y ambientales distintos.

- El desarrollo del presente trabajo de titulación permitió completar una de las primeras guías econométricas con ejercicios aplicados a la realidad del Ecuador, utilizando el software estadístico Stata 13 mediante procesos detallados y de fácil comprensión para el estudiante.

## RECOMENDACIONES

- Es importante validar el presente trabajo de titulación con estudiantes de econometría para conocer su grado de impacto y tener una retroalimentación por parte de los mismos. Con ello se podría determinar si la guía es de gran utilidad para los estudiantes tanto en el aprendizaje y aplicabilidad de la econometría con datos del Ecuador, como en la comprensión del manejo del software estadístico Stata. De esta manera, los resultados esperados del presente trabajo aspiran lograr una mayor comprensión en los estudiantes al aplicar la teoría econométrica, reflejando mejores resultados académicos en la materia de econometría, y demostrando habilidad y agilidad en el uso del programa Stata.
- Se debería profundizar en los métodos y técnicas econométricas que brinda el software estadístico Stata, ya que existen mejores alternativas de resolución econométrica de ejercicios a las que presenta Gujarati & Porter (2010) al utilizar el software estadístico E-Views. Estas alternativas acortan procesos densos, ejecutando resultados de manera más directa.
- El presente trabajo aplica los 16 primeros capítulos del libro “Econometría” de Gujarati & Porter (2010), por lo cual se recomienda que para futuras investigaciones se complemente este libro con ejercicios que abarquen los temas del capítulo 17 al 22, que presentan un mayor grado de complejidad pero igualmente importantes.
- Se debería continuar con estas iniciativas de trabajos prácticos de fin de titulación que puedan aportar significativamente al aprendizaje de los estudiantes como una herramienta adicional a la brindada por las universidades o distintos centros educativos.

## BIBLIOGRAFÍA

- Arce, R., Mahía, R., & Definición, I (2001). Conceptos básicos sobre la heteroscedasticidad en el modelo básico de regresión lineal tratamiento con E-Views.
- Arranz, J. M., & Zamora, M. M (2001). Análisis de Autocorrelación. 16-20.
- Banco Central del Ecuador (2016). *Cuentas Nacionales Regionales*. Recuperado de <https://www.bce.fin.ec/index.php/component/k2/item/293-cuentas-provinciales>
- Banco Central del Ecuador (2016). *Cuentas Nacionales Trimestrales*. Recuperado de <https://contenido.bce.fin.ec/home1/estadisticas/cntrimestral/CNTrimestral.jsp>
- Banco Central del Ecuador (2016). *Sistema de información macroeconómica*. Recuperado de <http://sintesis.bce.ec:8080/BOE/BI/logon/start.do?ivsLogonToken=bceqsappbo01%3A6400%40471683Jlk3lspKU3Kk0Oy40UPg89Y471681JISujlQGTEMz2jmGmPB2U0C>
- Banco Mundial (2015). *Ecuador*. Recuperado de <http://datos.bancomundial.org/pais/ecuador>
- Banco Mundial (2015). *World Development Indicators*. Recuperado de <http://data.worldbank.org/data-catalog/world-development-indicators>
- Baum, C & Schaffer, M (2013). A general approach to testing for autocorrelation. Presented at the Stata Conference, New Orleans.
- Cox, N. J (2005). A brief history of Stata on its 20th anniversary. *Stata Journal*, (1), 2–18.
- Dirección Nacional de Migración (2007). *Oficina de Estadísticas*. Recuperado de <http://www.ministeriointerior.gob.ec/migracion/>
- David Hendry, (1980). *Econometrics-Alchemy or Science?* Retrieved from <http://www.jstor.org/stable/2553385>
- Fuentes, J., & Palma, A (2003). *Manual STATA*. Chile. Retrieved from [http://www.dev-out.cl/sites/default/files/Manual Stata.pdf](http://www.dev-out.cl/sites/default/files/Manual%20Stata.pdf)
- Enguñados, A. M (1994). Los modelos de predicción del fracaso empresarial: una aplicación empírica del logit. *Revista Española de Financiación y Contabilidad*, 203-233.
- Escobar, M., Fernández, E., & Bernardi, F (2012). *Análisis de Datos con Stata (2ª ed.)*. Madrid: Cuadernos Metodológicos.
- Fuentes, J., & Palma, A (2003). *Manual STATA*. Chile. Recuperado de [http://www.dev-out.cl/sites/default/files/Manual Stata.pdf](http://www.dev-out.cl/sites/default/files/Manual%20Stata.pdf)



- Gould, W., & Cox, N (2015). *Stata | FAQ: History of Stata*. Recuperado de <http://www.stata.com/support/faqs/resources/history-of-stata/>
- Gujarati, M., & Porter, D (2010). *Econometría (5ª ed.)*. México: MC-Graw-Hill.
- Hendry, D. F (1980). Econometrics-alchemy or science?. *Economica*, 387-406.
- Instituto Nacional de Estadísticas y Censos INEC (2010). *Censo Nacional Económico*. Recuperado de <http://www.ecuadorencifras.gob.ec/banco-de-informacion/>
- Instituto Nacional de Estadísticas y Censos INEC (2010). *VII Censo de población y VI de vivienda*. Recuperado de <http://www.ecuadorencifras.gob.ec/banco-de-informacion/>
- Instituto Nacional de Estadísticas y Censos INEC (2014). *Encuesta de Condiciones de Vida*. Recuperado de <http://www.ecuadorencifras.gob.ec/banco-de-informacion/>
- Instituto Nacional de Estadísticas y Censos INEC (2014). *Encuesta de Nacional de Empleo, Desempleo y Subempleo*. Recuperado de <http://www.ecuadorencifras.gob.ec/banco-de-informacion/>
- Jiménez-Martín, S (2001). Breve curso de introducción a la programación en stata (6.0).
- Labra, R., & Torrecillas, C (2014). Guía CERO para datos de panel. Un enfoque práctico. *UAM-Accenture Working Papers*, (2014/16), 1-57.
- Legendre, P (1993). Spatial autocorrelation: trouble or new paradigm?. *Ecology*, 74(6), 1659-1673.
- Manoleter (2013). *Tablas estadísticas (normal, t student, chi-cuadrado, fisher, binomial, poisson)* [diapositivas de PowerPoint]. Recuperado de <http://es.slideshare.net/manoleter/tablas-estadsticas-normal-t-student-chicuadrado-fisher-binomial-poisson>
- Martín, S. J (2001). Breve curso de introducción a la programación en Stata (6.0). *Documentos-Instituto de Estudios Fiscales*, (7), 1-38.
- Mestre, R (1994). Stata, un paquete sencillo pero potente. *Revista de Economía Aplicada*, II(1), 163–172.
- Mincer, J (1974). Schooling, Experience, and Earnings. *Human Behavior & Social Institutions* No. 2.
- Ministerio de Agricultura, Ganadería, Acuacultura y Pesca (2012). *Costos Producción*. Recuperado de <http://sinagap.agricultura.gob.ec/productos/costos-de-produccion>

- MULTION (2015). *Stata*. Recuperado de <http://www.multon.com.mx/micrositios/stata/xqUsarStata.php>
- Muñoz, J (s.f.). *Curso de introducción a STATA*, 1–11. Recuperado de [http://gent.uab.cat/jmunoz/sites/gent.uab.cat/jmunoz/files/Curso de introducción a STATA\\_S1.pdf](http://gent.uab.cat/jmunoz/sites/gent.uab.cat/jmunoz/files/Curso%20de%20introducci3n%20a%20STATA_S1.pdf)
- Trapero, J., Dorado, J., Frutos, M., Rey, M., & Sanz, M (1999). *Cien ejercicios de Econometría*.
- Otero, J. V (1998). *Econometría y contrastación empírica: concepto e historia*. Universidad Autónoma, Instituto de Predicción Económica Laurence R. Klein.
- Pena, J. B., Estavillo, J., Galindo, M., Leceta, M., & Zamora, M (1999). *Cien ejercicios de Econometría*. Madrid: Pirámide.
- Polisofando (2013). *Tablas distribuciones* [diapositivas de PowerPoint]. Recuperado de <http://www.slideshare.net/pilosofando/tablas-distribuciones>
- Quintana, L., & Mendoza, M (2008). *Econometría Básica*. México: Plaza y Valdés.
- Rojo, J (2008). *Curso de introducción al paquete Stata*. Centro de Ciencias Humano Y Sociales. Recuperado de <http://investigadores.cide.edu/aparicio/data/IntroduccionStata9.pdf>
- Schumpeter, J (1933). The common sense of econometrics. *Econometrica: Journal of the Econometric Society*, 5-12.
- StataCorp (2007). *Getting Started with Stata for UNIX*. Stata Press.
- StataCorp (2009). *Análisis de microdatos de la ENIGH con STATA*. Recuperado de [http://www.stata.com/meeting/mexico09/mex09sug\\_jfi.pdf](http://www.stata.com/meeting/mexico09/mex09sug_jfi.pdf)
- StataCorp (2013). *Stata base reference manual*. Stata Press. Retrieved from <http://www.stata.com/manuals13/u.pdf>
- Tintner, G (1953). The definition of econometrics. *Econometrica: Journal of the Econometric Society*, 31-40.
- Williams, R (2015). *Heteroscedasticidad*. Universidad de Notre Dame.

## **ANEXOS**

## Anexo 1. Encuesta

### UNIVERSIDAD TECNICA PARTICULAR DE LOJA

La Universidad Católica de Loja

Titulación de Economía

“La presente encuesta es realizada para determinar el grado de dificultad en la aplicación de conceptos econométricos y el manejo del software STATA. Esta información será de uso exclusivo para fines académicos y su contenido será anónimo.”

#### 1. Datos informativos

##### 1.1 Género

1 ( ☐ ) F                      2 ( ☐ ) M

##### 1.2 Edad

- 1. ( ☐ ) 18-20
- 2. ( ☐ ) 21-23
- 3. ( ☐ ) 24-26

##### 1.3 Ciclo que cursa actualmente

- 1. ( ☐ ) Séptimo
- 2. ( ☐ ) Noveno

##### 1.4 ¿Cuál es el último componente de econometría que usted ha aprobado?

- 1. ( ☐ ) Econometría I
- 2. ( ☐ ) Econometría II
- 3. ( ☐ ) Econometría III

##### 1.5. ¿Ha reprobado algún componente de econometría? ¿Cuál?

*(Señale cual/es componentes ha reprobado)*

- 1. ( ☐ ) Econometría I
- 2. ( ☐ ) Econometría II
- 3. ( ☐ ) Econometría III
- 4. ( ☐ ) Ninguna

#### 2. Cuando empezó a utilizar el software estadístico STATA, ¿qué grado de dificultad le ocasionó el manejo del mismo?

- 1 ( ☐ ) Fácil
- 2 ( ☐ ) Intermedio
- 3 ( ☐ ) Difícil
- 4 ( ☐ ) Muy difícil

*Si su respuesta es fácil pase a la pregunta 4*

3. ¿Por qué le dificultó utilizar el software STATA? *(puede escoger más de una respuesta)*

1 ( ) Porque no tenía una guía instructiva qué seguir.

2 ( ) Porque STATA está en idioma inglés.

3 ( ) Porque los manuales de STATA en línea no tienen ejercicios resueltos claramente.

4 ( ) Otro

¿Cuál? \_\_\_\_\_

4. ¿Considera que aplica con facilidad los conocimientos aprendidos en econometría?

1 ( ) Si

2 ( ) No

5. La creación de modelos econométricos con bases de datos de Ecuador le resulta difícil? ¿Por qué?

1 ( ) Si

2 ( ) No

Porqué? \_\_\_\_\_  
\_\_\_\_\_

6. ¿Cree que es importante aplicar bases de datos de Ecuador en modelos econométricos?

1 ( ) Si

2 ( ) No

7. ¿Considera necesaria la creación de una guía pedagógica de econometría utilizando el software STATA que utilice datos del Ecuador?

1 ( ) Si

2 ( ) No

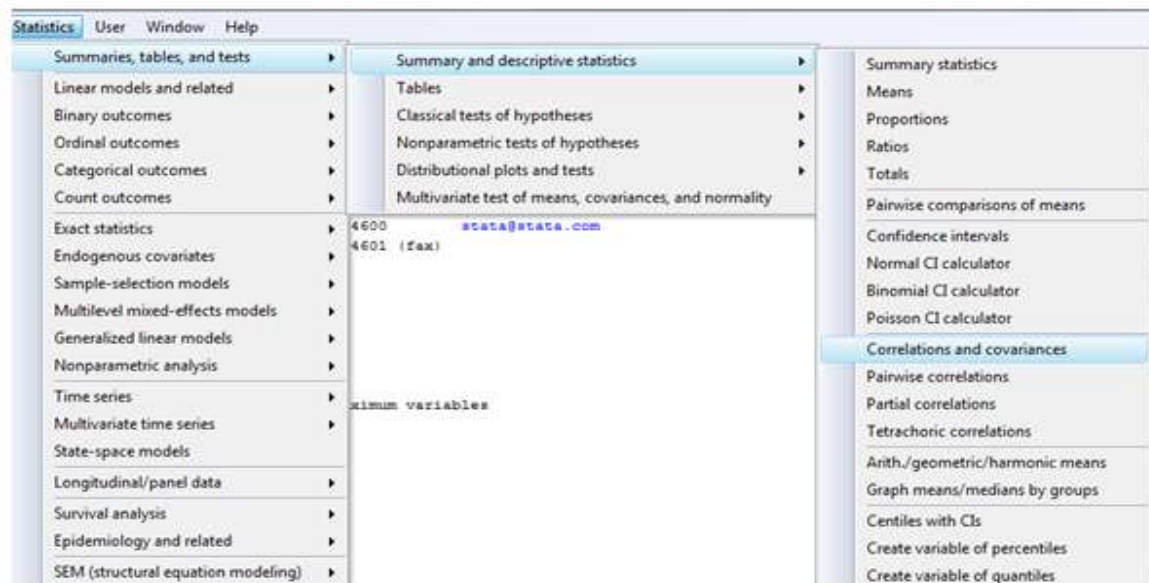
8. Seleccione la opción metodológica de resolución econométrica que considere sea la mejor en términos explicativos y de proceso.

## Detección de Multicolinealidad mediante una matriz de correlación

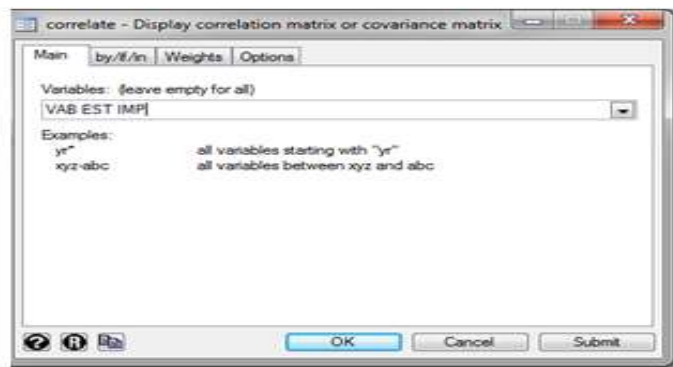
( ) OPCIÓN 1

### Correlograma

Para obtener la matriz de correlación se debe ir al menú *Statistics* y elegir la opción *Summaries, tables, and tests* y dentro de la misma se dirige a *Summary and descriptive statistics* para finalmente seleccionar la opción de *Correlations and covariances*.



Luego aparecerá una ventana en la cual se introducirán las variables a constar en la matriz de correlación.



Finalmente se mostrará en la pantalla una matriz de correlación de todas las variables seleccionadas con sus respectivos coeficientes de correlación  $r$ .

Un método más sencillo y directo para obtener esta matriz es escribir el siguiente comando en la barra inferior de la pantalla:

- correlate VAB EST IMP

Command				
correlate VAB EST IMP				
	VAB	EST	IMP	
VAB	1.0000			
EST	0.9887	1.0000		
IMP	0.9517	0.9318	1.0000	

( ) OPCIÓN 2

**Correlograma**

Para obtener la matriz de correlación se debe ir al menú *Statistics* y elegir la opción *Summaries, tables, and tests* y dentro de la misma se dirige a *Summary and descriptive statistics* para finalmente seleccionar la opción de *Correlations and covariances*.

Luego aparecerá una ventana en la cual se introducirán las variables a constar en la matriz de correlación.

Finalmente se mostrará en la pantalla una matriz de correlación de todas las variables seleccionadas con sus respectivos coeficientes de correlación  $r$ .

	VAB	EST	IMP
VAB	1.0000		
EST	0.9887	1.0000	
IMP	0.9517	0.9318	1.0000

( ) OPCIÓN 3

**Correlograma**

Es uno de los métodos más utilizados para detectar multicolinealidad entre variables explicativas.

Comandos a utilizar:

- *correlate*: Permite crear una matriz de correlación entre un conjunto de variables determinadas, con el fin de evaluar el grado de correlación que existe entre éstas.  
Su estructura es:  
*correlate variable1 variable 2 variable n*

Desarrollo:

En el archivo *do* o en la parte inferior de la ventana principal en *Command* se escribirá lo siguiente:

*correlate* VAB EST IMP

Ejecutamos este comando y se obtendrá el correlograma o matriz de correlación entre las tres variables (VAB, EST, IMP):

	VAB	EST	IMP
VAB	1.0000		
EST	0.9887	1.0000	
IMP	0.9517	0.9318	1.0000

Interpretación:

De acuerdo a los datos mostrados por el correlograma, se observa que el grado de correlación entre las variables "EST" y la variable "VAB" es de 0.98, entre la variable "IMP" y la variable "VAB" es de 0.95, y entre la variable "IMP" y la variable "EST" es de 0.93, estos valores al ser muy elevados indican que en estos tres casos hay una relación lineal casi perfecta, presenciando así multicolinealidad entre estas variables.

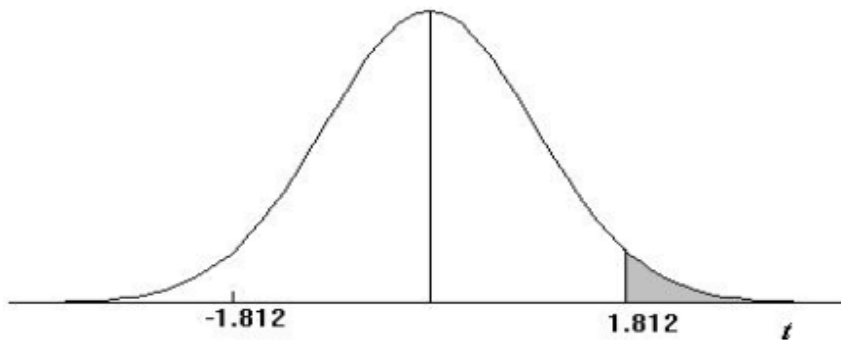
Gracias por su colaboración....!

**Anexo 2.** Cuadro de equivalencia de contenidos del capítulo III

<b>Temas del capítulo III</b>	<b>Capítulos del libro “Econometría” de Gujarati y Porter (2010)</b>
3.1 Aspectos básicos	-
3.2 Naturaleza del análisis de regresión	Capítulo 1. Naturaleza del análisis de regresión.
3.3 Análisis de regresión con dos variables: alguna ideas básicas	Capítulo 2. Análisis de regresión con dos variables: alguna ideas básicas
3.4 Análisis de regresión con dos variables: problema de estimación	Capítulo 3. Análisis de regresión con dos variables: problema de estimación
	Capítulo 4. Modelo clásico de regresión lineal normal (MCRL)
omitido	Capítulo 5. Regresión con dos variables: estimación por intervalos y pruebas de hipótesis
3.5. Extensiones del modelo de regresión lineal con dos variables	Capítulo 6. Extensiones del modelo de regresión lineal con dos variables
3.6 Análisis de regresión múltiple	Capítulo 7. Análisis de regresión múltiple: el problema de estimación
	Capítulo 8. Análisis de regresión múltiple: el problema de la inferencia
3.7 Modelos de regresión con variables dicótomas	Capítulo 9. Modelos de regresión con variables dicótomas
3.8 Multicolinealidad	Capítulo 10. Multicolinealidad: ¿qué pasa si las regresoras están correlacionadas?
3.9 Heteroscedasticidad	Capítulo 11. Heteroscedasticidad: ¿qué pasa si la varianza del error no es constante?
3.10 Autocorrelación	Capítulo 12. Autocorrelación: ¿qué pasa si los términos de error están correlacionados?
3.11 Creación de modelos econométricos	Capítulo 13. Creación de modelos econométricos: especificación del modelo y pruebas de diagnóstico
3.12 Modelos de regresión no lineales	Capítulo 14. Modelos de regresión no lineales
3.13 Modelos de regresión de respuesta cualitativa	Capítulo 15. Modelos de regresión de respuesta cualitativa
3.14 Modelos de regresión con datos de panel	Capítulo 16. Modelos de regresión con datos de panel



### Anexo 3. Tabla de distribución t de student



#### Ejemplo

Para  $r = 10$  grados de libertad:

$$P[t > 1.812] = 0.05$$

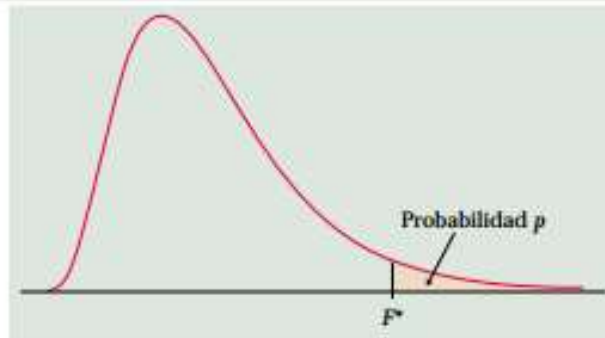
$$P[t < -1.812] = 0.05$$

$\alpha$ $r$	0,25	0,2	0,15	0,1	0,05	0,025	0,01	0,005	0,0005
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,656	636,578
2	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,600
3	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,768
24	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,689
28	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,660
30	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,679	0,848	1,045	1,296	1,671	2,000	2,390	2,660	3,460
120	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
$\infty$	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,290

**Fuente:** Manoleter (2013). *Tablas estadísticas (normal, t student, chi-cuadrado, fisher, binomial, poisson)* [diapositivas de PowerPoint]. Recuperado de <http://es.slideshare.net/manoleter/tablas-estadsticas-normal-t-student-chicuadrado-fisher-binomial-poisson>

$r = gl$  = grados de libertad  
 $\alpha$  = nivel de significancia

#### Anexo 4. Tabla de distribución F de Fisher

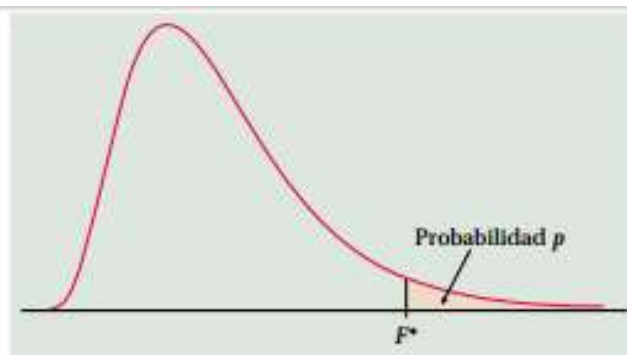


El valor de la tabla para  $p$  es el valor crítico  $F^*$  que deja la probabilidad  $p$  a la derecha

**TABLA D Valores críticos de la distribución F de Fisher**

		Grados de libertad en el numerador								
		1	2	3	4	5	6	7	8	9
Grados de libertad en el denominador	$p$									
	.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
	.050	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
	.025	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28
	.010	4052.2	4999.5	5403.4	5624.6	5763.6	5859.0	5928.4	5981.1	6022.5
	.001	405284	500000	540379	562500	576405	585937	592873	598144	602284
	.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
	.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
	.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
	.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
	.001	998.50	999.00	999.17	999.25	999.30	999.33	999.36	999.37	999.39
	.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
	.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
	.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
	.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
	.001	167.03	148.50	141.11	137.10	134.58	132.85	131.58	130.62	129.86
	.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
	.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
	.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
	.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
	.001	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47
	.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
	.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
	.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
	.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
	.001	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65	27.24
	.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
	.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
	.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
	.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
	.001	35.51	27.00	23.70	21.92	20.80	20.03	19.46	19.03	18.69
	.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
	.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
	.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
	.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
	.001	29.25	21.69	18.77	17.20	16.21	15.52	15.02	14.63	14.33

#### Anexo 4. Tabla de distribución F de Fisher (Continuación)



El valor de la tabla para  $p$  es el valor crítico  $F^*$  que deja la probabilidad  $p$  a la derecha

**TABLA D** Valores críticos de la distribución F de Fisher (cont.)

Grados de libertad del numerador										
10	12	15	20	25	30	40	50	60	120	1000
60.19	60.71	61.22	61.74	62.05	62.26	62.53	62.69	62.79	63.06	63.30
241.88	243.91	245.95	248.01	249.26	250.10	251.14	251.77	252.20	253.25	254.19
968.63	976.71	984.87	993.10	998.08	1001.4	1005.6	1008.1	1009.8	1014.0	1017.7
6055.8	6106.3	6157.3	6208.7	6239.8	6260.6	6286.8	6302.5	6313.0	6339.4	6362.7
605621	610668	615764	620908	624017	626099	628712	630285	631337	633972	636301
9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.47	9.48	9.49
19.40	19.41	19.43	19.45	19.46	19.46	19.47	19.48	19.48	19.49	19.49
39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.48	39.49	39.50
99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.48	99.49	99.50
999.40	999.42	999.43	999.45	999.46	999.47	999.47	999.48	999.48	999.49	999.50
5.23	5.22	5.20	5.18	5.17	5.17	5.16	5.15	5.15	5.14	5.13
8.79	8.74	8.70	8.66	8.63	8.62	8.59	8.58	8.57	8.55	8.53
14.42	14.34	14.25	14.17	14.12	14.08	14.04	14.01	13.99	13.95	13.91
27.23	27.05	26.87	26.69	26.58	26.50	26.41	26.35	26.32	26.22	26.14
129.25	128.32	127.37	126.42	125.84	125.45	124.96	124.66	124.47	123.97	123.53
3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.80	3.79	3.78	3.76
5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.70	5.69	5.66	5.63
8.84	8.75	8.66	8.56	8.50	8.46	8.41	8.38	8.36	8.31	8.26
14.55	14.37	14.20	14.02	13.91	13.84	13.75	13.69	13.65	13.56	13.47
48.05	47.41	46.76	46.10	45.70	45.43	45.09	44.88	44.75	44.40	44.09
3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.15	3.14	3.12	3.11
4.74	4.68	4.62	4.56	4.52	4.50	4.46	4.44	4.43	4.40	4.37
6.62	6.52	6.43	6.33	6.27	6.23	6.18	6.14	6.12	6.07	6.02
10.05	9.89	9.72	9.55	9.45	9.38	9.29	9.24	9.20	9.11	9.03
26.92	26.42	25.91	25.39	25.08	24.87	24.60	24.44	24.33	24.06	23.82
2.94	2.90	2.87	2.84	2.81	2.80	2.78	2.77	2.76	2.74	2.72
4.06	4.00	3.94	3.87	3.83	3.81	3.77	3.75	3.74	3.70	3.67
5.46	5.37	5.27	5.17	5.11	5.07	5.01	4.98	4.96	4.90	4.86
7.87	7.72	7.56	7.40	7.30	7.23	7.14	7.09	7.06	6.97	6.89
18.41	17.99	17.56	17.12	16.85	16.67	16.44	16.31	16.21	15.98	15.77
2.70	2.67	2.63	2.59	2.57	2.56	2.54	2.52	2.51	2.49	2.47
3.64	3.57	3.51	3.44	3.40	3.38	3.34	3.32	3.30	3.27	3.23
4.76	4.67	4.57	4.47	4.40	4.36	4.31	4.28	4.25	4.20	4.15
6.62	6.47	6.31	6.16	6.06	5.99	5.91	5.86	5.82	5.74	5.66
14.08	13.71	13.32	12.93	12.69	12.53	12.33	12.20	12.12	11.91	11.72

**Fuente:** Polisofando (2013). *Tablas distribuciones* [diapositivas de PowerPoint]. Recuperado de <http://www.slideshare.net/pilosofando/tablas-distribuciones>

**Anexo 5.** Tablas de datos

(Se adjunta en CD)