



UTPL
La Universidad Católica de Loja

Modalidad Abierta y a Distancia

Econometría Básica

Guía didáctica



Facultad de Ciencias Económicas y Empresariales

Departamento de Economía

Econometría Básica

Guía didáctica

Carrera	PAO Nivel
▪ Economía	IV

Autor:

Ochoa Ordoñez Oswaldo Francisco



Asesoría virtual
www.utpl.edu.ec

Universidad Técnica Particular de Loja

Econometría Básica

Guía didáctica

Ochoa Ordoñez Oswaldo Francisco

Diagramación y diseño digital:

Ediloja Cía. Ltda.

Telefax: 593-7-2611418.

San Cayetano Alto s/n.

www.ediloja.com.ec

edilojacialtda@ediloja.com.ec

Loja-Ecuador

ISBN digital - 978-9942-25-743-7



**Reconocimiento-NoComercial-CompartirIgual
4.0 Internacional (CC BY-NC-SA 4.0)**

Usted acepta y acuerda estar obligado por los términos y condiciones de esta Licencia, por lo que, si existe el incumplimiento de algunas de estas condiciones, no se autoriza el uso de ningún contenido.

Los contenidos de este trabajo están sujetos a una licencia internacional Creative Commons – **Reconocimiento-NoComercial-CompartirIgual 4.0 (CC BY-NC-SA 4.0)**. Usted es libre de **Compartir** – copiar y redistribuir el material en cualquier medio o formato. **Adaptar** – remezclar, transformar y construir a partir del material citando la fuente, bajo los siguientes términos: **Reconocimiento**– debe dar crédito de manera adecuada, brindar un enlace a la licencia, e indicar si se han realizado cambios. Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciatante. **No Comercial**-no puede hacer uso del material con propósitos comerciales. **Compartir igual**-Si remezcla, transforma o crea a partir del material, debe distribuir su contribución bajo la misma licencia del original. No puede aplicar términos legales ni medidas tecnológicas que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia. <https://creativecommons.org/licenses/by-nc-sa/4.0/>

Índice

1. Datos de información.....	9
1.1. Presentación de la asignatura	9
1.2. Competencias genéricas de la UTPL	9
1.3. Competencias específicas de la carrera.....	10
1.4. Problemática que aborda la asignatura.....	10
2. Metodología de aprendizaje.....	11
3. Orientaciones didácticas por resultados de aprendizaje.....	12
Primer bimestre	12
Resultado de aprendizaje 1	12
Contenidos, recursos y actividades de aprendizaje	12
Semana 1	12
Unidad 1. Introducción y naturaleza de los modelos de regresión	12
1.1. ¿Qué es la econometría y por qué es una disciplina aparte?	12
1.2. Metodología de la econometría	13
1.3. Origen histórico del término regresión	13
1.4. Interpretación moderna de la regresión.....	13
1.5. Relaciones estadísticas y relaciones deterministas.....	13
1.6. Regresión, causalidad y correlación	13
1.7. Naturaleza y fuentes de datos para el análisis económico.....	13
Actividades de aprendizaje recomendadas	15
Autoevaluación 1	17
Semana 2	19
Unidad 2. Análisis de regresión con 2 variables	19
2.1. Ejemplo hipotético	19
2.2. Concepto de función de regresión poblacional (FRP)	19
2.3. Significado del término lineal en variables y parámetros.....	19
2.4. Especificación estocástica de la FRP	19
2.5. Importancia del término de perturbación estocástica	19
2.6. Función de regresión muestral (FRM).....	19
2.7. Ejemplos ilustrativos.....	19
Actividades de aprendizaje recomendadas	24
Autoevaluación 2	25

Semana 3	27
 Unidad 3. Modelo de regresión con 2 variables: problema de estimación....	27
3.1. Método de mínimos cuadrados ordinarios (MCO)	27
3.2. Modelo clásico de regresión lineal: fundamentos del método de mínimos cuadrados. Advertencia sobre estos supuestos	27
3.3. Precisión o errores estándar de las estimaciones de mínimos cuadrados	27
3.4. Propiedades de los estimadores de mínimos cuadrados: teorema de Gauss-Markov.....	27
Actividades de aprendizaje recomendadas	32
Semana 4	32
3.5. Coeficiente de determinación r ² : una medida de la “bondad del ajuste”	32
3.6. Ejemplo numérico	32
3.7. Ejemplos ilustrativos.....	33
3.8. Una observación sobre los experimentos Montecarlo	33
Actividades de aprendizaje recomendadas	38
Autoevaluación 3	39
Semana 5	41
 Unidad 4. Modelo de regresión lineal normal (MCRLN)	41
4.1. Distribución de probabilidad de las perturbaciones ui	41
4.2. Supuesto de normalidad de ui. ¿Por qué debe formularse el supuesto de normalidad?	41
4.3. Propiedades de los estimadores de MCO según el supuesto de normalidad.....	41
Actividades de aprendizaje recomendadas	45
Autoevaluación 4	46
Resultado de aprendizaje 2 y 3	48
Contenidos, recursos y actividades de aprendizaje	48
Semana 6	48
 Unidad 5. Estimación por intervalos y pruebas de hipótesis	48
5.1. Requisitos estadísticos.....	48
5.2. Intervalos de confianza: ideas básicas.....	48
5.3. Intervalos de confianza para los coeficientes de regresión.....	48

5.4. Intervalo de confianza para la varianza	48
5.5. Pruebas de hipótesis: comentarios generales	48
Actividades de aprendizaje recomendadas	54
Semana 7	55
5.6. Pruebas de hipótesis: método del intervalo de confianza.....	55
5.7. Pruebas de hipótesis: enfoque de la prueba de significancia.....	55
5.8. Pruebas de hipótesis: algunos aspectos básicos.....	55
5.9. Análisis de regresión y de varianza.....	55
5.10. Aplicación del análisis de regresión: problema de predicción	55
Actividades de aprendizaje recomendadas	56
Autoevaluación 5	58
Semana 8	60
Actividades finales del bimestre.....	60
Actividades de aprendizaje recomendadas	60
Segundo bimestre	62
Resultado de aprendizaje 2	62
Contenidos, recursos y actividades de aprendizaje	62
Semana 9	62
Unidad 6. Extensiones del modelo de regresión lineal con dos variables....	62
6.1. Regresión a través del origen, r ² para el modelo de regresión a través del origen	62
6.2. Formas funcionales de los modelos de regresión.....	62
6.3. Cómo medir la elasticidad: modelo log-lineal	62
6.4. Modelos semilogarítmicos: log-lin y lin-log. El modelo lin-log	62
6.5. Modelos recíprocos	62
6.6. Elección de la forma funcional.....	62
Actividades de aprendizaje recomendadas	74
Autoevaluación 6	76
Resultado de aprendizaje 4	78
Contenidos, recursos y actividades de aprendizaje	78
Semana 10	78

Unidad 7. Análisis de regresión múltiple: el problema de estimación	78
7.1. Modelo con tres variables: notación y supuestos.....	78
7.2. Interpretación de la ecuación de regresión múltiple.....	78
7.3. Significado de los coeficientes de regresión parcial	78
7.4. Estimación de MCO y MV de los coeficientes de regresión parcial	78
7.5. R2 y R2 ajustada	78
Actividades de aprendizaje recomendadas	81
Semana 11	82
7.6. La función de producción Cobb-Douglas: más sobre la forma funcional	82
7.7. Modelos de regresión polinomial.....	82
7.8. Coeficientes de correlación parcial.....	82
Actividades de aprendizaje recomendadas	86
Autoevaluación 7	87
Semana 12	89
Unidad 8. Análisis de regresión múltiple: el problema de inferencia	89
8.1. Pruebas de hipótesis sobre coeficientes de regresión individuales	89
8.2. Prueba de significancia general de la regresión muestral.....	89
8.3. Prueba de igualdad de dos coeficientes de regresión.....	89
8.4. Mínimos cuadrados restringidos: pruebas de restricciones de igualdad lineales.....	89
8.5. Predicción con regresión múltiple.....	89
8.6. Prueba para la estabilidad estructural de los modelos de regresión: la prueba de Chow.....	89
8.7. Prueba de la forma funcional de la regresión: elección entre modelos de regresión lineal y log-lineal.....	89
Actividades de aprendizaje recomendadas	90
Resultado de aprendizaje 5.....	91
Contenidos, recursos y actividades de aprendizaje	91
Semana 13	91
Unidad 9. Modelos de regresión con variables dicótomas	91
9.1. Naturaleza de las variables dicótomas.....	91
9.2. Modelos ANOVA.....	91
9.3. Modelos ANCOVA	91
9.4. La variable dicótoma alternativa a la prueba de Chow	91

Actividades de aprendizaje recomendadas	95
Semana 14	95
9.5. Efectos de interacción al utilizar variables dicótomas	95
9.6. Uso de las variables dicótomas en el análisis estacional	95
9.7. Regresión lineal por segmentos.....	95
9.8. Algunos aspectos técnicos de la técnica con variables dicótomas	96
Actividades de aprendizaje recomendadas	98
Autoevaluación 8	100
Resultado de aprendizaje 6	102
Contenidos, recursos y actividades de aprendizaje	102
Semana 15	102
Unidad 10. Multicolinealidad: ¿qué pasa si las regresoras están correlacionadas?	102
10.1. Naturaleza de la multicolinealidad.....	102
10.2. Estimación en presencia de multicolinealidad.....	102
10.3. Consecuencias prácticas de la multicolinealidad.....	102
10.4. Detección de la multicolinealidad	102
10.5. Ejemplo ampliado.....	102
Actividades de aprendizaje recomendadas	103
Autoevaluación 9	104
Semana 16	106
Actividades finales del bimestre.....	106
Actividades de aprendizaje recomendadas	106
4. Solucionario	108
5. Referencias bibliográficas	117
6. Anexos	118



1. Datos de información

1.1. Presentación de la asignatura



1.2. Competencias genéricas de la UTPL

Son aquellas capacidades (actitudes, habilidades y conocimientos) comunes a todas las profesiones que se ofrece en la UTPL. Constituye una parte fundamental del perfil que el estudiante debe desarrollar durante su formación.

- Capacidad de abstracción, análisis y síntesis.
- Capacidad de aplicar los conocimientos en la práctica.
- Capacidad para organizar y planificar el tiempo.
- Conocimiento sobre el área de estudio y la profesión.
- Habilidades en el uso de las tecnologías de la información y de la comunicación.
- Capacidad de investigación.
- Habilidades para buscar, procesar y analizar información procedente de fuentes diversas.
- Capacidad crítica y autocrítica.
- Capacidad creativa.

- Capacidad para identificar, plantear y resolver problemas.
- Capacidad para tomar decisiones.
- Habilidades interpersonales.
- Compromiso ético.

1.3. Competencias específicas de la carrera

ECONOMÍA

- Analiza e interpreta el entorno económico social.
- Desarrolla el pensamiento matemático y estadístico para la aplicación y análisis de aspectos económicos.

1.4. Problemática que aborda la asignatura

En la actualidad parte de las teorías económicas no aplican componentes matemáticos o estadísticos a la hora de resolver problemas económicos. Es por esto que muchos investigadores poseen débiles competencias numéricas para demostrar empíricamente la evolución de los fenómenos económicos, que ayudan en gran parte a entender y pronosticar la evolución de estas economías; es por esto que la asignatura de Econometría Básica es importante a la hora de poder resolver los problemas antes mencionados, ya que permite al estudiante comprender los métodos y modelos econométricos para la representación y solución de problemas económicos y financieros.



2. Metodología de aprendizaje

A través del aprendizaje por descubrimiento y el aprendizaje basado en problemas, usted va a comprender y analizar cada uno de los elementos que se contemplan en la planificación semanal.

Para ello, lo invito a revisar los materiales y recursos educativos que se han previsto en cada uno de los temas a desarrollarse en las semanas que comprende el período académico.

Mediante la lectura de los documentos elaborados y también con las orientaciones académicas que reciba por parte de su profesor, podrá descubrir la utilidad de las medidas e indicadores que le permitan lograr los resultados de aprendizaje y por tanto el desarrollo de las competencias profesionales.

La aplicabilidad se la descubre mediante el desarrollo de problemas, de ahí que es importante que vaya revisando los ejercicios desarrollados en el texto básico, así como también desarrolle otros ejercicios que se encuentran propuestos en el texto.

Si desea conocer lo que significa esta metodología de aprendizaje, lo invito a revisar su explicación en los siguientes enlaces:

- [Aprendizaje por descubrimiento](#)
- [Aprendizaje basado en problemas](#)



3. Orientaciones didácticas por resultados de aprendizaje



Primer bimestre

Resultado de aprendizaje 1

- Analiza la relación lineal entre dos variables y predice la dependencia de una con respecto a la otra.

Para lograr este primer resultado de aprendizaje, es necesario partir de la identificación de los elementos conceptuales de lo que comprende el ámbito de un modelo de regresión, por ello, lo invito a que asuma con claridad los diferentes conceptos, comprendiendo su significado, alcance y aplicabilidad en cada uno de los momentos de aprendizaje.

Cuando usted ha logrado comprender cada uno de los elementos conceptuales, podrá entender a que se refieren los modelos de regresión, su utilidad y estructura.

El análisis de la regresión le permite conocer la relación existente entre dos variables, y a través de ciertos estadísticos la significatividad de los resultados de esta regresión.

Contenidos, recursos y actividades de aprendizaje



Semana 1

Unidad 1. Introducción y naturaleza de los modelos de regresión

1.1. ¿Qué es la econometría y por qué es una disciplina aparte?

1.2. Metodología de la econometría

1.3. Origen histórico del término regresión

1.4. Interpretación moderna de la regresión

1.5. Relaciones estadísticas y relaciones deterministas

1.6. Regresión, causalidad y correlación

1.7. Naturaleza y fuentes de datos para el análisis económico

Estimado alumno, empezamos el estudio de esta asignatura revisando la primera parte. Aquí se busca entender el concepto de econometría, así como una descripción general de la metodología econométrica. A continuación, se detallan algunos.

“La econometría se basa en métodos estadísticos para estimar relaciones económicas, poner a prueba teorías económicas y evaluar y poner en práctica políticas gubernamentales y comerciales” (Wooldridge, 2001).

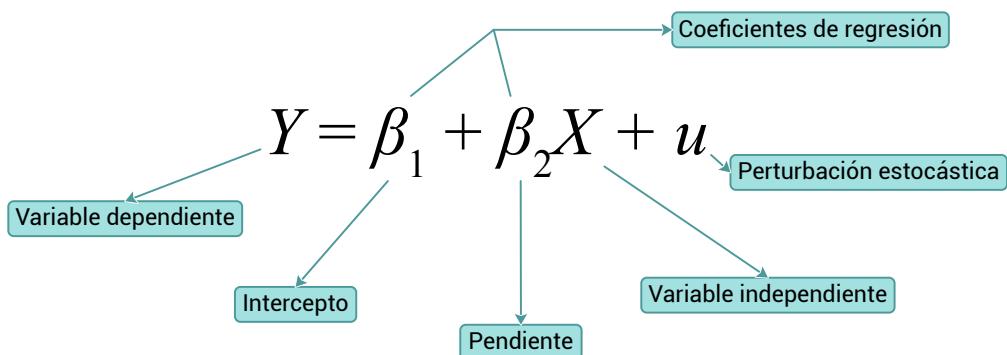
“La econometría puede ser definida como la ciencia social en la cual las herramientas de la teoría económica, las matemáticas y la inferencia estadística son aplicadas al análisis de los fenómenos económicos” (Golberger, 1964).

Para realizar el análisis de un problema económico se presenta la metodología tradicional o clásica que puede ser considerada como el camino a seguir para estructurar un modelo de regresión simple, es decir, aquel modelo que solo tiene 2 variables cuantitativas, el cual está estructurado de la siguiente forma.

Comenzaremos entendiendo cuál es el formato general de una ecuación de regresión simple y los factores que la componen:

Figura 1.

Estructura de un modelo de regresión simple



A continuación, le invito a revisar el recurso interactivo sobre Metodología econométrica.

Metodología econométrica

Para entender mejor esta metodología realizaremos un ejercicio práctico, el cual se describe a continuación: [Ejercicio práctico](#).

Interpretación moderna de la regresión

Comenzaremos analizando un concepto básico de regresión.

El análisis de regresión trata del estudio de la dependencia de la variable dependiente, respecto a una o más variables (las variables explicativas), con el objetivo de estimar y/o predecir la media o valor promedio poblacional de la primera en términos de los valores conocidos o fijos (en muestras repetidas) de las últimas. (Gujarati y Porter, 2010)

Con el siguiente ejemplo clarificaremos la importancia del concepto de regresión.

Un economista puede estar interesado en estudiar la dependencia del gasto de consumo personal en el ingreso personal neto disponible (después de impuestos). Un análisis de este tipo puede ser útil para estimar la propensión marginal a consumir (PMC), esto es, el cambio promedio en el gasto de consumo ante un cambio de un dólar. (Gujarati y Porter, 2010)

Los temas conceptuales sobre regresión, causalidad y correlación revíselos en el texto básico a partir de la página 19 del capítulo 1.

Recursos de aprendizaje

Los recursos que va a utilizar para este resultado de aprendizaje son:

Lecturas

- Gujarati, D. & Porter, D. (2010). *Econometría*. México: McGraw Hill. Quinta edición.

 El libro le proveerá de todo el marco teórico acerca del concepto de regresión y su aplicabilidad dentro de las ciencias sociales. Deberá leer la parte de introducción y el capítulo 1.

- Wooldridge, J. (2010). *Introducción a la econometría*. México: Thomson – Learning.

El libro le ayudará a entender el tema de los modelos de regresión, su concepto y utilidad, para esto revise el capítulo 1.

Una vez revisado el texto básico y texto complementario, continuemos con la revisión de la guía virtualizada, la misma que nos indica a través de ejemplos prácticos el desarrollo de esta unidad 1.



Actividades de aprendizaje recomendadas

▪ Actividad 1

Reflexione sobre los primeros temas trabajados en la guía virtualizada y en el capítulo 1 del texto básico.

Procedimiento: Para que comprenda los temas desarrollados le aconsejo que realice cuadros sinópticos o resúmenes en los que tome las ideas principales con la finalidad de que tenga un documento de trabajo que posteriormente le permita revisar y comprender cada

tema. Utilice las técnicas que de acuerdo a su estilo de aprendizaje le sean de mayor utilidad.

Nota. conteste las actividades en un cuaderno de apuntes o en un documento Word en caso de ser necesario.

- **Actividad 2**

Autoevaluación 1, estimado estudiante, para evaluar su aprendizaje puede realizar la autoevaluación de la unidad 1.

Procedimiento: En esta actividad debe contestar las preguntas al final de la unidad 1, esta tarea tiene la finalidad de comprobar que usted está teniendo sintonía con los temas tratados.

Recuerde que la autoevaluación no es obligatoria, sin embargo, le ayuda a reforzar sus conocimientos y lo prepara para la evaluación bimestral.

¡Vamos a trabajar!



Autoevaluación 1

Considerando el siguiente enunciado, seleccione la alternativa correcta.

1. En el análisis de regresión lo que interesa es la dependencia:
 - a. Funcional.
 - b. Determinista.
 - c. Estadística.

2. En las relaciones estadísticas entre variables se analizan:
 - a. Variables aleatorias.
 - b. Errores estándar.
 - c. Formas funcionales.

3. El análisis de correlación mide:
 - a. Fuerza de asociación.
 - b. Normalidad.
 - c. Errores estándar.

4. A la variable dependiente también se la conoce como variable:
 - a. Respuesta.
 - b. Estímulo.
 - c. Covariante.

5. A la variable independiente también se la conoce como variable:
 - a. Explicada.
 - b. Controlada.
 - c. Predictora.

6. Información de series de tiempo es:
 - a. Salarios cantón Loja 2016.
 - b. PIB anual 1980 – 2016.
 - c. Inflación de países latinoamericanos 2016.

7. Información de corte transversal es:
- a. Inflación mensual 2000 – 2016.
 - b. Precios diarios de las acciones de la bolsa de valores.
 - c. Nivel de escolaridad cantón Loja año 2017.
8. Información de panel es:
- a. Información en un mismo periodo de tiempo.
 - b. Información en distintos periodos de tiempo.
 - c. Información combinada de corte transversal y serie de tiempo.

[Ir al solucionario](#)



Unidad 2. Análisis de regresión con 2 variables

2.1. Ejemplo hipotético

2.2. Concepto de función de regresión poblacional (FRP)

2.3. Significado del término lineal en variables y parámetros

2.4. Especificación estocástica de la FRP

2.5. Importancia del término de perturbación estocástica

2.6. Función de regresión muestral (FRM)

2.7. Ejemplos ilustrativos

Lo que se va a revisar a continuación es la teoría básica del análisis de regresión de la forma más sencilla posible, es decir, la regresión bivariante o con dos variables, en la cual la variable dependiente (la regresada) se relaciona con una sola variable explicativa (la regresora).

Conocemos que para establecer un modelo de regresión simple se deben definir las variables, sobre todo, cuál es la variable controlada y cuál la de control.

La ecuación aplicada para la función de regresión poblacional es:

Consumo en función del ingreso:

$$C = f(Y)$$

La regresión lineal permite obtener los valores esperados o la media poblacional del consumo ante una variación en los ingresos.

$$E(y/X) = f(X)$$

$$E(y/x_i) = \beta_1 + \beta_2 X_i + \mu$$

Los resultados obtenidos en esta regresión fueron los promedios de consumo de cada familia ante una variación en sus ingresos, los mismos que mantienen una relación directa con la variable dependiente, que basado en la teoría económica manifiesta que las personas ante un mayor nivel de ingreso tienden a aumentar su nivel de consumo.

Además, revisaremos el término de linealidad, el cual es aplicable tanto para los parámetros como para las variables, y la importancia de que esta linealidad se mantenga en los parámetros para poder realizar los modelos de regresión. Revise lo siguiente.

Linealidad en las variables (X o Y)

Para que un modelo sea considerado como lineal en sus variables, estas no pueden estar elevadas a un exponente mayor a uno, ni multiplicada o dividida para otra variable. Para que el modelo sea lineal este debe darse en todas las variables.

$$Y_i = \beta_1 + \beta_2 + X_i$$

Ejemplos de linealidad:

$$Y_i = \beta_1 + \beta_2 X_i^2 \quad Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 X_i^3$$

Ejemplos de no linealidad:

$$Y_i = e^{\beta_1 + \beta_2 X_i} \quad Y_i = \beta_1 + \beta_2 \left(\frac{1}{x_i} \right) + \mu$$

$$\ln Y_i = \beta_1 + \beta_2 X_i^2 + \mu$$

Linealidad en los parámetros β_1, β_2

Aquí se aplica el mismo concepto de linealidad solo que ahora es para los parámetros.

$$E(Y|Xi) = \beta_1 + \beta_2 Xi \quad E(Y|X_i) = \beta_1 + \beta_2 X_i^2$$

Ejemplo 1

Aquí algunos ejemplos de distintos tipos de regresiones:

- Lineal. $Y = \beta_1 + \beta_2 X$
- Logarítmica. $Y = \beta_1 + \beta_2 \ln X$
- Inversa. $Y = \beta_1 + \frac{\beta_2}{X}$
- Cuadrático. $Y = \beta_1 + \beta_2 X + \beta_3 X^2$
- Cúbico. $Y = \beta_1 + \beta_2 X + \beta_3 X^2 + \beta_4 X^3$
- Potencia. $Y = \beta_1 X^{\beta_2}$ o $\ln Y = \ln \beta_1 + \beta_2 \ln X$
- Compuesto. $Y = \beta_1 \beta_2^X$ o $\ln Y = \ln \beta_1 + X \ln \beta_2$
- Curva-S. $e^{\beta_1} + \frac{\beta_2}{X}$ o $\ln Y = \beta_1 + \frac{\beta_2}{X}$
- Crecimiento. $e^{\beta_1} + \frac{\beta_2}{X}$ o $\ln Y = \beta_1 + \frac{\beta_2}{X}$
- Exponencial. $Y = \beta_1 e^{\beta_2 X}$ o $\ln Y = \ln \beta_1 + \beta_2 X$

Continuando con el estudio del capítulo es importante conocer cómo funciona el término de perturbación estocástica, el mismo que indica la dispersión o desviación de un E individual alrededor de su valor esperado. Estas desviaciones son causadas por otras variables que afectan a la variable dependiente cuyos efectos en su conjunto son expresados por el error estocástico (Ochoa,2011.).

$$\begin{aligned}\mu_i &= Y_i - E(Y/X_i) \\ Y_i &= E(Y/X_i) + \mu_i\end{aligned}$$

Para obtener resultados confiables ($\hat{\beta}_1$ y $\hat{\beta}_2$) en la regresión, se debe suponer que no existe ninguna relación entre variable X y el valor esperado de la variable u.

$$E(\mu) = 0$$

$$E(\mu/x) = E(\mu) = 0$$

Como u y X son variables aleatorias, podemos definir la distribución condicional de u dado cualquier valor de X . En particular para cualquier X , podemos obtener el valor esperado o promedio de u para aquel sector de la población que describe esa X . La suposición crucial es que el valor promedio de u no depende de X .

Ejemplo 2

$$Y_i = \beta_1 + \beta_2 X_i + \mu$$

$$\text{cosecha} = \beta_1 + \beta_2 \text{fertilizante} + \mu$$

- En la que u podría representar variables como la lluvia, el sol, los años de vida del terreno, etc., todas aquellas variables que ejercen algún efecto sobre Y.

$$Y_i = \beta_1 + \beta_2 X_i + \mu$$

$$\text{salario} = \beta_1 + \beta_2 \text{educación} + \mu$$

- En la que u podría representar variables como años de experiencia, número de cursos, sexo, religión etc., todas aquellas variables que ejercen algún efecto sobre Y.

$$Y_i = \beta_1 + \beta_2 X_i + \mu$$

$$\text{calificación} = \beta_1 + \beta_2 \text{asistencia} + \mu$$

- En la que u podría representar variables como conducta, participación, trabajo entregados etc., todas aquellas variables que ejercen algún efecto sobre Y (Wooldridge, 2009).

Como acabamos de revisar, el término “perturbación estocástica”, u , es un sustituto para todas aquellas variables que son omitidas en el modelo, pero que de manera conjunta ejercen un efecto sobre la variable dependiente.

Para conocer cuáles son las razones por las que no se desarrolla un modelo con todas estas variables que agrupa el error estocástico, revise el tema 2.5 de su texto básico.

Por último, veremos que para la aplicación o construcción de modelos de regresión debemos trabajar con una función de regresión muestral (FRM), es decir, trabajar con información a partir de una muestra, la que buscará a través de sus resultados ser lo más cercana a una función de regresión poblacional (FRP).

Ahora analizaremos la Función de Regresión Muestral (FRM), la misma que parte de la ecuación de regresión poblacional, es decir, estimaremos la Función de Regresión Poblacional (FRP), con base en información muestral.

$$Y_i = \beta_1 + \beta_2 X_i + \mu_i \quad FRP$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\mu}_i \quad FRM$$

Este es el procedimiento normal que se utiliza para la estimación de un modelo de regresión, en la cual es importante que las variables que se utilice tengan la información necesaria para poder ser aplicadas. Para entender un poco más este tema revise los ejemplos ilustrativos del literal 2.7 de su texto básico.

Recursos de aprendizaje

Los recursos que va a utilizar para este resultado de aprendizaje son:

Lecturas

- Gujarati, D. & Porter, D. (2010). *Econometría*. México: McGraw Hill. Quinta edición.

El libro le proveerá de todo el marco teórico acerca del concepto de regresión y su aplicabilidad dentro de las ciencias sociales. Deberá leer el capítulo 2.

- Wooldridge, J. (2010). *Introducción a la econometría*. México: Thomson–Learning.

El libro le ayudará a entender el tema de los modelos de regresión, su concepto y utilidad, para esto revise nuevamente el capítulo 1.



Actividades de aprendizaje recomendadas

■ Actividad 1

Reflexione sobre los primeros temas trabajados en la guía didáctica y en el capítulo 2 del texto básico.

Procedimiento: Para que comprenda los temas desarrollados le aconsejo que realice cuadros sinópticos o resúmenes en los que tome las ideas principales con la finalidad de que tenga un documento de trabajo que posteriormente le permita revisar y comprender cada tema.

Nota. conteste las actividades en un cuaderno de apuntes o en un documento Word en caso de ser necesario.

■ Actividad 2

Estimado estudiante, para evaluar su aprendizaje puede realizar la autoevaluación de la unidad 2.

Procedimiento: En esta actividad debe contestar las preguntas al final de la unidad 2, esta tarea tiene la finalidad de comprobar que usted está teniendo sintonía con los temas tratados.

Recuerde que la autoevaluación no es obligatoria, sin embargo, le ayuda a reforzar sus conocimientos y lo prepara para la evaluación bimestral.

¡Vamos a trabajar!



Autoevaluación 2

Considerando el siguiente enunciado, seleccione la alternativa correcta.

1. Linealidad es aquella en que la esperanza condicional de Y es:
 - a. Función lineal de Y.
 - b. Función lineal de u .
 - c. Función lineal de X.
2. ¿Cuál de las siguientes funciones es lineal en sus parámetros?
 - a. $\ln Y = B_1 + B_2 X + u$.
 - b. $\ln Y = B_1 + B_2 X + u$.
 - c. $Y = \ln B_1 + B_2 + u$.
3. La función de regresión poblacional nos muestra que el valor condicional de Y dado X es:
 - a. Una función de X.
 - b. Una función de Y.
 - c. Una función de X y Y.
4. Uno de los coeficientes de la ecuación de regresión es:
 - a. Coeficiente de correlación.
 - b. El intervalo de confianza.
 - c. El intercepto.
5. Vaguedad de la teoría se refiere a:
 - a. A la falta de datos.
 - b. A que u agrupe variables faltantes u omitidas.
 - c. Modelo mal especificado.
6. Las variables proxy se refiere a variables:
 - a. Representantes.
 - b. Logarítmicas.
 - c. Recíprocas.

7. El principio de parsimonia manifiesta que:
 - a. Gran cantidad de datos.
 - b. Modelo lo más sencillo posible.
 - c. Modelo mal especificado.

8. Cuando hablamos de la forma funcional incorrecta, nos referimos a:
 - a. Relación funcional entre las variables.
 - b. Aplicar pruebas de hipótesis.
 - c. Establecer malos intervalos de confianza.

[Ir al solucionario](#)



Unidad 3. Modelo de regresión con 2 variables: problema de estimación

3.1. Método de mínimos cuadrados ordinarios (MCO)

3.2. Modelo clásico de regresión lineal: fundamentos del método de mínimos cuadrados. Advertencia sobre estos supuestos

3.3. Precisión o errores estándar de las estimaciones de mínimos cuadrados

3.4. Propiedades de los estimadores de mínimos cuadrados: teorema de Gauss-Markov

En la siguiente unidad estudiaremos el método generalmente utilizado para la estimación de modelos de regresión, cabe resaltar que existe una gran variedad de métodos para realizar regresiones entre ellos tenemos el de mínimos cuadrados ordinarios, el de máxima verosimilitud (MV), el método de momentos generalizados, entre otros.

Para el estudio de la econometría, utilizaremos el método de mínimos cuadrados ordinarios (MCO), ya que, con base en la teoría, proporciona resultados estadísticos significativos y está regido por una serie de supuestos y propiedades que lo vuelven bastante confiable a la hora de realizar regresiones. Adicionalmente, revisaremos el cálculo de los residuos los mismos que se generan al momento de estimar la regresión.

En resumen, dentro de esta unidad revisaremos dos puntos importantes tal como interpretar los resultados de la regresión y si los resultados obtenidos son estadísticamente significativos, es decir, son confiables para ser utilizados en una investigación.

Para entender este tema lo haremos a través de algunos ejercicios.

Ejemplo 3

Ejercicio 4.1. Gasto de consumo en personal (Gujarati, 2010)

El siguiente ejercicio muestra información sobre Y (gasto de consumo personal) y X (PIB), desde 1982-1996, en miles de millones de dólares de 1992. Tabla I.1 del texto básico, página 6. Aplicaremos la metodología econométrica conforme vayamos avanzando en esta unidad, inicialmente lo haremos con lo siguiente:

1. Planteo teórico o de hipótesis. Gasto de consumo de personal= f (PIB)
o $Y = f(X)$
2. Especificación del modelo matemático $Y = B_1 + B_2 X$
3. Especificación modelo econométrico $Y = B_1 + B_2 X + u$
4. Obtención de datos. Tabla I.1
5. Estimación de la regresión.

Tabla 1.

Información del gasto de personal y el PIB (miles de millones de dólares 1992)

Obs	Y	X	$(X)^2$	XY	\hat{Y}	\hat{u}	\hat{u}^2	$(X)^2$	$(Y)^2$	
1982	3081,5	4620,3	21347172,1	14237454,5	3077,01	4,49	20,14	1567236,92	778965,11	
1983	3240,6	4803,7	23075533,7	15566870,2	3206,46	34,14	1165,6	1141678	523437,78	
1984	3407,6	5140,1	26420628	17515404,8	3443,9	-36,3	1317,41	535960,65	309681,12	
1985	3566,5	5323,5	28339652,3	18986262,8	3573,34	-6,84	46,83	301064,37	158077,81	
1986	3708,7	5487,7	30114851,3	20352233	3689,24	19,46	378,75	147835,12	65224,05	
1987	3822,3	5649,5	31916850,3	21594083,9	3803,44	18,86	355,71	49592,32	20104,40	
1988	3972,7	5865,2	34400571	23300680	3955,68	17,02	289,52	48,91	74,13	
1989	4064,6	6062	36747844	24639605,2	4094,59	-29,99	899,38	36026,57	10102,26	
1990	4132,2	6136,3	37654177,7	25356418,9	4147,03	-14,83	219,98	69752,33	28260,97	
1991	4105,8	6079,4	36959104,4	24960800,5	4106,87	-1,07	1,15	42934,6	20081,72	
1992	4219,8	6244,4	38992531,4	26350119,1	4223,33	-3,53	12,47	138537,8	65387,60	
1993	4343,6	6389,6	40826988,2	27753866,6	4325,82	17,78	316,29	267709,66	144027,84	
1994	4486	6610,7	43701354,5	29655600,2	4481,87	4,13	17,04	545392,1	272390,05	
1995	4595,3	6742,1	45455912,4	30981972,1	4574,62	20,68	427,82	756737,61	398426,06	
1996	4714,1	6928,4	48002726,6	32661170,4	4706,11	7,99	63,84	1115572,52	562515,00	
Σ	59461,3	88082,9	523955898	353912542	59409,31			5531,93	6716079,49	3356755,92

Nota. Adaptado de Gujarati, D. (2010). *Econometría*. México: McGraw Hill. Quinta edición.

Debemos iniciar calculando las sumatorias de (Y, X, XY, X^2) descritos en la tabla 1. Una vez obtenidos procedemos al cálculo de los coeficientes de la regresión. Para esto utilizamos las siguientes fórmulas:

$$\hat{\beta}_1 = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n \sum X_i^2 - (\sum X_i)^2}$$

$$\hat{\beta}_1 = \frac{(523955897,65)(59461,30) - (88082,90)(353912542,09)}{(15)(523955897,65) - (88082,90)^2}$$

$$\hat{\beta}_1 = -299.59$$

La interpretación de este parámetro es que, si el ingreso o el PIB fueran de cero, el gasto de consumo de personal sería de -299 mil millones de dólares.

$$\hat{\beta}_2 = \frac{n \sum X_i Y_i - \sum Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2}$$

$$\hat{\beta}_2 = \frac{(15)(353912542,09) - (59461,30) - (88082,90)(59461,30)}{(15)(523955897,65) - (88082,90)^2}$$

$$\hat{\beta}_2 = 0,71$$

La interpretación de este parámetro indica que, si el PIB se incrementa en un dólar, el gasto de consumo personal promedio aumenta en casi 71 centavos.

Estos valores obtenidos se conocen como estimadores de mínimos cuadrados, en donde $\hat{\beta}_1$ representa el intercepto y $\hat{\beta}_2$ la pendiente. Con los valores de los coeficientes procedemos a establecer los resultados de la ecuación de regresión. Como lo dijimos en la unidad anterior debemos utilizar la función de regresión muestral (FRM).

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$$

$$\hat{Y}_i = -299,59 + 0,71 X_i + \mu_i$$

Los residuos se generan por cada uno de los datos de la variable dependiente, pero antes de poder calcularlos debemos generar los valores de la Y estimada (\hat{Y}), para esto utilizamos la ecuación de regresión y reemplazamos el valor de X de la siguiente manera:

$$\begin{aligned}\hat{Y}_1 &= -299,59 + 0,71X_1 \\ \hat{Y}_1 &= -299,59 + 0,71(4620,30) \\ \hat{Y}_1 &= 2980,82 \\ \hat{Y}_2 &= -299,59 + 0,71X_2 \\ \hat{Y}_2 &= -299,59 + 0,71(4803,70) \\ \hat{Y}_2 &= 3111,03\end{aligned}$$

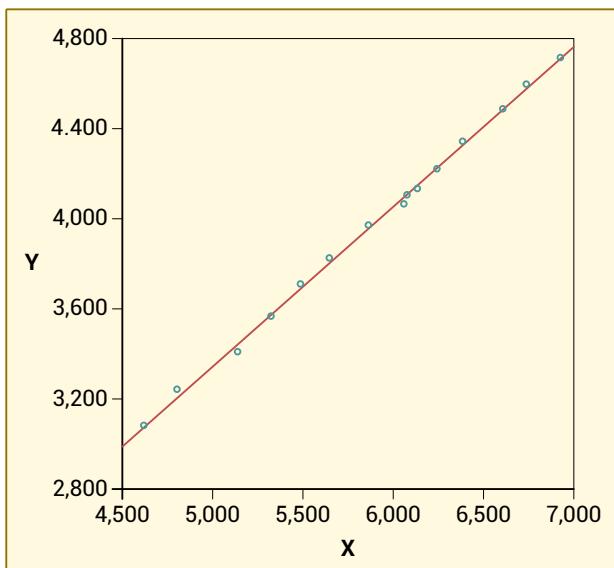
De igual manera se procede con los demás datos de X, lo que al final nos permite es obtener los valores de las Y estimadas y posteriormente la $\Sigma \hat{Y}$. Luego de haber obtenido las Y estimadas, procedemos al cálculo de los residuos:

$$\hat{u} = Y - \hat{Y}$$

$$\begin{aligned}\hat{u}_1 &= 3081,50 - 2980,82 \\ \hat{u}_1 &= 100,68 \\ \hat{u}_2 &= 3240,60 - 3111,03 \\ \hat{u}_2 &= 129,57\end{aligned}$$

De igual manera se procede con los demás datos de Y, lo que al final nos permite es obtener los valores de las u estimadas y posteriormente la $\Sigma \hat{u}$.

Figura 2.
Línea de regresión



Hemos realizado la gráfica para determinar a simple vista como los datos se ajustan a la línea de regresión. En este ejemplo podemos determinar que el PIB explica en gran medida las variaciones que se dan en el gasto de consumo de personal.

Ahora procederemos a realizar este mismo ejercicio en el software STATA, los pasos a seguir serán los siguientes: [Ejercicio en el software STATA](#).

Recursos de aprendizaje

Los recursos que va a utilizar para este resultado de aprendizaje son:



Lecturas

- Gujarati y Porter (2010). *Econometría*. México: McGraw Hill. Quinta edición.

El libro le proveerá de todo el marco teórico acerca del concepto de regresión y su aplicabilidad dentro de las ciencias sociales. Deberá leer el capítulo 3.

- Wooldridge (2010). Introducción a la econometría. México: Thomson – Learning.

El libro le ayudará a entender el tema de los modelos de regresión, su concepto y utilidad, para esto revise nuevamente el capítulo 2.

Video



- Cuellar L. (2019) [Regresión lineal en STATA: interpretación de salida.](#)



Actividades de aprendizaje recomendadas

▪ Actividad 1

Reflexione sobre los temas trabajados en la guía virtualizada y en el capítulo 3 del texto básico.

Procedimiento: Para que comprenda los temas desarrollados le aconsejo que realice cuadros sinópticos o resúmenes en los que tome las ideas principales con la finalidad de que tenga un documento de trabajo que posteriormente le permita revisar y comprender cada tema.

Nota. conteste las actividades en un cuaderno de apuntes o en un documento Word en caso de ser necesario.



Semana 4

3.5. Coeficiente de determinación r²: una medida de la “bondad del ajuste”

3.6. Ejemplo numérico

3.7. Ejemplos ilustrativos

3.8. Una observación sobre los experimentos Montecarlo

PRECISIÓN O ERRORES ESTÁNDAR DE LOS MÍNIMOS CUADRADOS ESTIMADOS

Dentro de los resultados de la ecuación de regresión, se busca alguna medida de confiabilidad o precisión de los coeficientes de β_1 y β_2 . Por lo tanto, en la estadística la precisión de un valor estimado está dado por los errores estándar (ee).

Gujarati y Porter (2010) manifiestan que el error estándar del valor estimado o el error estándar de la regresión (ee), es simplemente la desviación estándar de los valores y alrededor de la recta de regresión estimada, la cual es utilizada frecuentemente como una medida resumen de la “bondad de ajuste” de dicha recta.

Adicional, los errores estándar nos permiten generar intervalos de confianza para procedimientos econométricos posteriores. Para que un modelo sea adecuado debe tener errores mínimos, lo que da confiabilidad a la estimación. Para que un error sea mínimo este debe ser máximo la tercera parte del valor del parámetro, es decir, dividir el valor del parámetro para tres y compararlo con el del error. Si este es menor a la tercera parte del valor del parámetro es confiable, caso contrario no lo es.



Finalizado este tema, tendremos toda la información necesaria para continuar con al paso 7 de la metodología econométrica, ya que nos permite ahora realizar la VERIFICACIÓN ESTADÍSTICA.

Para entender este tema de mejor manera continuemos con la revisión del ejercicio planteado en la unidad anterior.

Continuación del ejercicio 4.1

Aplicando las fórmulas anteriormente descritas procedemos a calcular los errores estándar de esta regresión:

$$\begin{array}{lll}
\hat{\delta}^2 = \frac{\sum \hat{\mu}^2_i}{n-K} & Var(\hat{\beta}_1) = \frac{\sum X^2_i}{n \sum X^2} \delta^2 & Var(\hat{\beta}_2) = \frac{\delta^2}{\sum X^2} \\
\hat{\delta}^2 = \frac{5531,93}{15-2} & Var(\hat{\beta}_1) = \frac{523955897,7}{15(6716079,49)} (425,53) & Var(\hat{\beta}_2) = \frac{425,53}{6716079,49} \\
\hat{\delta}^2 = 425,53 & Var(\hat{\beta}_1) = 2213,186 & Var(\hat{\beta}_2) = 0,0000633 \\
& ee(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)} & ee(\hat{\beta}_2) = \sqrt{Var(\hat{\beta}_2)} \\
& ee(\hat{\beta}_1) = \sqrt{2213,186} & ee(\hat{\beta}_2) = \sqrt{0,0000633} \\
& ee(\hat{\beta}_1) = 28,76 & ee(\hat{\beta}_2) = 0,0044
\end{array}$$

Para constatar si los errores estándares calculados son confiables, en primer lugar, procedemos a obtener la tercera parte de los parámetros obtenidos en la regresión.

$$\hat{Y}_i = -299,59 + 0,72X_i + \mu_i$$

$$\begin{array}{ll}
\hat{\beta}_1 = -299,59 / 3 & \hat{\beta}_2 = 0,72 / 3 \\
\hat{\beta}_1 = -99,86 & \hat{\beta}_2 = 0,23 \\
ee(\hat{\beta}_1) 28,76 \prec -99,86 & ee(\hat{\beta}_2) 0,0044 \prec 0,23 \\
Es confiable & Es confiable
\end{array}$$

Con base en los resultados podemos concluir que los estimadores de la regresión son confiables ya que sus (ee) han sido menores a la tercera parte del valor del parámetro. Para realizar las comparaciones entre los valores no se debe tomar en cuenta los signos de los parámetros.

En el software STATA serían los siguientes resultados:

Figura 3.

Resultados 1, el software STATA

The screenshot shows the Stata/SE 14.0 interface with the following details:

- Menu Bar:** Archivo, Edición, Datos, Gráficos, Estadísticas, Usuario, Ventana, Ayuda.
- Toolbar:** Includes icons for file operations, search, and help.
- Command Window:** Displays the command history and output. A red box highlights the coefficient table for the first regression.
- Output Window:** Shows the regression results for PCEY on GDPX, including the model summary and coefficient table.
- Variables View:** Lists variables with their names and labels (Year, PCEY, GDPX, RES, YEST).
- Properties View:** Shows detailed properties for the selected variable GDPX.
- File Path:** C:\Users\oficial\Documents

Below is the text from the Command Window:

```
# Comando _rc
1 import excel "..."
2 reg PCEY GDPX
3 destring Year, ...
4 tset Year, yearly
5 reg PCEY GDPX
6 gen RES, residuals 100
7 predict RES, re...
8 predict YEST, xb
9 twoway scatter...
```

Below is the text from the Output Window:

```
. reg PCEY GDPX
Source          SS           df        MS      Number of obs   =       46
               Model    144125684     1  144125684  Prob > F        =  0.0000
                  Residual  230131.858    44  5412.08768  R-squared       =  0.9984
                                         Adj R-squared =  0.9983
                                         Root MSE       =    73.567
PCEY |      Coef.  Std. Err.      t  P>|t| [95% Conf. Interval]
GDPX | .7210343  .0044233  163.19  0.000  .7129197  .730749
_cons | -299.5913 28.76494 -10.42  0.000 -357.5632 -241.6194
.
gen RES, residuals
```

Los mismos resultados calculados anteriormente se reflejan en el software STATA, tanto los valores de los coeficientes como sus errores estándar se reflejan en el recuadro color rojo.

COEFICIENTE DE DETERMINACIÓN (R^2)

Hasta aquí no tenemos forma de medir qué tan bien explica la variable independiente (X), a la variable dependiente (Y). Por lo tanto, es beneficioso calcular un número que indique el grado en que la línea de regresión MCO coincide con los datos de las variables. Si los puntos asociados a los datos están en la misma línea, los MCO proporcionan un ajuste perfecto. Esta medida de bondad de ajuste se mide a través del **coeficiente determinación** (R^2). Este indicador mide el porcentaje en que la variable dependiente está siendo explicada por la variable independiente. Su valor fluctúa entre 0 y 1.

- Si $R^2 = 1$, el ajuste es perfecto.
- Si $R^2 = 0$, no hay relación entre la variable dependiente y la variable explicativa.

Ahora lo revisaremos en el *software STATA*.

Figura 4.

The screenshot shows the Stata/SE 14.0 interface with the following details:

- Menu Bar:** Archivo, Edición, Datos, Gráficos, Estadísticas, Usuario, Ventana, Ayuda.
- Toolbar:** Includes icons for opening files, saving, and various statistical operations.
- Command Window:** Shows the command history and output. The commands run are:
 - # Comando _rc
 - 1 import excel "..."
 - 2 reg PCEY GDPX
 - 3 destring Year, replace
 - 4 tset Year, yearly
 - 5 reg PCEY GDPX
 - 6 gen RES, resid... 100
 - 7 predict RES, re...
 - 8 predict YEST, xb
 - 9 twoway scatter...

Output from command 2:

PCEY	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
GDPX	.7218343	.0044233	163.19	0.000	.7129197 .730749
_cons	-299.5913	28.76494	-10.42	0.000	-357.5632 -241.6194

Output from command 5:

Source	SS	df	MS	Number of obs	= 46
Model	144125684	1	144125684	F(1, 44)	= 26630.33
Residual	238131.858	44	5412.08768	Prob > F	= 0.0000
				R-squared	= 0.9984
				Adj R-squared	= 0.9983
Total	144363816	45	3208084.79	Root MSE	= 73.567

Output from command 6:

PCEY	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
GDPX	.7218343	.0044233	163.19	0.000	.7129197 .730749
_cons	-299.5913	28.76494	-10.42	0.000	-357.5632 -241.6194

Output from command 7:

. gen RES, residuals

Variables View: Shows variables PCEY, GDPX, RES, and YEST with their respective labels and types.

Properties View: Shows properties for variable GDPX, including Name (GDPX), Label (GDP(X)), Type (double), Format (%10.0g), and Value Labels.

File Path: C:\Users\lochoa\Documents

Su interpretación sería que el 99% de las variaciones en el GCP se explica por el cambio en el PIB. El ajuste es casi del 100% por lo que la variable PIB es adecuada para explicar al GCP.



Actividades de aprendizaje recomendadas

Nota. conteste las actividades en un cuaderno de apuntes o en un documento Word en caso de ser necesario.

- **Actividad 1**

Reflexione sobre los temas trabajados en la guía virtualizada y en el capítulo 3 del texto básico.

Procedimiento: Para que comprenda los temas desarrollados le aconsejo que realice cuadros sinópticos o resúmenes en los que tome las ideas principales con la finalidad de que tenga un documento de trabajo que posteriormente le permita revisar y comprender cada tema.

- **Actividad 2**

Estimado estudiante, para evaluar su aprendizaje puede proceder a realizar la autoevaluación de la unidad 3.

Procedimiento: En esta actividad debe contestar las preguntas al final de la unidad 3, esta tarea tiene la finalidad de comprobar que usted está teniendo sintonía con los temas tratados.

Recuerde que la autoevaluación no es obligatoria, sin embargo, le ayuda a reforzar sus conocimientos y lo prepara para la evaluación bimestral.

¡Vamos a trabajar!



Autoevaluación 3

Considerando el siguiente enunciado, seleccione la alternativa correcta.

1. El supuesto de homocedasticidad se refiere a:
 - a. Igual varianza.
 - b. Desigual varianza.
 - c. Linealidad en las variables.
2. El valor del coeficiente de determinación se encuentra entre:
 - a. -1 y 1
 - b. 1 y 3
 - c. 0 y 1
3. La metodología empleada para estimar los valores de B1 y B2 es:
 - a. Mínimos cuadrados ordinarios MCO
 - b. El teorema del límite central.
 - c. Cálculo de valores máximos.
4. El coeficiente de correlación (r) nos indica:
 - a. Cómo está relacionada X a Y.
 - b. La variación de los residuos con respecto al valor medio.
 - c. La forma funcional del modelo.
5. Diferente varianza significa:
 - a. Homocedasticidad.
 - b. Heteroscedasticidad.
 - c. Autocorrelación.
6. Una de las propiedades estadísticas de los estimadores de MCO es:
 - a. Que B1 siempre sea mayor a B2.
 - b. Que sean estimadores puntuales.
 - c. Que el coeficiente de determinación sea negativo.

7. ¿Cuántos supuestos plantea la teoría econométrica?
- a. 5
 - b. 10
 - c. 7
8. Uno de los supuestos del MCRL, manifiesta que los valores de X deben ser:
- a. Independientes del término de error
 - b. Dependientes del término de error.
 - c. Iguales al término de error.

[Ir al solucionario](#)



Unidad 4. Modelo de regresión lineal normal (MCRLN)

4.1. Distribución de probabilidad de las perturbaciones u_i

4.2. Supuesto de normalidad de u_i . ¿Por qué debe formularse el supuesto de normalidad?

4.3. Propiedades de los estimadores de MCO según el supuesto de normalidad

Si a los supuestos del modelo clásico de regresión lineal (MCRL), se le añade el supuesto de normalidad para u , tenemos el modelo clásico de regresión lineal normal (MCRLN). Para que un modelo sea considerado como adecuado, es necesario que sus residuos estén normalmente distribuidos, es decir que:

- Media: $E(u) = 0$
- Varianza: $E[u - E(u)]^2 = E(u^2) = \hat{\sigma}^2$
- Covarianza $E(u_i, u_j) = 0$

Es decir, están normal e independientemente distribuidas con media 0 y varianza $\hat{\sigma}^2$.

En base a los siguientes indicadores, y con ayuda del software STATA, podrá determinar si los residuos dentro del modelo están normalmente distribuidos.

- **Simetría:** Esta medida nos permite determinar si los datos se distribuyen de manera uniforme alrededor del valor medio o promedio.
- **Curtosis:** Determina el grado de concentración de los datos alrededor de la región central de la distribución. Al igual que la simetría, esta presenta también tres estados.

Ahora determinaremos la normalidad de los residuos, que es parte de uno de los supuestos que debe cumplir la regresión calculada a través de la metodología de MCO. Lo haremos por medio del STATA y lo primero que debemos hacer es obtener los residuos de la regresión, estos ya fueron calculados anteriormente usando el comando: **predict RES, residuals**.

Ahora utilizaremos una prueba de normalidad que el STATA nos provee, usando el siguiente comando: **sktest RES**.

Figura 5.

Comando: sktest RES

The screenshot shows the Stata/SE 14.0 interface with the following details:

- Menu Bar:** Archivo, Edición, Datos, Gráficos, Estadísticas, Usuario, Ventana, Ayuda.
- Toolbar:** Includes icons for opening files, saving, and other common functions.
- Code Editor:** Shows the command history:

```
1 import excel "..."  
2 reg PCEY GDPX  
3 desring Year, ...  
4 tset Year, yearly  
5 reg PCEY GDPX  
6 gen RES, resid... 100  
7 predict RES, re...  
8 predict YEST, xb  
9 twoway scatte...  
10 sktest RES
```
- Output Window:** Displays the results of the regression and the sktest command.

Regression Results:

	Model	1	Residual	F(1, 44)	= 26630.33
	144125684	144125684	238131.858	Prob > F	= 0.0000
			44	R-squared	= 0.9984
				Adj R-squared	= 0.9983
Total	144363816	45	3208084.79	Root MSE	= 73.567

Coeficients:

	PCEY	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
GDPX	.7218343	.0044233	163.19	0.000	.7129197	.730749
_cons	-299.5913	28.76494	-10.42	0.000	-357.5632	-241.6194

Skewness/Kurtosis tests for Normality:

Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	Prob>chi2
RES	46	0.1685	0.7187	2.14	0.3424
- Variables View:** Lists variables: Year, PCEY, GDPX, RES, YEST.
- Properties View:** Shows properties for variable RES, including Name: RES, Label: Residuals, Type: double, Format: %10.0g, and Observations: 46.

De los resultados obtenidos, vemos que utiliza la simetría y la curtosis para determinar si los residuos de la regresión están normalmente distribuidos. Para conocer si se cumple con el supuesto de normalidad, debemos hacerlos a través de hipótesis las mismas que son:

H₀: Los residuos de la regresión están normalmente distribuidos.

H₁: Los residuos de la regresión NO están normalmente distribuidos.

Para contrastar la H₀, debemos utilizar el p-valor del estadístico Chi², cuyo valor es de 0,34 o 34%. Como el valor de 34% es superior al 5% del nivel de significancia, entonces no podemos rechazar la H₀, y aceptaríamos que los residuos están normalmente distribuidos.

Las razones para formular el supuesto de normalidad se exponen en su libro básico en el capítulo 4, es necesario que las revise para un mejor entendimiento.

Recursos de aprendizaje

Los recursos que va a utilizar para este resultado de aprendizaje son:

Lectura

- Gujarati y Porter (2010). *Econometría*. México: McGraw Hill. Quinta edición.
El libro le proveerá de todo el marco teórico acerca del concepto de regresión y su aplicabilidad dentro de las ciencias sociales. Capítulo 4.



Video

- Cuellar L. (2019) [Evaluando normalidad en STATA](#).





Actividades de aprendizaje recomendadas

■ Actividad 1

Lea detenidamente los contenidos del capítulo 4 del texto básico y revise los ejemplos y ejercicios de la guía didáctica.

Procedimiento: Para una mejor comprensión de los temas desarrollados, es importante que revise varias fuentes de información, sobre los temas vistos y realice comparaciones en cuanto a conceptos, terminologías y procedimientos.

Nota. conteste las actividades en un cuaderno de apuntes o en un documento Word en caso de ser necesario.

■ Actividad 2

Estimado estudiante, para evaluar su aprendizaje puede realizar la autoevaluación de la unidad 4.

Procedimiento: En esta actividad debe contestar las preguntas al final de la unidad 4, esta tarea tiene la finalidad de comprobar que usted está teniendo sintonía con los temas tratados.

Recuerde que la autoevaluación no es obligatoria, sin embargo, le ayuda a reforzar sus conocimientos y lo prepara para la evaluación bimestral.

¡Vamos a trabajar!



Autoevaluación 4

Considerando el siguiente enunciado, seleccione la alternativa correcta.

1. La distribución normal es una distribución comparativa sencilla y requiere solo dos parámetros:
 - a. La media.
 - b. Hipótesis.
 - c. Coeficiente de determinación.
2. La suposición de normalidad permite utilizar pruebas o indicadores estadísticos:
 - a. La media y mediana.
 - b. t, F y X^2
 - c. Coeficiente de determinación.
3. ¿Cuál de las siguientes propiedades se da bajo el supuesto de normalidad?
 - a. Insesgamiento.
 - b. Varianza máxima.
 - c. Inconsistencia.
4. El método de máxima verosimilitud tiene propiedades:
 - a. Menos fuertes que los MCO.
 - b. Igual que los MCO.
 - c. Más fuertes que los MCO.
5. El supuesto de normalidad requiere que la varianza sea:
 - a. De cero.
 - b. Igual a 1.
 - c. Constante.

6. La justificación teórica del supuesto de normalidad es:
 - a. MCO.
 - b. Teorema del límite central.
 - c. Pruebas de hipótesis.
7. Sin el supuesto de normalidad el teorema de Gauss-Markov demostró de que los estimadores de MCO son:
 - a. Mejores estimadores lineales insesgados.
 - b. Mejores estimadores lineales sesgados.
 - c. Malos estimadores lineales y sesgados.
8. Una alternativa al método de MCO es:
 - a. Intervalos de confianza.
 - b. Sumatoria de los residuos.
 - c. Máxima verosimilitud.

[Ir al solucionario](#)

Resultado de aprendizaje 2 y 3

- Comprende la utilidad de un análisis de regresión lineal simple.
- Conocer como los cambios de una variable afectan a otra es importante a la hora de establecer un adecuado modelo de regresión.

Contenidos, recursos y actividades de aprendizaje



Semana 6

Unidad 5. Estimación por intervalos y pruebas de hipótesis

5.1. Requisitos estadísticos

5.2. Intervalos de confianza: ideas básicas

5.3. Intervalos de confianza para los coeficientes de regresión

5.4. Intervalo de confianza para la varianza

5.5. Pruebas de hipótesis: comentarios generales

A más de la utilización de los errores estándar y del coeficiente de determinación se incluyen el cálculo de los intervalos de confianza, tema que veremos a continuación.

Como parte de la verificación estadística que constituye el paso 7 de la metodología econométrica, a más de los errores estándar y del coeficiente de determinación se incluyen el cálculo de los intervalos de confianza, tema que veremos a continuación.

Como lo vimos anteriormente la confiabilidad de un estimador se mide por su error estándar, por lo tanto, para no depender de un solo indicador, se puede construir un intervalo que tenga la probabilidad de incluir dentro de sus límites el verdadero valor del parámetro.

Se trata de encontrar que tan cerca está $\hat{\beta}_1$ de $\hat{\beta}_2$, para esto buscamos dos números σ y α este último situado entre 0 y 1.

Para esto utilizamos la siguiente ecuación tanto para $\hat{\beta}_1$ y $\hat{\beta}_2$

$$\Pr(\hat{\beta}_2 - \delta \leq \beta_2 \leq \hat{\beta}_2 + \delta) = 1 - \alpha$$

La ecuación anterior muestra un estimador de intervalo, construido de manera que contenga dentro de sus límites el valor verdadero del parámetro. El estimador de intervalos proporciona entonces un recorrido de valores dentro de los cuales puede encontrarse el verdadero $\hat{\beta}_2$. Adicional a esto el cálculo de intervalos nos permiten poder aceptar o rechazar cualquier hipótesis que se haya planteado con respecto a los resultados del modelo.

Para esto vamos a utilizar cuatro métodos prácticos para aceptar o rechazar hipótesis:

- Intervalo de confianza
- Nivel de significancia
- Prueba t
- Prueba F

De los métodos antes mencionados los que vamos a utilizar son los que utiliza cualquiera de los estadísticos F, t o X^2 , al ser los que generalmente utiliza el software STATA. Lo que se busca es determinar si los coeficientes de regresión son significativos tanto a nivel individual como a nivel conjunto o global del modelo.

A nivel individual las hipótesis planteadas para poder utilizar los cuatro métodos son las siguientes:

$$H_0: \beta_1 = 0$$

No hay intercepto.

$$H_1: \beta_1 \neq 0$$

Si hay intercepto.

$$H_0: \beta_2 = 0$$

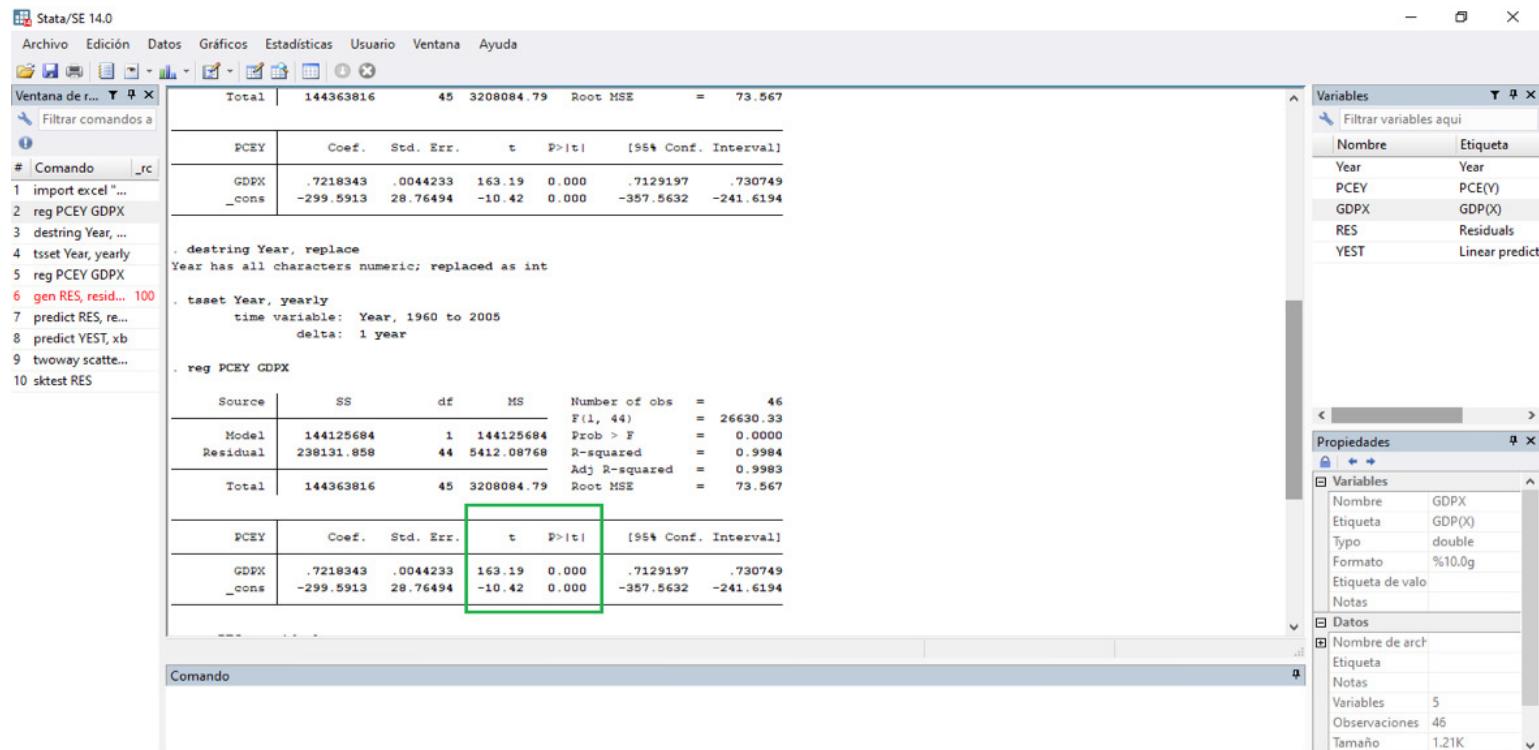
No hay relación lineal entre la variable X y la variable Y.

$$H_1: \beta_2 \neq 0$$

Si hay relación lineal entre la variable X y la variable Y.

Del ejercicio que estamos utilizando de ejemplo tenemos los siguientes resultados en STATA, los cuales ya fueron obtenidos en la estimación del modelo:

Figura 6.
Resultados en STATA



Para probar si los coeficientes son significativos o no, vamos a utilizar el valor p del estadístico t ($p>t$). La regla de decisión para poder contrastar la H_0 , es la siguiente:

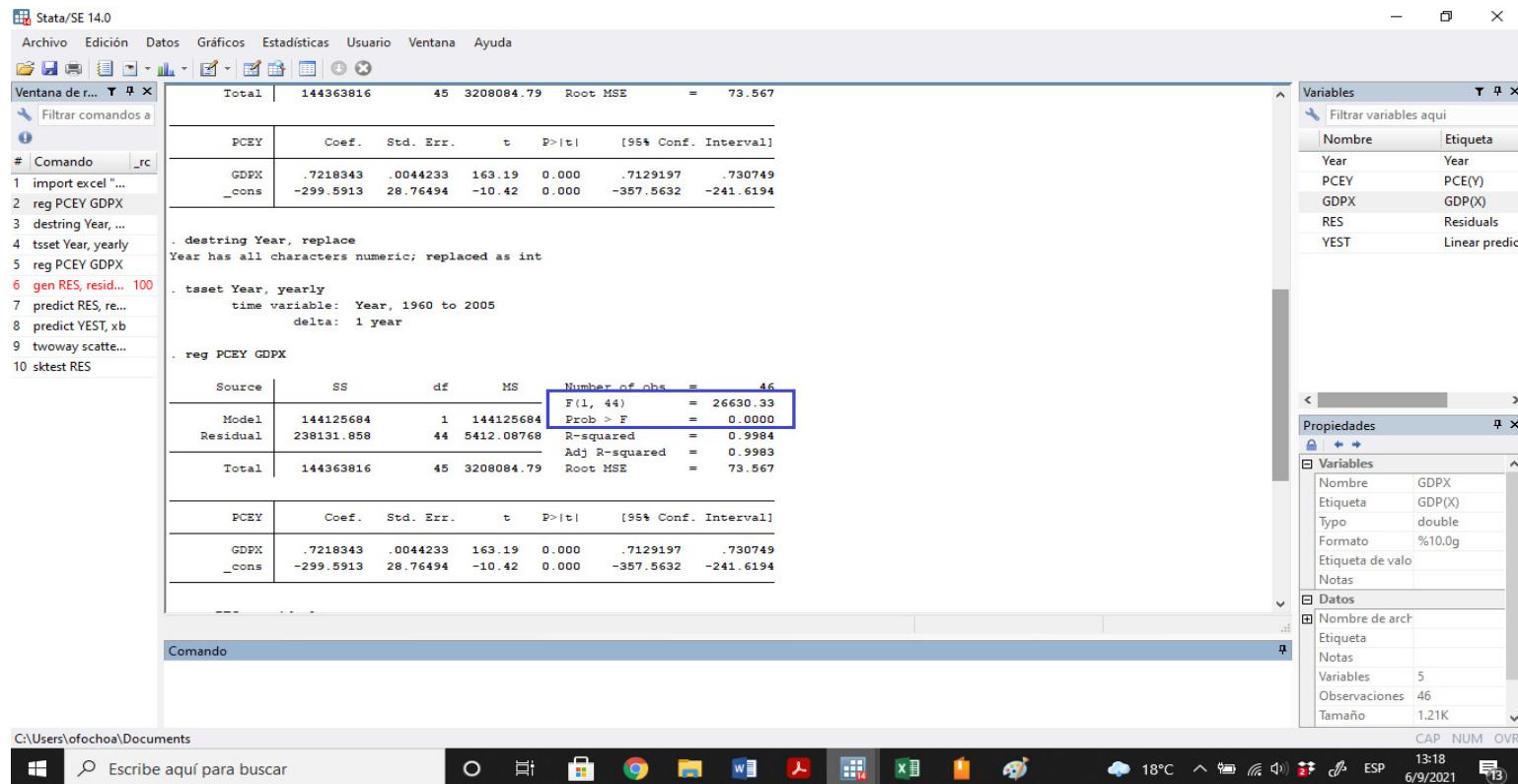
Si el p-valor del estadístico, es menor al nivel de significancia (5%), entonces podemos rechazar la H_0 , caso contrario no podemos rechazarla.

El p-valor de t para B1 y B2 son: de 0,0000 o del 0%, por ende, menores al 5%, lo que nos permite rechazar la H_0 y aceptar que los coeficientes individualmente son significativos.

El p-valor del estadístico F, en cambio nos permite conocer si el modelo en su conjunto es significativo, la regla de decisión es la misma que aplicamos anteriormente solo que ahora utilizamos el estadístico F y no el t.

Revise la figura del software STATA.

Figura 7.



Como podemos observar del rectángulo azul, el valor p del estadístico F, es de 0,00 o 0%, menor al 5%, por lo que rechazamos H₀ y aceptamos que el modelo en su conjunto es significativo. Para un mayor entendimiento del tema, revise el capítulo 5 de su texto base.

Lecturas

- Gujarati y Porter (2010). *Econometría*. México: McGraw Hill. Quinta edición.



El libro le proveerá de todo el marco teórico acerca del concepto de regresión y su aplicabilidad dentro de las ciencias sociales capítulo 5.

- Wooldridge (2009). *Introducción a la econometría: un enfoque moderno*. Cuarta edición. CENGAGE Learning.

Revise el capítulo 4 que le permitirá entender cómo funcionan los intervalos de confianza para modelos de regresión múltiple.



Actividades de aprendizaje recomendadas

- **Actividad 1**

Lea detenidamente los contenidos del capítulo 5 del texto básico y revise los ejemplos y ejercicios de la guía didáctica.

Procedimiento: Para una mejor comprensión de los temas desarrollados, es importante que revise varias fuentes de información, sobre los temas vistos y realice comparaciones en cuanto a conceptos, terminologías y procedimientos. Utilice las técnicas que de acuerdo con su estilo de aprendizaje le sean de mayor utilidad.

Nota. conteste las actividades en un cuaderno de apuntes o en un documento Word en caso de ser necesario.



5.6. Pruebas de hipótesis: método del intervalo de confianza

5.7. Pruebas de hipótesis: enfoque de la prueba de significancia

5.8. Pruebas de hipótesis: algunos aspectos básicos

5.9. Análisis de regresión y de varianza

5.10. Aplicación del análisis de regresión: problema de predicción

Intervalo de Confianza: esta prueba inicialmente nos permite obtener el intervalo dentro del cual se encuentra el verdadero valor de los parámetros $\hat{\beta}_1$ y $\hat{\beta}_2$, adicional a esto nos permite aceptar o rechazar las hipótesis planteadas para cada parámetro de manera individual.

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

La interpretación de este intervalo sería que, en 95 de cada 100 casos, intervalos como este contendrán el verdadero valor de $\hat{\beta}_1$.

$$H_0: \beta_2 = 0$$

$$H_1: \beta_2 \neq 0$$

La interpretación de este intervalo es que, en 95 de cada 100 casos, intervalos como este contendrán el verdadero valor de $\hat{\beta}_2$.

Nivel de significancia: Es similar a la prueba de intervalo de confianza, con la diferencia que $\hat{\beta}_1$ y $\hat{\beta}_2$ toman el valor de cero dentro de la fórmula y se los identifica ahora como β^*_1 y β^*_2 .

$$H_0: \beta^*_1 = 0 \quad H_0: \beta^*_2 = 0$$

$$H_1: \beta^*_1 \neq 0 \quad H_1: \beta^*_2 \neq 0$$

Para reforzar los temas finales revise estos temas que corresponde al capítulo 5.

Lecturas

- Gujarati y Porter (2010). *Econometría*. México: McGraw Hill. Quinta edición.

El libro le proveerá de todo el marco teórico acerca del concepto de regresión y su aplicabilidad dentro de las ciencias sociales. Capítulo 5.



- Wooldridge (2009). *Introducción a la econometría: un enfoque moderno*. Cuarta edición. CENGAGE Learning.

Revise el capítulo 4 que le permitirá entender cómo funcionan los intervalos de confianza para modelos de regresión múltiple.



Actividades de aprendizaje recomendadas

▪ Actividad 1

Lea detenidamente los contenidos del capítulo 5 del texto básico y revise los ejemplos y ejercicios de la guía didáctica.

Procedimiento: Para una mejor comprensión de los temas desarrollados, es importante que revise varias fuentes de información, sobre los temas vistos y realice comparaciones en cuanto a conceptos, terminologías y procedimientos. Utilice las técnicas que de acuerdo con su estilo de aprendizaje le sean de mayor utilidad.

Nota. conteste las actividades en un cuaderno de apuntes o en un documento Word en caso de ser necesario.

- **Actividad 2**

Estimado estudiante, para evaluar su aprendizaje puede realizar la autoevaluación de la unidad 5.

Procedimiento: En esta actividad debe contestar las preguntas al final de la unidad 5, esta tarea tiene la finalidad de comprobar que usted está teniendo sintonía con los temas tratados.

Recuerde que la autoevaluación no es obligatoria, sin embargo, le ayuda a reforzar sus conocimientos y lo prepara para la evaluación bimestral.

¡Vamos a trabajar!



Autoevaluación 5

Considerando el siguiente enunciado, seleccione la alternativa correcta.

1. Si la hipótesis alternativa está planteada de la siguiente manera: $H_1: B_2=1$, significa que esta es:
 - a. Compuesta.
 - b. Simple.
 - c. Ninguna de las anteriores.
2. Un estimador de intervalo tiene una probabilidad específica de:
 - a. $(0.5 - 0.25)$
 - b. $(1 - 0.75)$
 - c. $(1 - \alpha)$
3. Una hipótesis alternativa compuesta se la conoce como:
 - a. Hipótesis bilateral.
 - b. Hipótesis nula.
 - c. Hipótesis simple.
4. Error de tipo 1 es:
 - a. Aceptar una hipótesis verdadera.
 - b. Aceptar una hipótesis falsa.
 - c. Rechazar una hipótesis verdadera.
5. Probar hipótesis de manera conjunta se lo hace a través de:
 - a. La prueba t.
 - b. La prueba χ^2 .
 - c. La prueba F.
6. Una prueba para probar hipótesis es:
 - a. Intervalos de confianza.
 - b. Correlación.
 - c. Mínimos cuadrados ordinarios MCO.

7. La hipótesis nula se prueba frente a:
 - a. Al coeficiente de determinación.
 - b. A la hipótesis alternativa.
 - c. Los errores estándar.

8. En un estadístico, el valor calculado debe ser mayor que el valor crítico o de la tabla para poder:
 - a. Rechazar H_0 .
 - b. Aceptar H_0 .
 - c. Establecer un intervalo de confianza.

[Ir al solucionario](#)



Semana 8



Actividades finales del bimestre

Revisión de las unidades 1 a la 5

Debe revisar nuevamente los contenidos de los capítulos 1 al 5, para un reforzamiento de los contenidos y como preparación previa al examen presencial.

Lectura

- Gujarati y Porter (2010). *Econometría*. México: McGraw Hill. Quinta edición.



El libro le proveerá de todo el marco teórico acerca del concepto de regresión y su aplicabilidad dentro de las ciencias sociales.



Actividades de aprendizaje recomendadas

▪ Actividad 1

Lea los contenidos del texto básico en su capítulo del 1 al 5 y las unidades 1 al 5 de la guía didáctica adicional a las explicaciones sobre los temas desarrollados.

Procedimiento: Para el desarrollo de esta actividad es importante que realice una lectura comprensiva de todos los temas abordados y que vaya trabajando mediante cuadros sinópticos o resúmenes para que luego pueda revisarlos con mayor facilidad. Recuerde que lo importante es que usted quede lo suficientemente claro con todas las temáticas.

Nota. conteste las actividades en un cuaderno de apuntes o en un documento Word en caso de ser necesario.

- **Actividad 2**

Desarrolle las actividades recomendadas.

Procedimiento: Es importante que aun no siendo obligatoria su presentación, usted desarrolle las actividades recomendadas ya que de esa manera puede desarrollar las habilidades y destrezas en la aplicación de los contenidos en los aspectos de la vida práctica.



Segundo bimestre

Resultado de aprendizaje 2 ▪ Comprende la utilidad de un análisis de regresión lineal simple.

Contenidos, recursos y actividades de aprendizaje



Semana 9

Unidad 6. Extensiones del modelo de regresión lineal con dos variables

6.1. Regresión a través del origen, r^2 para el modelo de regresión a través del origen

6.2. Formas funcionales de los modelos de regresión

6.3. Cómo medir la elasticidad: modelo log-lineal

6.4. Modelos semilogarítmicos: log-lin y lin-log. El modelo lin-log

6.5. Modelos recíprocos

6.6. Elección de la forma funcional

Luego de haber revisado lo concerniente a un modelo de regresión simple, en los cuales se trabajó con modelos lineales, es decir, que tanto sus variables como parámetros estaban expresadas de manera lineal, es muy poco probable que la teoría económica siempre se rija a este tipo de modelos.

En esta unidad vamos a revisar que muchos fenómenos económicos no pueden ser expresados a través de modelos lineales, sino que la misma determinación de información propondrá que estos modelos sean planteados de diversas formas. Esto es lo que se conoce como formas funcionales de los modelos, en los cuales las variables dejan de ser lineales, pero no los parámetros.

Regresión a través del origen: Wooldridge (2001) la denomina regresión a través del origen, porque la línea pasa por el origen $x=0$ y $\bar{y}=0$.

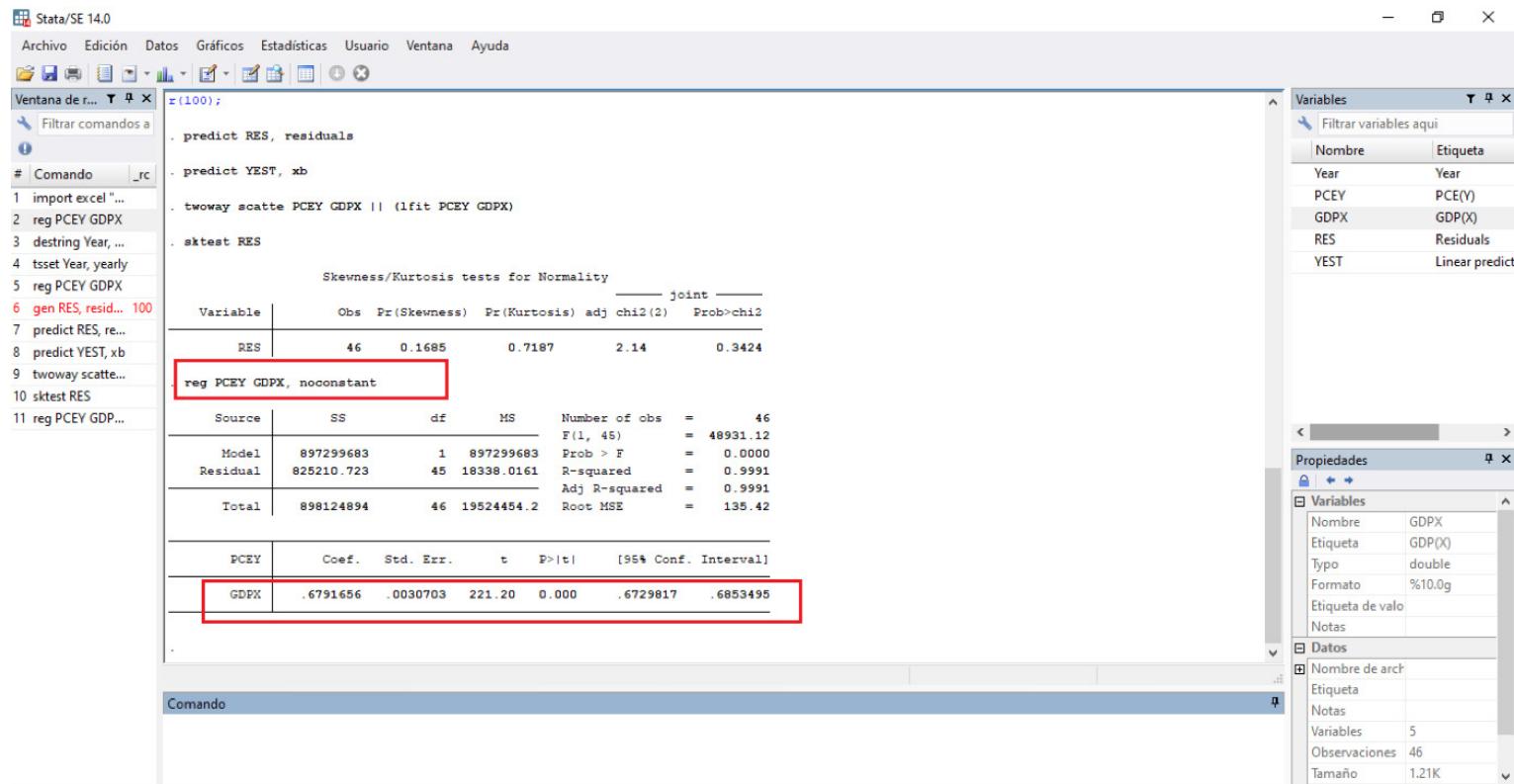
La ecuación para este tipo de regresión es:

$$Y_i = \hat{\beta}_2 \hat{X}_i + \hat{u}_i$$

Esta ecuación carece del intercepto o esta toma el valor de cero, es decir el $\hat{\beta}_1$.

Este tipo de modelo presenta ciertas desventajas como la suma de residuos no siempre es cero, y el coeficiente de determinación puede ser negativo. El uso de este modelo solo sería indispensable si existe expectativa a priori muy fuerte que justifique su uso. Algunas de las razones que justifican que el intercepto esté ausente del modelo, es que la teoría económica requiere o exige esto, casos como la elasticidad-oferta unitaria, puede llevarnos a excluirlo. Para el ejercicio práctico vamos a seguir con la tabla I.1 del texto básico. La variable Y en esta tabla es el gasto de consumo personal (GCP) agregado (para la economía en su conjunto), y la variable X, el producto interno bruto (PIB), una medida del ingreso agregado, ambos medidos en miles de millones de dólares de 2000. Lo que vamos a realizar es una regresión a través del origen, es decir, sin intercepto, el comando: **reg PCEY GDPX, noconstant**.

Figura 8.



Como podemos observar se obtiene una regresión a través del origen en la cual no consta el intercepto o constante.

Modelo Log – lineal o doble log o Log – Log: conocido como modelo de regresión exponencial, el mismo que se los puede linealizar a través de logaritmos naturales. A estos modelos también se los conoce como modelos de elasticidad constante. La ecuación de regresión es la siguiente:

$$y = \beta_1 X^{\beta_2} e^u \longrightarrow \ln y = \ln \beta_1 + \beta_2 \ln x + u$$

- Lineal en los parámetros.
- Lineal en los logaritmos de las variables Y y X, debido a estos se denominan modelos log-log; doble log o log-lineales.
- El modelo log-log es importante, ya que el coeficiente de la pendiente B2 mide la elasticidad de Y con respecto a X, es decir, el cambio porcentual en Y ante un pequeño cambio porcentual en X.

En STATA si queremos trabajar con un modelo doble LOG, debemos utilizar los siguientes comandos: **modelo doble LOG**.

1. Debemos transformar las variables a logaritmo, para esto debemos generar nuevas variables: **gen LPCEY= log(PCEY)** y **gen LGDPX= log(GDPX)**, con esto generamos las variables Y y X, pero en logaritmos.

Figura 9.
Modelo doble LOG

Stata/SE 14.0

Archivo Edición Datos Gráficos Estadísticas Usuario Ventana Ayuda

Ventana de r... T X

```

. predict YEST, xb
. twoway scatter PCEY GDPX || (lfit PCEY GDPX)
. sktest RES
      Skewness/Kurtosis tests for Normality
      joint
      Variable Obs Pr(Skewness) Pr(Kurtosis) adj chi2(2) Prob>chi2
      RES 46 0.1685 0.7187 2.14 0.3424
. reg PCEY GDPX, noconstant
      Source SS df MS Number of obs = 46
      Model 897299683 1 897299683 Prob > F = 0.0000
      Residual 825210.723 45 18338.0161 R-squared = 0.9991
      Total 898124894 46 19524454.2 Adj R-squared = 0.9991
      Root MSE = 135.42
. predict RES, re...
. gen RES, resid...
. predict YEST, xb
. twoway scatter...
. sktest RES
. reg PCEY GDP...
. gen LPCEY=lo...
. gen LGDPX=lo...
. gen LPCEY=log(PCEY)
. gen LGDPX=log(GDPX)

```

Variables

Nombre	Etiqueta
Year	Year
PCEY	PCE(Y)
GDPX	GDP(X)
RES	Residuals
YEST	Linear predict
LPCEY	
LGDPX	

Propiedades

Variables
Nombre: GDPX
Etiqueta: GDP(X)
Tipo: double
Formato: %10.0g
Etiqueta de valor:
Notas:

Datos
Nombre de archivo:
Etiqueta:
Notas:
Variables: 7
Observaciones: 46
Tamaño: 1.57K

Comando

2. Luego volvemos a realizar la regresión con estas nuevas variables: **reg LPCEY LGDPX**

Figura 10.

Regresión con variables: *reg LPCEY LGDPX*

The screenshot shows the Stata/SE 14.0 interface with the following details:

- Variables pane:** Shows the variable definitions:
 - Year
 - PCEY
 - GDPX
 - RES
 - YEST
 - LPCEY
 - LGDPX** (selected)
- Properties pane:** Shows the properties for the selected variable LGDPX:
 - Nombre: LGDPX
 - Etiqueta:
 - Tipo: float
 - Formato: %9.0g
 - Etiqueta de valor:
 - Notas:
- Log Output:** Displays the command history and log output for the regression command.

```
1 import excel "..."  
2 reg PCEY GDPX  
3 desiring Year, ...  
4 tset Year, yearly  
5 reg PCEY GDPX  
6 gen RES, resid... 100  
7 predict RES, re...  
8 predict YEST, xb  
9 twoway scatter...  
10 sktest RES  
11 reg PCEY GDP...  
12 gen LPCEY=lo...  
13 gen LGDPX=lo...  
14 reg LPCEY LG...  
  
reg LPCEY LGDPX
```

Regression Results:

Source	SS	df	MS	Number of obs	=	46
Model	9.43853406	1	9.43853406	F(1, 44)	=	48481.89
Residual	.008565992	44	.000194682	Prob > F	=	0.0000
Total	9.44710005	45	.209935557	R-squared	=	0.9991
				Adj R-squared	=	0.9991
				Root MSE	=	.01395

LPCEY	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
LGDPX	1.073644	.0048761	220.19	0.000	1.063817 1.083471
_cons	-1.045081	.042068	-24.84	0.000	-1.129863 -.9602981
- Command pane:** Shows the last command entered: `reg LPCEY LGDPX`.

Lo que diferencia a un modelo de regresión lineal simple, es que la interpretación de los coeficientes va a cambiar. En el caso de B1, se dice que si la variable X=0, la variable LGPCY disminuye en promedio en un 1.04%, y en cuanto al B2, si el LGDPX aumenta en 1%, el LGPCY aumenta en promedio en un 1.07%.

Modelos semilogarítmicos: en este tipo de modelos tenemos dos variaciones.

Modelo Log – Lin

$$\ln Y_t = \beta_1 + \beta_2 t + u_t$$

Generalmente utilizado para medir tasas de crecimiento instantánea o compuesta. Este modelo es logarítmico en la variable dependiente, pero lineal en la variable independiente que además está dada por el tiempo. En este modelo el coeficiente de la pendiente mide el cambio proporcional constante o relativo en (Y), para un cambio absoluto dado en el valor de (X). Al multiplicar por 100 ($\hat{\beta}_2$), da como resultado la tasa de crecimiento en (Y).

Figura 11.
Modelo Log_Lin

Stata/SE 14.0

Archivo Edición Datos Gráficos Estadísticas Usuario Ventana Ayuda

Ventana de r... T X

Filtrar comandos a

Comando _rc

```

1 import excel "... "
2 reg PCEY GDPX
3 descript Year, ...
4 ttest Year, yearly
5 reg PCEY GDPX
6 gen RES, resid..., 100
7 predict RES, re...
8 predict YEST, xb
9 twoway scatter...
10 sktest RES
11 reg PCEY GDP...
12 gen LPCEY=lo...
13 gen LGDPX=lo...
14 reg LPCEY LG...
15 reg LPCEY GD...

```

Variables

Nombre	Etiqueta
Year	Year
PCEY	PCE(Y)
GDPX	GDP(X)
RES	Residuals
YEST	Linear predict
LPCEY	
LGDPX	

Propiedades

Variables

Nombre	Etiqueta
GDPX	GDP(X)
Etiqueta	
Tipo	double
Formato	%10.0g
Etiqueta de valor	
Notas	

Datos

Nombre de archivo	Etiqueta
Notas	
Variables	7
Observaciones	46
Tamaño	1,57K

Comando

C:\Users\ofochoa\Documents

Output for reg LPCEY GDPX:

Source	SS	df	MS	Number of obs	=	46
Model	9.43853406	1	9.43853406	F(1, 44)	=	48481.89
Residual	.008565992	44	.000194682	Prob > F	=	0.0000
Total	9.44710005	45	.209935557	R-squared	=	0.9991
				Adj R-squared	=	0.9991
				Root MSE	=	.01395

LPCEY	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
GDPX	.1.073644	.0048761	220.19	0.000	1.063817 1.083471
_cons	-1.045081	.042068	-24.84	0.000	-1.129863 -.9602981

Source	SS	df	MS	Number of obs	=	46
Model	9.09928187	1	9.09928187	F(1, 44)	=	1151.09
Residual	.347818176	44	.007904959	Prob > F	=	0.0000
Total	9.44710005	45	.209935557	R-squared	=	0.9632
				Adj R-squared	=	0.9623
				Root MSE	=	.08891

LPCEY	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
GDPX	.0001814	5.35e-06	33.93	0.000	.0001706 .0001921
_cons	7.114236	.0347641	204.64	0.000	7.044174 7.184299

Modelo Lin – Log

$$Y_i = \beta_1 + \beta_2 \ln X_i + u_i$$

A diferencia del modelo anterior, lo que se trata de conocer ahora es el cambio absoluto en Y, debido a un cambio porcentual en X. Esta ecuación plantea que el cambio absoluto en (Y), es igual a B2 veces el cambio relativo en (X).

Así, si X cambia en 0,01 unidades o 1%, el cambio absoluto en Y es $(0,01) * B2$ o lo mismo si divido $100/B2$.

Figura 12.
Modelo Lin-Log

Stata/SE 14.0

Archivo Edición Datos Gráficos Estadísticas Usuario Ventana Ayuda

Ventana de r... T X

Filtrar comandos a

Comando _rc

```

1 import excel "... "
2 reg PCEY GDPX
3 descript Year, yearly
4 tset Year, yearly
5 reg PCEY GDPX
6 gen RES, resid..., 100
7 predict RES, re...
8 predict YEST, xb
9 twoway scatte...
10 sktest RES
11 reg PCEY GDP...
12 gen LPCEY=lo...
13 gen LGDPX=lo...
14 reg LPCEY LG...
15 reg LPCEY GD...
16 reg PCEY LGD...

```

Variables

Nombre	Etiqueta
Year	Year
PCEY	PCE(Y)
GDPX	GDP(X)
RES	Residuals
YEST	Linear predict
LPCEY	
LGDPX	

Propiedades

Variables

Nombre	LGDPX
Etiqueta	
Tipo	float
Formato	%9.0g
Etiqueta de valor	
Notas	

Datos

Nombre de archivo	
Etiqueta	
Notas	
Variables	7
Observaciones	46
Tamaño	1.57K

Comando

Source SS df MS Number of obs = 46
F(1, 44) = 1151.09
Prob > F = 0.0000
R-squared = 0.9632
Adj R-squared = 0.9623
Root MSE = .08691

LPCEY Coef. Std. Err. t P>|t| [95% Conf. Interval]
GDPX .0001814 5.35e-06 33.93 0.000 .0001706 .0001921
_cons 7.114236 .0347641 204.64 0.000 7.044174 7.184299

reg PCEY LGDPX

Source SS df MS Number of obs = 46
F(1, 44) = 878.05
Prob > F = 0.0000
R-squared = 0.9523
Adj R-squared = 0.9512
Root MSE = 395.69

PCEY Coef. Std. Err. t P>|t| [95% Conf. Interval]
LGDPX 4097.505 138.2804 29.63 0.000 3818.82 4376.191
_cons -31260.69 1193.005 -26.20 0.000 -33665.03 -28856.35

Para un mayor entendimiento de este tema le invito a revisar el capítulo 6 de su texto básico, para que revise como son las interpretaciones de este tipo de modelos.

Modelo recíproco: este modelo es generalmente utilizado cuando al graficar X y Y, el gráfico determina una hipérbola. Adicionalmente este modelo está relacionado con la curva de Engel.

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

En STATA se procede de la siguiente forma: se debe generar una variable que cumpla la condición $1/x$, el comando para esto sería: **gen RGDPX=(1/GDPX)**, el nombre que le dé a la nueva variable depende de Usted. Se vuelve a realizar la regresión con esta nueva variable.

Figura 13.

Comando_gen RGDPX

The screenshot shows the Stata/SE 14.0 interface with the following details:

- Menu Bar:** Archivo, Edición, Datos, Gráficos, Estadísticas, Usuario, Ventana, Ayuda.
- Toolbar:** Includes icons for opening files, saving, and other common functions.
- Command History:** Shows the command sequence used to generate RGDPX. The last two commands are highlighted with red boxes:
 - `gen RGDPX=(1/GDPX)`
 - `reg PCEY RGDPX`
- Output Window:** Displays the regression results for the second command. The output includes:

PCEY	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
RGDPX	-1.91e+07	1278547	-14.95	0.000	-2.17e+07 -1.65e+07
_cons	7830.464	275.1617	28.46	0.000	7275.912 8385.016
- Variables View:** A list of variables with their names and labels. The variable `RGDPX` is highlighted with a red box in the list.
- Properties View:** A detailed view of the `RGDPX` variable properties, including Name, Label, Type (float), Format (%9.0g), and Value Labels.
- File Path:** C:\Users\lofchoa\Documents

La interpretación de B2 para este tipo de modelo se lo dejo como consulta.

Lectura

- Gujarati y Porter (2010). *Econometría*. México: McGraw Hill. Quinta edición.



El libro le proveerá de todo el marco teórico acerca del concepto de regresión y su aplicabilidad dentro de las ciencias sociales. Capítulo 6.

Video



- Revisa el siguiente video acerca de los [modelos logarítmicos y como realizarlos STATA](#).



Actividades de aprendizaje recomendadas

▪ Actividad 1

Lea detenidamente los contenidos del capítulo 6 del texto básico y revise los ejemplos y ejercicios de la guía didáctica.

Procedimiento: Para una mejor comprensión de los temas desarrollados, es importante que revise varias fuentes de información sobre los temas vistos y realice comparaciones en cuanto a conceptos, terminologías y procedimientos. Utilice las técnicas que de acuerdo a su estilo de aprendizaje le sean de mayor utilidad.

Nota: conteste *las actividades en un cuaderno de apuntes o en un documento Word en caso de ser necesario*.

▪ Actividad 2

Estimado estudiante, para evaluar su aprendizaje puede proceder a realizar la autoevaluación de la unidad 6.

Procedimiento: En esta actividad debe contestar las preguntas al final de la unidad 6, esta tarea tiene la finalidad de comprobar que usted está teniendo sintonía con los temas tratados.

Recuerde que la autoevaluación no es obligatoria, sin embargo, le ayuda a reforzar sus conocimientos y lo prepara para la evaluación bimestral.

¡Vamos a trabajar!



Autoevaluación 6

Considerando el siguiente enunciado, seleccione la alternativa correcta.

1. Una propiedad de la variable estandarizada es:
 - a. Suma de cuadrados ajustadas.
 - b. Suma de cuadrados estandarizados.
 - c. Suma de cuadrados simples.
2. Con regresión a través del origen el coeficiente de determinación puede ser:
 - a. Negativo.
 - b. Igual a cero.
 - c. Igual a uno.
3. Ajustar una regresión a través del origen implica:
 - a. Pérdida de datos.
 - b. Error de especificación.
 - c. Modelo perfecto.
4. La transformación de escala de las variables:
 - a. No afecta a las propiedades de los MCO.
 - b. Afecta a las propiedades de los MCO.
 - c. Genera modelos logarítmicos.
5. Una propiedad de la variable estandarizada es:
 - a. Media = 1.
 - b. Media = 0.
 - c. Desviación estándar = 0.
6. Modelo log - lineal es:
 - a. $\log Y = B_0 + B_1 \log X + u_i$.
 - b. $Y = B_0 + B_1 \log X + u_i$.
 - c. $\log Y = B_0 + B_1 X + u_i$.

7. Las elasticidades pueden ser medida por un:
 - a. Modelo semi logarítmico.
 - b. Modelo doble log.
 - c. Modelo recíproco.

8. ¿Qué modelo mide la tasa de crecimiento?
 - a. Modelo semilogarítmico.
 - b. Modelo log – log.
 - c. Modelo log – lin.

[Ir al solucionario](#)

Resultado de aprendizaje 4

- Analiza la relación lineal entre una variable dependiente en función de otras denominadas independientes.

Contenidos, recursos y actividades de aprendizaje



Semana 10

Unidad 7. Análisis de regresión múltiple: el problema de estimación

7.1. Modelo con tres variables: notación y supuestos

7.2. Interpretación de la ecuación de regresión múltiple

7.3. Significado de los coeficientes de regresión parcial

7.4. Estimación de MCO y MV de los coeficientes de regresión parcial

7.5. R^2 y R^2 ajustada

No es muy común que dentro del análisis econométrico se dé modelos de regresión simple, por lo tanto, es necesario ampliar este modelo a modelos de regresión con varias variables explicativas.

El modelo de regresión con tres variables se puede expresar así:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

Al incluirse más de dos variables explicativas se debe trabajar con la suposición de que no existe relación lineal exacta entre X_2, X_3 .

Igual que el modelo de regresión lineal se sigue trabajando bajo los supuestos del modelo clásico de regresión (MCRL). Adicionalmente partimos de una regresión muestral.

Todas las fórmulas aquí utilizadas son tomadas de su texto guía.

Significado de los coeficientes de regresión parcial

$$Y = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{u}$$

Ahora en el estudio de los modelos de regresión múltiple tenemos varias variables que intervienen en los modelos, por tal razón un análisis de la relación que existe entre estas variables es importante para que el modelo pueda presentar resultados con mayor confiabilidad. El coeficiente de correlación parcial es aquel que mide la correlación neta entre la variable dependiente y una variable explicativa manteniendo.

La interpretación general de los coeficientes de regresión parcial son las siguientes:

β_1 = Es el término del intercepto, es el valor medio de Y cuando X_2 y X_3 , son iguales a cero.

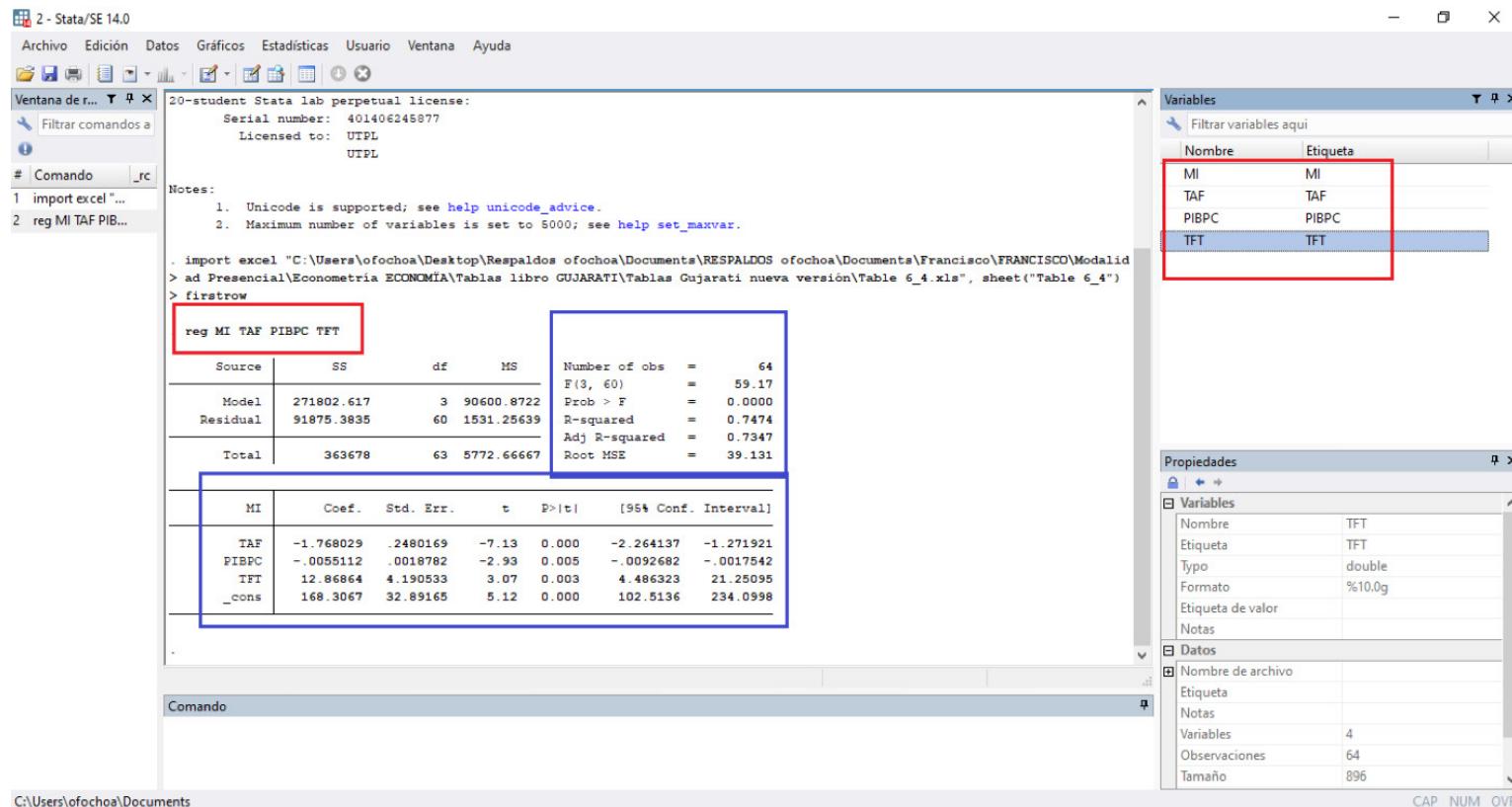
β_2 = Mide las variaciones de Y ante el incremento de 1 unidad en X_2 , manteniendo constante X_3 .

β_3 = Mide las variaciones de Y ante el incremento de 1 unidad en X_3 , manteniendo constante X_2 .

Vamos a trabajar con la tabla 6.4 de su texto básico en STATA para realizar un modelo de regresión múltiple.

Figura 14.

Modelo de regresión múltiple



Regresión múltiple, con base en la teoría realice las interpretaciones de los coeficientes de regresión parcial.

Lecturas

- Gujarati y Porter (2010). *Econometría*. México: McGraw Hill. Quinta edición.



El libro le proveerá de todo el marco teórico acerca del concepto de regresión y su aplicabilidad dentro de las ciencias sociales. Capítulo 7.

- Wooldridge (2009). Introducción a la econometría: un enfoque moderno. Cuarta edición. CENGAGE Learning.

Revise el capítulo 3 y 4 que le permitirá entender cómo funcionan los modelos de regresión múltiple.

Video



- Revisa el siguiente video para comprender como funcionan los [modelos de regresión múltiple](#).



Actividades de aprendizaje recomendadas

▪ Actividad 1

Lea detenidamente los contenidos del capítulo 7 del texto básico y revise los ejemplos y ejercicios de la guía virtualizada.

Procedimiento: Para una mejor comprensión de los temas desarrollados, es importante que revise varias fuentes de información, sobre los temas vistos y realice comparaciones en cuanto a conceptos, terminologías y procedimientos.

Nota: conteste *las actividades en un cuaderno de apuntes o en un documento Word en caso de ser necesario*.



7.6. La función de producción Cobb-Douglas: más sobre la forma funcional

7.7. Modelos de regresión polinomial

7.8. Coeficientes de correlación parcial

La función de producción de Cobb-Douglas

Es una extensión del modelo log-lineal o doble log, en el cual se parte de una función exponencial a una logarítmica.

$$y = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} + e^{ui} \longrightarrow \ln y_i = \ln \beta_1 + \beta_2 \ln x_{2i} + \beta_3 \ln x_{3i} + u_i$$

La función estudiada a continuación nos permite determinar elasticidades a través de sus coeficientes de regresión parcial, adicionalmente nos permite el cálculo de los rendimientos a escala dentro del modelo de producción. Para entender mejor este tema realicemos el siguiente ejercicio.

Se presenta información de la referente al sector agrícola de Taiwán durante 1958-1972. La tabla 7.3 fue tomada de su texto básico.

Y= Producto bruto real (P)

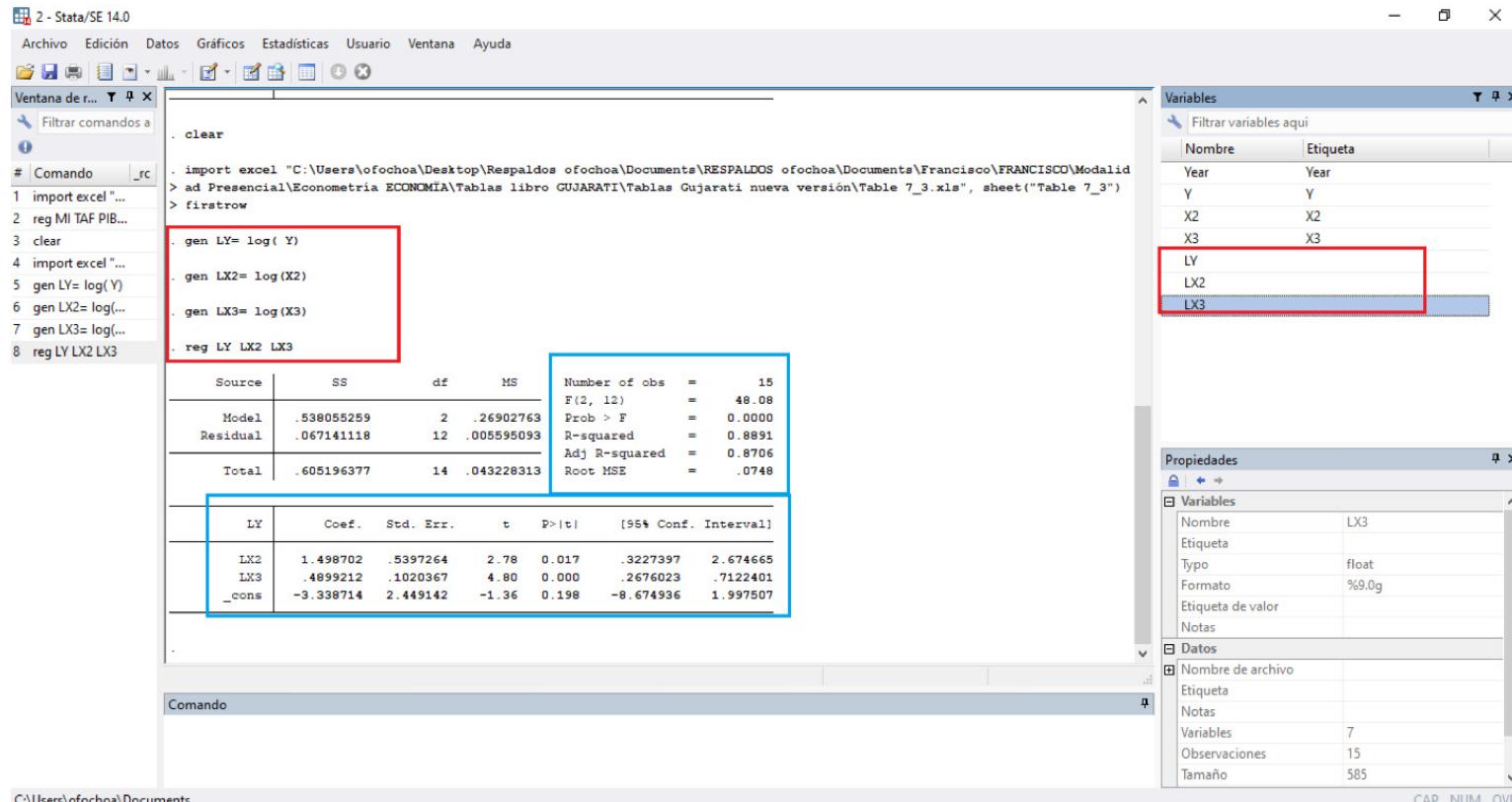
X2= Días laborales (L)

X3= Insumo capital real (K)

Ya que el *modelo de Cobb-Douglas* es netamente logarítmico, debemos generar nuevas variables para posterior realizar la regresión.

Figura 15.

Modelo de Cobb-Douglas



Interpretando los coeficientes tendríamos:

- $\hat{\beta}_1 = -3,33$

Cuando los días laborales y el insumo capital real son de cero, el producto bruto real es de -3,33 millones de nuevos dólares taiwaneses.

- $\hat{\beta}_2 = 1,49$

Significa que ante un incremento del 1% en el insumo trabajo condujo en promedio a un incremento de cerca del 1,5% en el producto, manteniendo constante el insumo capital. Adicionalmente este valor me da la elasticidad del trabajo con respecto al producto o elasticidad producto-trabajo.

- $\hat{\beta}_3 = 0,49$

Significa que ante un incremento del 1% en el insumo capital condujo en promedio a un incremento de cerca del 0,5% en el producto, manteniendo constante el insumo trabajo.

Adicionalmente este valor me da la elasticidad del capital con respecto al producto o la elasticidad producto-capital. La sumatoria de las dos elasticidades del producto me permite determinar los rendimientos a escala. Si este valor es mayor a uno los rendimientos son crecientes, si es menor a uno son decrecientes y si es igual a uno estos son constantes. En nuestro ejercicio la sumatoria de las dos elasticidades nos da 1,9887, lo que nos muestra que el sector agrícola en Taiwán desde 1958 hasta 1972 ha tenido rendimientos crecientes.

Modelos polinomiales: estos modelos son generalmente utilizados en la investigación econométrica en las funciones de costo y de producción.

Se plantean desde una regresión simple y al convertirse en un modelo polinomial se transforma en una regresión múltiple. Al dibujar en un plano cartesiano las variables (Y y X), obtenemos un gráfico, si este es una parábola nos demuestra que el modelo de regresión para estas variables es un modelo polinomial y su ecuación sería: $Y = ax^2 + bx + c$

El modelo econométrico planteado sería el siguiente:

$$Y = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u$$

Este es un polinomio de segundo grado, ya que X está elevado al cuadrado, si X estuviera elevado al cubo, este polinomio sería ahora de tercer grado. Tenga en cuenta que, en este tipo de regresiones polinomiales, solamente hay una variable explicativa al lado derecho, pero aparece elevada a distintas potencias.

Prueba de igualdad de los dos coeficientes de regresión: cabe resaltar que, en muchas investigaciones económicas se presentan variables que pueden tener similitud, tal es el caso de la función de demanda de un bien, el cual plantea que Y = cantidad de demanda del bien, está en función de X_2 = precio del bien, X_3 = ingreso del consumidor y X_4 = riqueza del consumidor. Se puede presentar la duda de que si existe una relación entre la variable ingreso y riqueza. El planteo de la hipótesis sujeta a comprobación en este caso sería que los coeficientes del ingreso y la riqueza son los mismos. Entonces tendríamos lo siguiente:

$$H_0: \beta_3 = \beta_4 \text{ o } (\beta_3 - \beta_4) = 0$$
$$H_1: \beta_3 \neq \beta_4 \text{ o } (\beta_3 - \beta_4) \neq 0$$

Para poder aceptar o rechazar el H_0 planteado nos fijamos en los valores t de estas variables (X_3, X_4), si estos t calculados son mayores a los t de las tablas rechazamos H_0 , es decir los dos coeficientes no son iguales. Caso contrario si el tc es menor al tt aceptamos que los coeficientes son iguales.

Lecturas

- Gujarati y Porter (2010). *Econometría*. México: McGraw Hill. Quinta edición. Revise los temas adicionales de esta unidad y complemente el estudio con la revisión de la parte teórica y de ejercicios.
- Wooldridge (2009). Introducción a la econometría: un enfoque moderno. Cuarta edición. CENGAGE Learning.

Revise el capítulo 3 y 4 que le permitirá entender cómo funcionan las formas funcionales para modelos de regresión simple.





Actividades de aprendizaje recomendadas

Nota: conteste las actividades en un cuaderno de apuntes o en un documento Word en caso de ser necesario.

- **Actividad 1**

Continúe con los contenidos del capítulo 7 del texto básico y revise los ejemplos y ejercicios de la guía virtualizada.

Procedimiento: Para una mejor comprensión de los temas desarrollados, es importante que revise varias fuentes de información, sobre los temas vistos y realice comparaciones en cuanto a conceptos, terminologías y procedimientos.

- **Actividad 2**

Estimado estudiante, para evaluar su aprendizaje puede proceder a realizar la autoevaluación de la unidad 7.

Procedimiento: En esta actividad debe contestar las preguntas al final de la unidad 7, esta tarea tiene la finalidad de comprobar que usted está teniendo sintonía con los temas tratados.

Recuerde que la autoevaluación no es obligatoria, sin embargo, le ayuda a reforzar sus conocimientos y lo prepara para la evaluación bimestral.

¡Vamos a trabajar!



Autoevaluación 7

Considerando el siguiente enunciado, seleccione la alternativa correcta.

1. Los coeficientes de correlación entre regresoras y regresores se denominan:
 - a. Parámetros.
 - b. Variables.
 - c. Coeficientes de correlación simple.
2. ¿Cuál es un coeficiente de primer orden?
 - a. r 1.23
 - b. r 12.3
 - c. r 12. 34
3. Si hay relación entre X2 y X3 se dice que son:
 - a. Colineales.
 - b. Insesgados.
 - c. Sesgados.
4. Una de las metodologías para estimar un modelo de regresión múltiple es:
 - a. Intervalo de confianza.
 - b. Nivel de significancia.
 - c. Máxima verosimilitud.
5. Una variable estandarizada es aquella que:
 - a. Obtiene la media de la regresión.
 - b. Expresa las variables en una sola medida.
 - c. Obtiene la varianza de una variable.
6. Una variable estandarizada tiene:
 - a. Varianza de cero.
 - b. Varianza infinita.
 - c. Varianzas unitarias.

7. A medida que aumenta el número de regresoras, R2:
 - a. Aumenta mínimamente.
 - b. Disminuye considerablemente.
 - c. Aumenta considerablemente.

8. Otro criterio para medir la bondad de ajuste de un modelo es:
 - a. Error estándar.
 - b. Criterio de Akaike.
 - c. Intervalo de confianza.

[Ir al solucionario](#)



Unidad 8. Análisis de regresión múltiple: el problema de inferencia

8.1. Pruebas de hipótesis sobre coeficientes de regresión individuales

8.2. Prueba de significancia general de la regresión muestral

8.3. Prueba de igualdad de dos coeficientes de regresión

8.4. Mínimos cuadrados restringidos: pruebas de restricciones de igualdad lineales

8.5. Predicción con regresión múltiple

8.6. Prueba para la estabilidad estructural de los modelos de regresión: la prueba de Chow

8.7. Prueba de la forma funcional de la regresión: elección entre modelos de regresión lineal y log-lineal

Los cinco primeros temas al igual que el modelo de regresión simple, buscan validar el modelo a través de la significativa de sus coeficientes de forma individual, así como su significativa a nivel de toda la regresión. Este tema es abordado a continuación y para una mayor comprensión revise el capítulo 8 de su texto base.

Prueba de estabilidad estructural para los modelos de regresión: la prueba de Chow.

Esta prueba generalmente busca determinar si en un modelo que trabaja con información de serie de tiempo, existió un cambio brusco en la información a lo largo del tiempo. Este cambio generalmente se debe a factores externos, y aquel punto del tiempo puede ser considerado como la fecha de cambio. Esto lo revisaremos con mayor profundidad en el capítulo 9, que corresponde a variables dicotomas.

Lecturas

- Gujarati y Porter (2010). *Econometría*. México: McGraw Hill. Quinta edición.

 El libro le proveerá de todo el marco teórico acerca del concepto de regresión múltiple y su aplicabilidad y validación de resultados dentro de las ciencias sociales.

- Wooldridge (2009). Introducción a la econometría: un enfoque moderno. Cuarta edición. CENGAGE Learning.

Revise el capítulo 3 y 4 que le permitirá entender cómo funcionan las formas funcionales para modelos de regresión simple.



Actividades de aprendizaje recomendadas

- **Actividad 1**

Lea detenidamente los contenidos del capítulo 8 del texto básico y revise los ejemplos y ejercicios de la guía virtualizada.

Procedimiento: Para una mejor comprensión de los temas desarrollados, es importante que revise varias fuentes de información, sobre los temas vistos y realice comparaciones en cuanto a conceptos, terminologías y procedimientos.

Nota: conteste *las actividades en un cuaderno de apuntes o en un documento Word en caso de ser necesario*.

Resultado de aprendizaje 5

- Comprende las características de una variable dicótoma (cualitativa) y su aplicación en modelos de regresión.

Contenidos, recursos y actividades de aprendizaje



Semana 13

Unidad 9. Modelos de regresión con variables dicótomas

9.1. Naturaleza de las variables dicótomas

9.2. Modelos ANOVA

9.3. Modelos ANCOVA

9.4. La variable dicótoma alternativa a la prueba de Chow

En el análisis de regresión no solo se trabaja con variables cuantitativas (por ejemplo: ingreso, producción, precios, costos, inflación, etc.), sino también con variables cualitativas, es decir, variables que expresan presencia o ausencia de una cualidad o característica y que no se expresan de manera numérica (por ejemplo: raza, religión, sexo, nacionalidad, región geográfica, cambios políticos, etc.). Para entender mejor este concepto revisemos el siguiente ejemplo:

- Un modelo en el cual el consumo está en función del ingreso, a más de esta variable, el consumo podría verse también condicionado por variables como la zona geográfica, o por la religión, o por la raza, cualquiera de estas tendría un efecto sobre las variaciones en la variable dependiente.

$$Consumo = f(Ingresa, raza, religion)$$

$$C = f(Y, Rz, Re)$$

Este modelo ha pasado de ser un modelo de regresión simple a unos de regresión múltiple, por la inclusión de estas variables denominadas dicótomas, es decir, variables expresadas en escala nominal. Para poder cuantificar estas variables y sus atributos, debemos generar variables artificiales que toman los valores de 0 y 1. Estas se representan de la siguiente manera:

1 = presencia o posesión de ese atributo o cualidad.

0 = su ausencia.

Para comprender mejor esta parte revisemos los siguientes ejemplos:

- Persona de sexo: 1 = femenino, 0 = masculino.
- Persona de religión: 1 = católico, 0 = no católico o de cualquier otra religión.
- Persona con: 1 = analfabeto, 0 = alfabeto.

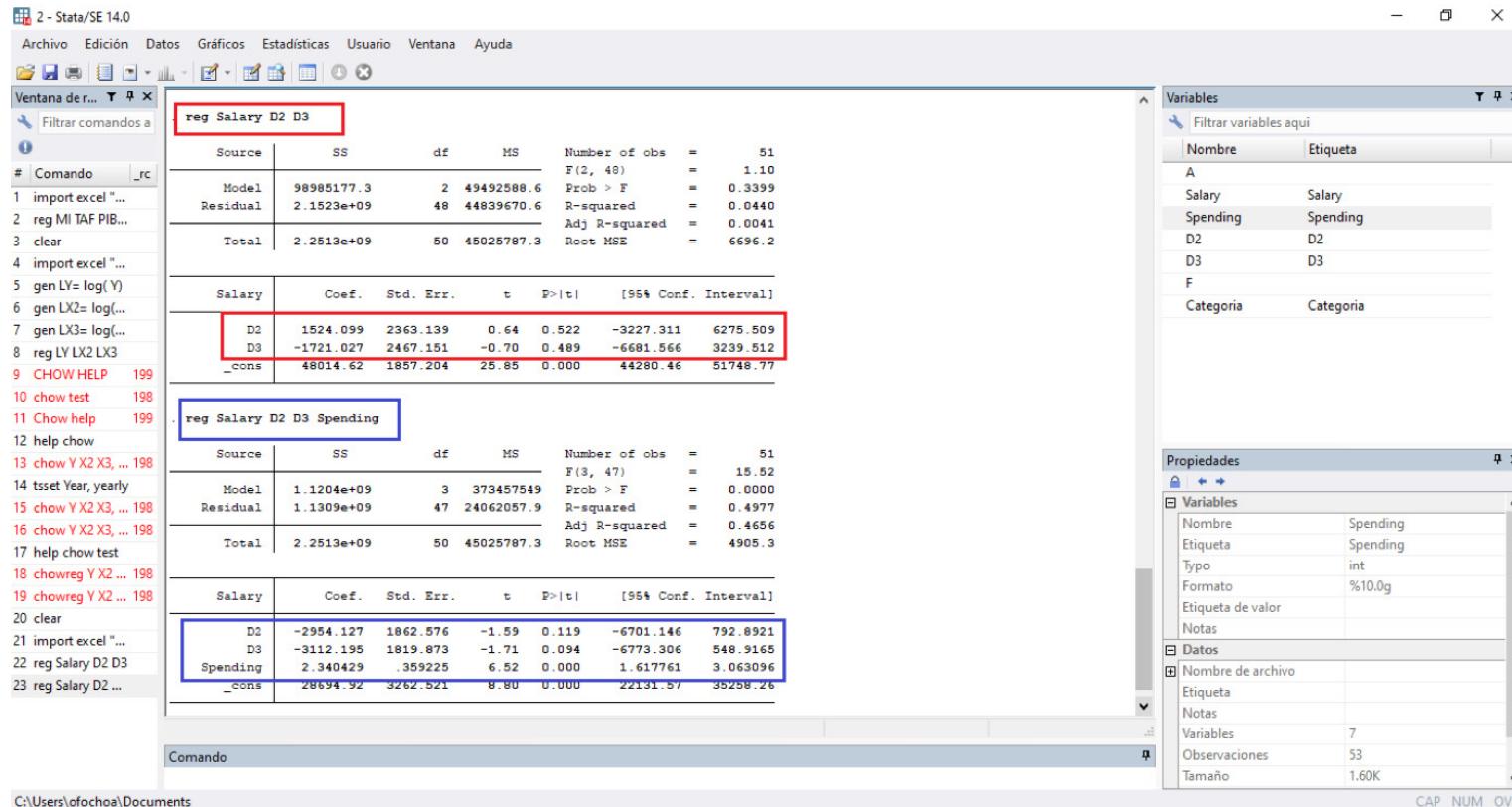
A continuación, le invito a revisar el siguiente recurso interactivo sobre los modelos ANOVA y ANCOVA.

[Modelos ANOVA y ANCOVA](#)

Para entender cómo funcionan este tipo de modelos dentro del STATA, vamos a trabajar con la tabla.

Figura 16.

Modelos ANOVA y ANCOVA



Como puede observar en STATA se realizaron los dos tipos de modelo el primero un ANOVA y el segundo un ANCOVA. La interpretación de resultados se deja para su revisión, y la asignación de valores de 1 y 0 para cada variable dicótoma.

El tema de variables dicótomas, como alternativa a la prueba de Chow, permite conocer si hubo un cambio radical a lo largo de la información y si este cambio afectó a los interceptos, coeficientes o ambos, en las regresiones antes y después del dato que produjo el cambio. Para esto se debe adoptar la siguiente ecuación:

$$Y_t = \alpha_1 + \alpha_2 D_t + \beta_1 X_t + \beta_2 (D_t X_i) + u_i$$

Donde,

α_2 = *intercepto diferencial(D)*

β_2 = *coeficiente de pendiente diferencial(DX)*

Para reforzar lo que acabamos de estudiar, revise el ejemplo 9.5 del texto base, el cual nos indica la interacción entre variables cuantitativas y cualitativas. Para un mayor entendimiento del tema, revise el capítulo 9 de su texto base.

Lecturas

- Gujarati y Porter (2010). *Econometría*. México: McGraw Hill. Quinta edición.

 El libro le proveerá de todo el marco teórico acerca del concepto de regresión y su aplicabilidad dentro de las ciencias sociales.

- Wooldridge (2009). *Introducción a la econometría: un enfoque moderno*. Cuarta edición. CENGAGE Learning.

Revise el capítulo 7 que le permitirá entender cómo funcionan los modelos de regresión con variables cualitativas.



Video

- Revisa cómo funcionan los modelos de regresión con variables dicótomas. [Regresión múltiple de variables dicótomas](#).



Actividades de aprendizaje recomendadas

▪ Actividad 1

Lea detenidamente los contenidos del capítulo 9 del texto básico y revise los ejemplos y ejercicios de la guía virtualizada.

Procedimiento: Para una mejor comprensión de los temas desarrollados, es importante que revise varias fuentes de información, sobre los temas vistos y realice comparaciones en cuanto a conceptos, terminologías y procedimientos. Utilice las técnicas que de acuerdo a su estilo de aprendizaje le sean de mayor utilidad.

Nota: conteste *las actividades en un cuaderno de apuntes o en un documento Word en caso de ser necesario.*



Semana 14

9.5. Efectos de interacción al utilizar variables dicótomas

9.6. Uso de las variables dicótomas en el análisis estacional

9.7. Regresión lineal por segmentos

9.8. Algunos aspectos técnicos de la técnica con variables dicótomas

Regresión lineal por segmentos: para entender este tema realicemos paso a paso el ejemplo 9.7 de su texto básico, el cual presenta un modelo que relaciona el costo total (Y) con la producción (X).

El modelo a desarrollarse para este ejercicio es el siguiente:

$$Y_i = \alpha_1 + \beta_1 X_i + \beta_2 (X_i - X^*) D_i + u_i$$

X^* viene a ser el cambio de la pendiente o de la producción a las 5500 unidades. Para poder realizar esta regresión debemos realizar lo siguiente en STATA.

- Obtener los valores para la variable dicótoma: $D=1$, si X mayor X^* y $D=0$ si X menor X^* , es decir, todos los valores por encima de 5500 toman el valor de 1 y los otros toman los valores de cero.
- Debemos restar a los valores de X el valor de X^* , es decir 5500, y estos valores multiplicarlos por la variable dicótoma. De esta manera obtengo los valores de la variable que permitirá medir el cambio en las pendientes.

Ahora corremos el modelo antes planteado y analizamos los resultados.

Figura 17.

Modelo de regresión

2 - Stata/SE 14.0

Archivo Edición Datos Gráficos Estadísticas Usuario Ventana Ayuda

Ventana de r... T X

Filtrar comandos a

Comando _rc ^

9 CHOW HELP 1...

10 chow test 1...

11 Chow help 1...

12 help chow

13 chow Y X2 X... 1...

14 tsset Year, y...

15 chow Y X2 X... 1...

16 chow Y X2 X... 1...

17 help chow t...

18 chowreg Y ... 1...

19 chowreg Y ... 1...

20 clear

21 import exce...

22 reg Salary D...

23 reg Salary D...

24 clear

25 import exce...

26 clear

27 import exce...

28 import exce... 4

29 clear

30 import exce...

31 clear

32 import exce...

33 gen X5500= ...

34 gen DX5500...

35 reg Totalcos...

C:\Users\ofochoa\Documents

Variables

Filtrar variables aquí

Nombre	Etiqueta
Totalcost	Total cost (\$)
Outputunits	Output (units)
Dicotoma	Dicotoma
X5500	
DX5500	

Propiedades

Variables

Nombre	DX5500
Etiqueta	
Tipo	float
Formato	%9.0g
Etiqueta de valor	
Notas	

Datos

Nombre de archivo	
Etiqueta	
Notas	
Variables	5
Observaciones	10
Tamaño	140

Comando

```
. clear

. import excel "C:\Users\ofochoa\Desktop\Respaldos ochoa\Documents\RESPALDOS ochoa\Documents\Francisco\FRANCISCO\Modalidad Presencial\Econometria ECONOMIA\Tablas libro GUJARATI\Tablas Gujarati nueva versión\Table 9_1.xls", sheet("Table 9_1")
> firstrow

. clear

. import excel "C:\Users\ofochoa\Desktop\Respaldos ochoa\Documents\RESPALDOS ochoa\Documents\Francisco\FRANCISCO\Modalidad Presencial\Econometria ECONOMIA\Tablas libro GUJARATI\Tablas Gujarati nueva versión\Table 9_6.xls", sheet("Table 9_6")
> firstrow

gen X5500= ( Outputunits - 5500)

gen DX5500= ( X5500* Dicotoma)

reg Totalcost Outputunits DX5500
```

Source	SS	df	MS	Number of obs	=	10
Model	8832644.9	2	4416322.45	F(2, 7)	=	129.61
Residual	238521.502	7	34074.5002	Prob > F	=	0.0000
				R-squared	=	0.9737
				Adj R-squared	=	0.9662
Total	9071166.4	9	1007907.38	Root MSE	=	184.59

Totalcost	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Outputunits	.2791258	.0460081	6.07	0.001	.1703338 .3879177
DX5500	.0945	.0825524	1.14	0.290	-.1007054 .2897054
_cons	-145.7167	176.7341	-0.82	0.437	-563.6265 272.1932

Ahora basándose en el texto básico y el ejercicio, realice las interpretaciones de este ejercicio.

Lecturas

- Gujarati y Porter (2010). *Econometría*. México: McGraw Hill. Quinta edición.



El libro le proveerá de todo el marco teórico acerca del concepto de regresión y su aplicabilidad dentro de las ciencias sociales.

- Wooldridge (2009). Introducción a la econometría: un enfoque moderno. Cuarta edición. CENGAGE Learning.

Revise el capítulo 7 que le permitirá entender cómo funcionan las formas funcionales para modelos de regresión múltiple.



Actividades de aprendizaje recomendadas

▪ Actividad 1

Continúe leyendo los contenidos del capítulo 9 del texto básico y revise los ejemplos y ejercicios de la guía didáctica.

Procedimiento: Para una mejor comprensión de los temas desarrollados, es importante que revise varias fuentes de información, sobre los temas vistos y realice comparaciones en cuanto a conceptos, terminologías y procedimientos.

Utilice las técnicas que de acuerdo a su estilo de aprendizaje le sean de mayor utilidad.

▪ Actividad 2

Estimado estudiante, para evaluar su aprendizaje puede proceder a realizar la autoevaluación de la unidad 9.

Procedimiento: En esta actividad debe contestar las preguntas al final de la unidad 9, esta tarea tiene la finalidad de comprobar que usted está teniendo sintonía con los temas tratados.

Recuerde que la autoevaluación no es obligatoria, sin embargo, le ayuda a reforzar sus conocimientos y lo prepara para la evaluación bimestral.

¡Vamos a trabajar!



Autoevaluación 8

Considerando el siguiente enunciado, seleccione la alternativa correcta.

1. Una variable dicótoma también se denomina: (2 respuestas)
 - a. Variable categórica.
 - b. Variable cualitativa.
 - c. Variable cuantitativa.
 - d. Variable explicativa.
2. Una variable dicótoma en un modelo de regresión puede ocasionar: (2 respuestas)
 - a. Heterocedasticidad.
 - b. Homocedasticidad.
 - c. Igualdad de las varianzas de las variables.
 - d. Diferencia en las varianzas de las variables.
3. En el análisis de regresión, la variable dependiente a menudo causa influencia no solo de variables en escala de razón sino también de variables: (2 respuestas)
 - a. Cualitativa.
 - b. Dicótoma.
 - c. Explicada.
 - d. Dependiente.
4. Cuando se utiliza variables dicótomas en los modelos de regresión se tiene que tener en cuenta: (2 respuestas)
 - a. No cometer la trampa de la variable dicótoma.
 - b. Que la inclusión de la variable dicótoma no produzca heterocedasticidad.
 - c. Que no disminuyan el número de observaciones.
 - d. Que exista variación en las variables.

5. Para eludir la trampa de la variable dicótoma introduciendo tantas variables dicótomas como categorías se puede:
 - a. Eliminar datos.
 - b. Incrementar la muestra.
 - c. No se debe introducir el intercepto en dicho modelo.
6. Cuando se elimina el intercepto y se permite una variable dicótoma para cada categoría se obtiene directamente:
 - a. Los valores medios de las distintas categorías.
 - b. Las elasticidades.
 - c. Los errores estándar.
7. ¿Cómo se coloca la categoría base en una variable dicótoma?
 - a. La primera categoría siempre es la base.
 - b. Dependiendo del criterio del investigador.
 - c. La última categoría siempre es la base.

[Ir al solucionario](#)

Resultado de aprendizaje 6

- Aplica los supuestos básicos del modelo de regresión.

Contenidos, recursos y actividades de aprendizaje



Semana 15

Unidad 10. Multicolinealidad: ¿qué pasa si las regresoras están correlacionadas?

10.1. Naturaleza de la multicolinealidad

10.2. Estimación en presencia de multicolinealidad

10.3. Consecuencias prácticas de la multicolinealidad

10.4. Detección de la multicolinealidad

10.5. Ejemplo ampliado

La multicolinealidad es la relación lineal perfecta entre las variables explicativas de un modelo de regresión. Aunque en la práctica adicional a los modelos lineales se suele trabajar con modelos no lineales, en los cuales el fenómeno de la multicolinealidad también puede darse entre sus variables. Si un modelo se calcula en presencia de multicolinealidad, los estimadores dejan de ser confiables, lo que generaría que los resultados obtenidos no sean los más adecuados para explicar cierto fenómeno económico.

Vamos a revisar el siguiente ejercicio y veremos cómo detectar este problema a través del STATA. [La multicolinealidad](#).

Lectura

- Gujarati y Porter (2010). *Econometría*. México: McGraw Hill. Quinta edición.



El libro le proveerá de todo el marco teórico acerca del concepto de regresión y su aplicabilidad dentro de las ciencias sociales.

Video



- El siguiente video explica cómo funciona el [problema de multicolinealidad](#).



Actividades de aprendizaje recomendadas

▪ Actividad 1

Continúe leyendo los contenidos del capítulo 10 del texto básico y revise los ejemplos y ejercicios de la guía didáctica.

Procedimiento: Para una mejor comprensión de los temas desarrollados, es importante que revise varias fuentes de información, sobre los temas vistos y realice comparaciones en cuanto a conceptos, terminologías y procedimientos.

▪ Actividad 2

Estimado estudiante, para evaluar su aprendizaje puede realizar la autoevaluación de la unidad 10.

Procedimiento: En esta actividad debe contestar las preguntas al final de la unidad 10, esta tarea tiene la finalidad de comprobar que usted está teniendo sintonía con los temas tratados.

Recuerde que la autoevaluación no es obligatoria, sin embargo, le ayuda a reforzar sus conocimientos y lo prepara para la evaluación bimestral.

¡Vamos a trabajar!



Autoevaluación 9

Considerando el siguiente enunciado, seleccione la alternativa correcta.

1. ¿En qué consiste la multicolinealidad?
 - a. Consiste en la relación entre las variables independientes del modelo uniecuacional múltiple y su rezago en el tiempo.
 - b. Consiste en la existencia de relaciones lineales entre dos o más variables independientes del modelo uniecuacional múltiple.
 - c. Consiste en la diferencia de las varianzas entre dos o más variables independientes del modelo uniecuacional simple.
2. ¿Cuáles son las principales causas que produce multicolinealidad en un modelo?
 - a. Amplia diferencia entre las varianzas de las variables independientes.
 - b. Escasa variabilidad en las observaciones de las variables independientes.
 - c. No normalidad de las variables independientes.
3. ¿Qué es la multicolinealidad exacta o perfecta?
 - a. Es la existencia de una relación lineal exacta entre dos o más variables independientes.
 - b. Es la relación lineal aproximada entre dos o más variables independientes.
 - c. Es la relación entre las varianzas de dos o más variables independientes.
4. ¿Qué es la multicolinealidad leve o aproximada?
 - a. Es la relación lineal aproximada entre dos o más variables independientes.
 - b. Es la existencia de una relación lineal exacta entre dos o más variables independientes.
 - c. Es la relación lineal aproximada entre dos o más variables independientes.

5. Cuando la multicolinealidad es perfecta:
 - a. Se incumple la hipótesis básica del modelo uniecuacional múltiple: la matriz X no es de rango completo por columnas, esto es $rg(x) < K$.
 - b. Se incumple el supuesto de homocedasticidad de las varianzas esto es se incumple $G_1 = G_2$.
 - c. Se incumple el supuesto de normalidad, es decir $N < 30$.
6. Cuando la multicolinealidad es leve o aproximada
 - a. No se incumple la hipótesis básica del modelo uniecuacional múltiple: la matriz X es de rango completo por columnas, esto es: $rg(x) = K$.
 - b. Se incumple el supuesto de homocedasticidad de las varianzas esto se incumple $G_1 = G_2$.
 - c. Se incumple el supuesto de normalidad.
7. Cuando la multicolinealidad es leve o moderada se pueden presentar los siguientes problemas:
 - a. Existen coeficientes de determinación muy elevados.
 - b. Existen t muy significativas.
 - c. Existen chi 2 muy elevados.
8. ¿Qué métodos se pueden utilizar para la detección de multicolinealidad?
 - a. Prueba de White.
 - b. Correlograma.
 - c. Factor inflador de la varianza.

[Ir al solucionario](#)



Actividades finales del bimestre

Revisión de las unidades de la 6 a la 10

Debe revisar nuevamente los contenidos del capítulo 6 al 10, para un reforzamiento de los contenidos y como preparación previa al examen presencial.

Lectura

- Gujarati y Porter (2010). *Econometría*. México: McGraw Hill. Quinta edición.



El libro le proveerá de todo el marco teórico acerca del concepto de regresión y su aplicabilidad dentro de las ciencias sociales.



Actividades de aprendizaje recomendadas

Nota: conteste las actividades en un cuaderno de apuntes o en un documento Word en caso de ser necesario.

▪ Actividad 1

Lea los contenidos del texto básico del capítulo 6 al 10 para explicaciones sobre los temas desarrollados.

Procedimiento: Para el desarrollo de esta actividad es importante que realice una lectura comprensiva de todos los temas abordados y que vaya trabajando mediante cuadros sinópticos o resúmenes para que luego pueda revisarlos con mayor facilidad.

- **Actividad 2**

Desarrolle las actividades recomendadas.

Procedimiento: Es importante que aun no siendo obligatoria su presentación, usted desarrolle las actividades recomendadas ya que de esa manera puede desarrollar las habilidades y destrezas en la aplicación de los contenidos en los aspectos de la vida práctica.



4. Solucionario

Autoevaluación 1		
Pregunta	Respuesta	Retroalimentación
1	c	Se busca dependencia estadística.
2	a	La relación de las variables debe ser estocástica.
3	a	La correlación es la asociación entre variables.
4	a	Son sinónimos de la variable dependiente.
5	c	Son sinónimos de la variable independiente.
6	b	Información que se genera en distintos puntos del tiempo.
7	c	Información que se genera en un punto específico del tiempo.
8	c	Mix de información de serie de tiempo y corte transversal.

Ir a la
autoevaluación

Autoevaluación 2

Pregunta	Respuesta	Retroalimentación
1	c	La función se da sobre la variable independiente.
2	b	Los parámetros son lineales.
3	a	Es función de la variable X.
4	c	El intercepto o la constante.
5	b	El error estocástico agrupa aquellas variables que afectan a Y.
6	a	Variables muy similares a las originales.
7	b	Un modelo debe ser eficiente.
8	a	Es la relación que se da entre las variables.

Ir a la
autoevaluación

Autoevaluación 3		
Pregunta	Respuesta	Retroalimentación
1	a	En la homocedasticidad las varianzas de los residuos son iguales.
2	c	El coeficiente de determinación fluctúa entre 0 y 1.
3	a	Los MCO son una de varias metodologías para estimar una regresión.
4	a	Indica cómo se relacionan las variables.
5	b	Ante una diferente varianza de los residuos existe el problema de heterocedasticidad.
6	b	Deben ser lo más puntuales o confiables posibles.
7	c	Son siete supuestos los que rigen a la metodología de MCO.
8	a	Se debe dar la independencia de las variables.

[Ir a la autoevaluación](#)

Autoevaluación 4		
Pregunta	Respuesta	Retroalimentación
1	a	Se ve la distribución de los valores con respecto a la media.
2	b	Los estadísticos tal como: F, t, X ₂ , Z, sirven para probar hipótesis.
3	a	Una propiedad de la normalidad es que no exista sesgo en la distribución de los residuos.
4	c	El método de MV es mucho más fuerte que el de MCO, pero matemáticamente más complicado.
5	c	La varianza debe ser constante.
6	b	El supuesto de normalidad parte del teorema del límite central.
7	a	La metodología de MCO, permite obtener los mejores estimadores de una regresión.
8	c	Un método alternativo al de MCO, es el de MV.

Ir a la
autoevaluación

Autoevaluación 5

Pregunta	Respuesta	Retroalimentación
1	b	Es simple porque se prueba un único resultado, es decir la igualdad de un resultado.
2	c	Su probabilidad es de $1 - \alpha$.
3	a	Es bilateral porque hace referencia a las dos colas de la distribución.
4	c	Es rechazar una hipótesis siendo esta verdadera.
5	c	Para probar si un modelo en su conjunto es significativo, utilizamos el estadístico F.
6	a	Uno de los métodos para probar la significatividad de un resultado es a través de los intervalos de confianza.
7	b	La hipótesis nula siempre se prueba frente a la hipótesis alternativa.
8	a	Es la condición que nos permite contrastar la H_0 , en este caso el de rechazar la H_0 .

Ir a la
autoevaluación

Autoevaluación 6

Pregunta	Respuesta	Retroalimentación
1	c	Una variable estandarizada utiliza una suma de cuadrados simple.
2	a	Por su forma de cálculo se puede obtener un R ² negativo, pero solo si no se tiene el intercepto.
3	b	El tema de la falta de intercepto puede considerarse como un error de especificación.
4	a	El tema de las escalas no influye sobre las propiedades de los estimadores.
5	b	Una variable estandarizada tiene una media =0.
6	a	Un modelo log-lineal es aquel cuyas variables son logarítmicas.
7	b	Un modelo doble logarítmico nos ayuda a calcular elasticidades.
8	c	Generalmente un tema de crecimiento se ajusta mejor a un modelo Log-Lin.

Ir a la
autoevaluación

Autoevaluación 7

Pregunta	Respuesta	Retroalimentación
1	c	En un modelo de regresión múltiple se obtienen los coeficientes de correlación.
2	b	Es un coeficiente de primer orden porque refleja la relación de la variable Y con X2, dejando constante X3.
3	a	Cuando dos variables están relacionadas se dice que existe colinealidad entre ellas.
4	c	Al igual que en modelos de regresión simple, un modelo múltiple puede también ser estimado por MCO o por MV.
5	b	El estandarizar una variable permite mantener una única unidad de medida de las variables.
6	c	Una propiedad es que su varianza es igual a 1.
7	a	El incorporar muchas variables independientes altera mínimamente el valor del R ² .
8	b	El criterio de AKAIKE sirve también para elegir entre dos modelos.

Ir a la
autoevaluación

Autoevaluación 8

Pregunta	Respuesta	Retroalimentación
1	a - b	Varios sinónimos para una variable categórica o cualitativa.
2	a - d	Puede generar un problema de heterocedasticidad.
3	a - b	Un modelo de regresión puede contener también variables independientes cualitativas.
4	a - b	No asignar igual número de variables por categorías.
5	c	Se puede evitar la trampa de la variable dicótoma estableciendo un modelo de regresión a través del origen.
6	a	Los coeficientes de regresión que se obtienen de un modelo sin intercepto nos da las variaciones promedio de las categorías de las variables dicotómicas.
7	b	La elección de la categoría base depende del investigador.

[Ir a la autoevaluación](#)

Autoevaluación 9		
Pregunta	Respuesta	Retroalimentación
1	b	La multicolinealidad es la relación entre las variables independientes.
2	b	La no variabilidad de los datos de las variables puede ocasionar multicolinealidad.
3	a	Es la relación perfecta o del 100% entre variables independientes.
4	a	Es una relación menos que perfecta existente entre las variables independientes.
5	a	Cuando no se cumple la hipótesis de un modelo uniecuacional donde la matriz x no es de rango completo.
6	a	Cuando la relación entre las variables independientes es menor al 50%.
7	a	Un síntoma de un problema de multicolinealidad es un R2 alto.
8	c	El FIV nos permite detectar si las variables independientes son altamente colineales.

Ir a la
autoevaluación



5. Referencias bibliográficas

Cuellar L. (2019) Regresión lineal en STATA: interpretación de salida.
Recuperado de: [enlace web](#)

Cuellar L. (2019) Evaluando normalidad en STATA. Recuperado de: [enlace web](#)

Recuperado de Cuellar L. (2019). Modelo con variables categóricas. [enlace web](#)

Recuperado de Cuellar, L. (2019). Cómo funciona el problema de la multicolinealidad. [enlace web](#)

Gujarati, D. & Porter, D. (2010). *Econometría*. México: McGraw Hill. Quinta edición.

Modelos logarítmicos y como realizarlos STATA. Recuperado de Economics en tu PC. [enlace web](#)

Ochoa, O. (2009). *Guía de Econometría 1*. UTPL.

Ochoa, O. (2009). *Guía de Econometría 2*. UTPL.

Peña, J. Estavillo, J. (1999): *CIEN EJERCICIOS DE ECONOMETRÍA*, Madrid, Ediciones Pirámide.

Wooldridge, J. (2009). *Introducción a la econometría: un enfoque moderno*. Cuarta edición. CENGAGE Learning.



6. Anexos

Ejercicio práctico

Planteamiento de la teoría o hipótesis

- Consumo (C) en función del ingreso (Y)
- $C=f(Y)$
- Esta ecuación significa que el consumo está en función del ingreso, es decir, que los cambios o variaciones que se den en el ingreso repercuten en el consumo, estos pueden ser de manera directa o inversa.

Especificación modelo matemático

- $Y = a + bX$
- $Y = f(X)$
- $Y = B_0 + B_1 X$

Especificación econométrica

- $Y = B_0 + B_1 X + u$
- Se añade el término de error, que representa a todas aquellas variables que tienen alguna relación o influencia sobre la variable dependiente Y.

Obtención de datos

- Información de series de tiempo.
- Información de corte transversal.
- Información de panel o mixta

Estimación del modelo

- Se presentan los resultados de la regresión planteada, aquí se determinará si las variables que utilice son adecuadas o no con respecto a la investigación que se realizó, es decir, se podrá determinar si la variable independiente (X), es apropiada para explicar los cambios en la variable dependiente (Y).

Planteo de hipótesis

- Suponiendo que el modelo que se planteó es adecuado, lo que se buscaría ahora es determinar si los valores estimados concuerdan con las expectativas de la teoría que se está probando.

Verificación de los resultados

- Verificamos de forma económica, estadística y econométrica los resultados de la regresión.

Predicción

Nota. Gujarati, D & Porter, D. (2010). *Econometría*. México: McGraw Hill. Quinta edición. **Elaboración.** El autor

Ejercicio en el software STATA

- Vamos a importar la base de Excel de la tabla I.1, del texto básico al STATA, como lo muestra la figura 1.

Figura 1.

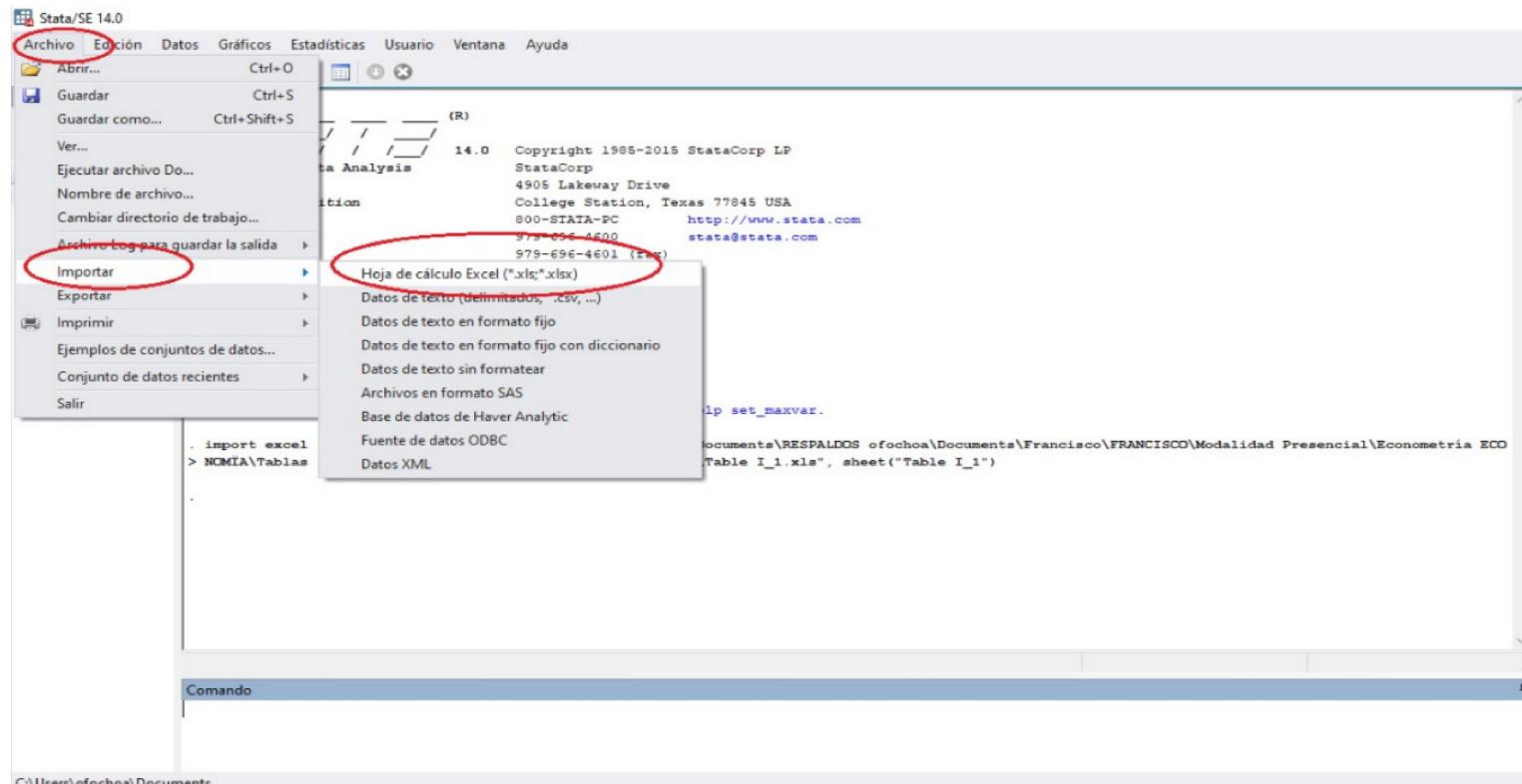
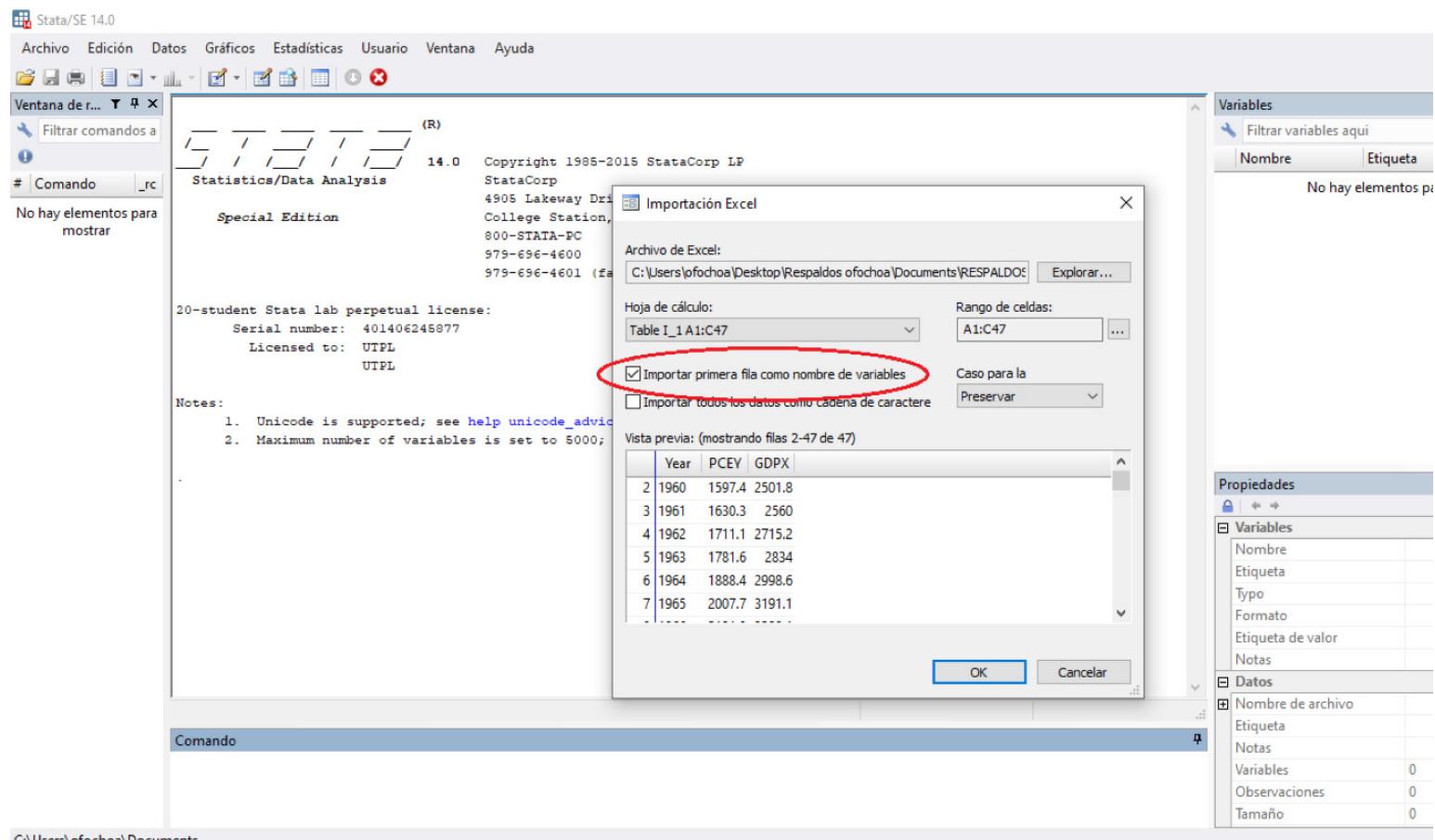


Figura 2.



2. Procedemos a seleccionar el archivo y aceptar (figura 2).

Figura 3.

The screenshot shows the Stata/SE 14.0 software interface. The menu bar includes Archivo, Edición, Datos, Gráficos, Estadísticas, Usuario, Ventana, and Ayuda. The toolbar has icons for file operations like Open, Save, and Print. The main window has a title bar "Ventana de r..." and a status bar "Comando". The command window displays the following text:

```
(R)
Statistics/Data Analysis
14.0 Copyright 1985-2015 StataCorp LP
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)

20-student Stata lab perpetual license:
Serial number: 401406245877
Licensed to: UTPL
UTPL

Notes:
1. Unicode is supported; see help unicode_advice.
2. Maximum number of variables is set to 5000; see help set maxvar.

import excel "C:\Users\ofochoa\Desktop\Respaldos ochoa\Documents\RESPALDOS ochoa\Documents\Francisco\FRANCISCO\Modalidad
> ad Presencial\Econometria ECONOMIA\Tablas libro GUJARATI\Tablas Gujarati nueva versión\Table I_1.xls", sheet("Table I_1")
> firstrow
```

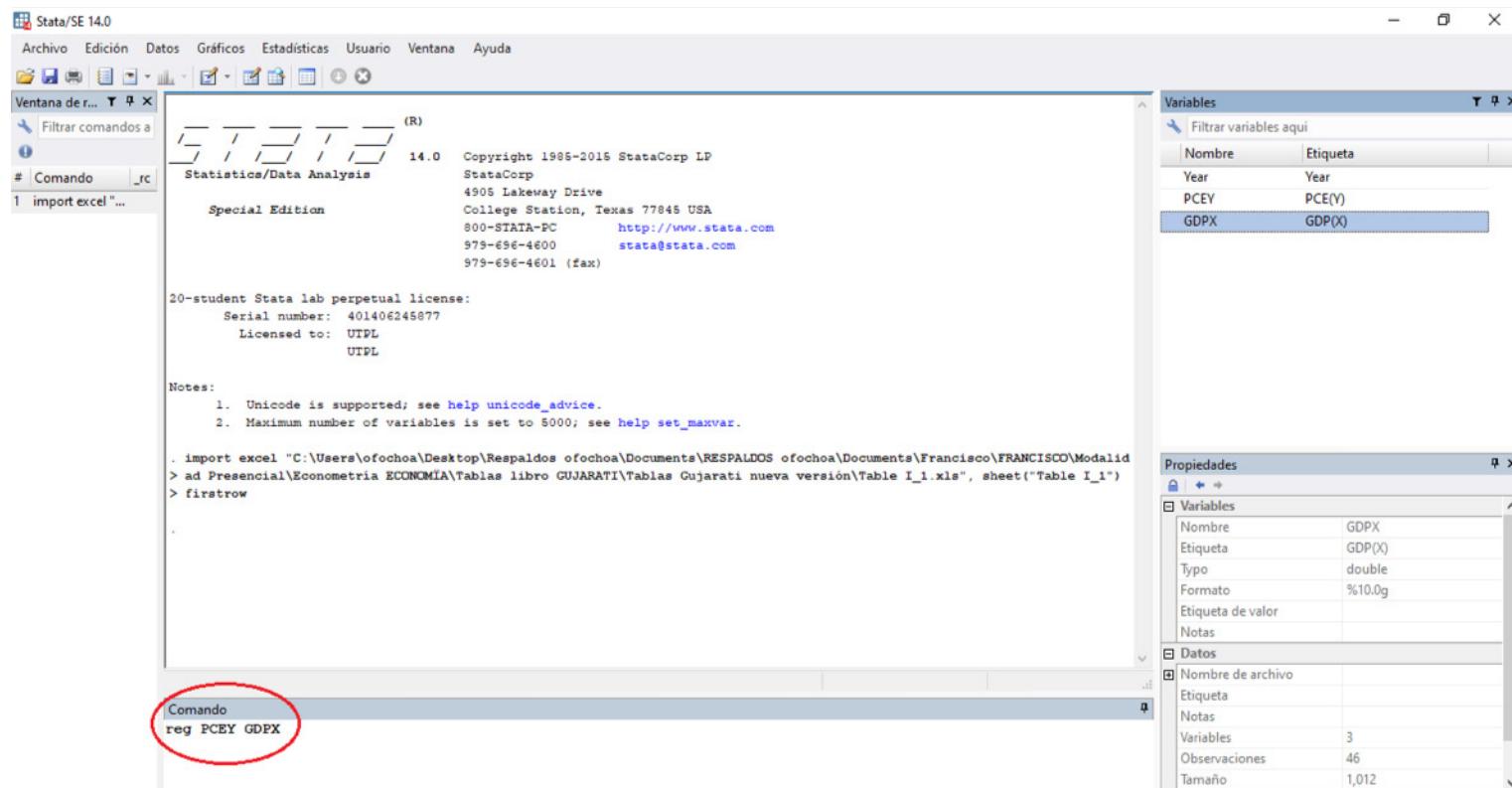
A red box highlights the command ". import excel" and its arguments. The variable viewer on the right shows a table of variables:

Nombre	Etiqueta
Year	Year
PCEY	PCE(Y)
GDPX	GDP(X)

A second red box highlights the "Variables" section of the properties panel, which lists the same variables with their types: Year (Year), PCEY (PCE(Y)), and GDPX (GDP(X)).

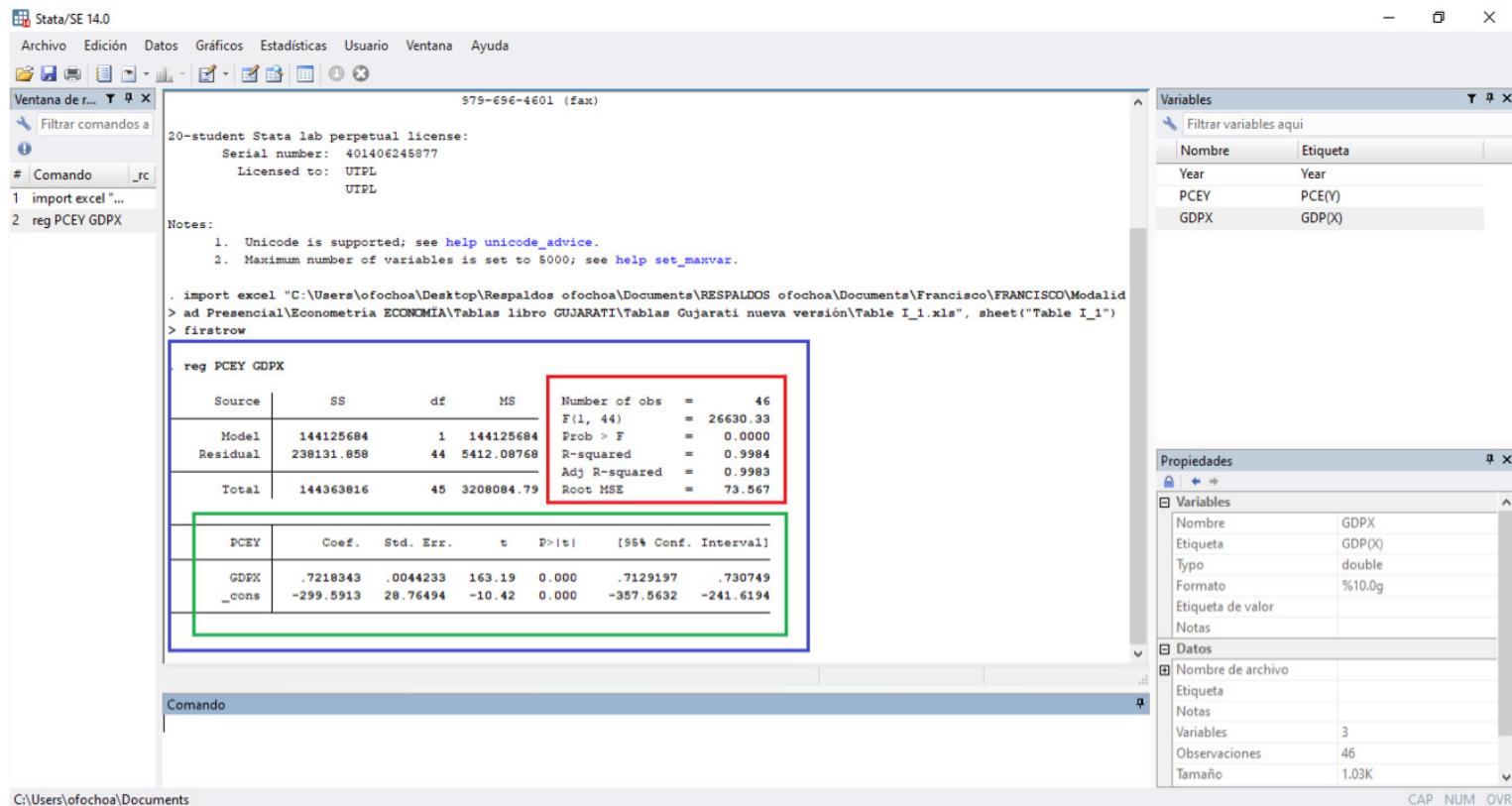
3. Una vez cargada la base de datos (figura 3), procedemos a realizar la regresión a través de la metodología de MCO, como se puede observar las variables ya están cargadas en la parte superior derecha.
4. Procedemos a realizar la regresión de la variable dependiente en función de la variable independiente: PCEY = f (GDPX). Vamos a utilizar el siguiente comando para realizar esta regresión, **reg PCEY GDPX**.

Figura 4.



Los resultados obtenidos de esta regresión son los siguientes:

Figura 5.

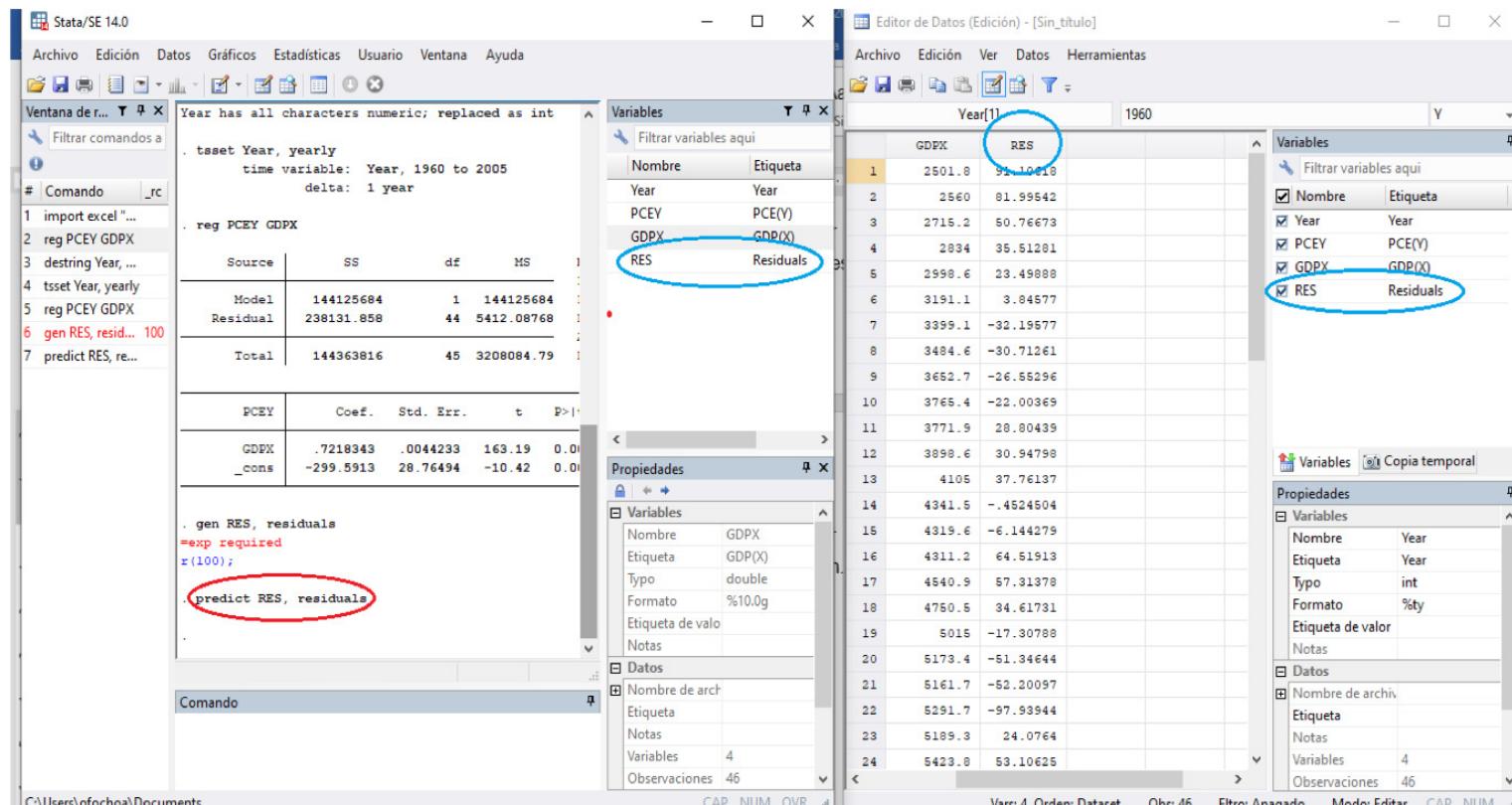


El recuadro en azul indica todos los resultados de esta regresión a través de la metodología de MCO, el recuadro en rojo indica información del modelo en su conjunto y el recuadro verde, en cambio, proporciona información de las variables del modelo en forma individual. Las interpretaciones son similares a la del ejemplo 3.

5. Ahora calculamos los residuos de la regresión, es decir, obtener los valores residuales de la regresión, para esto debemos utilizar los siguientes comandos.

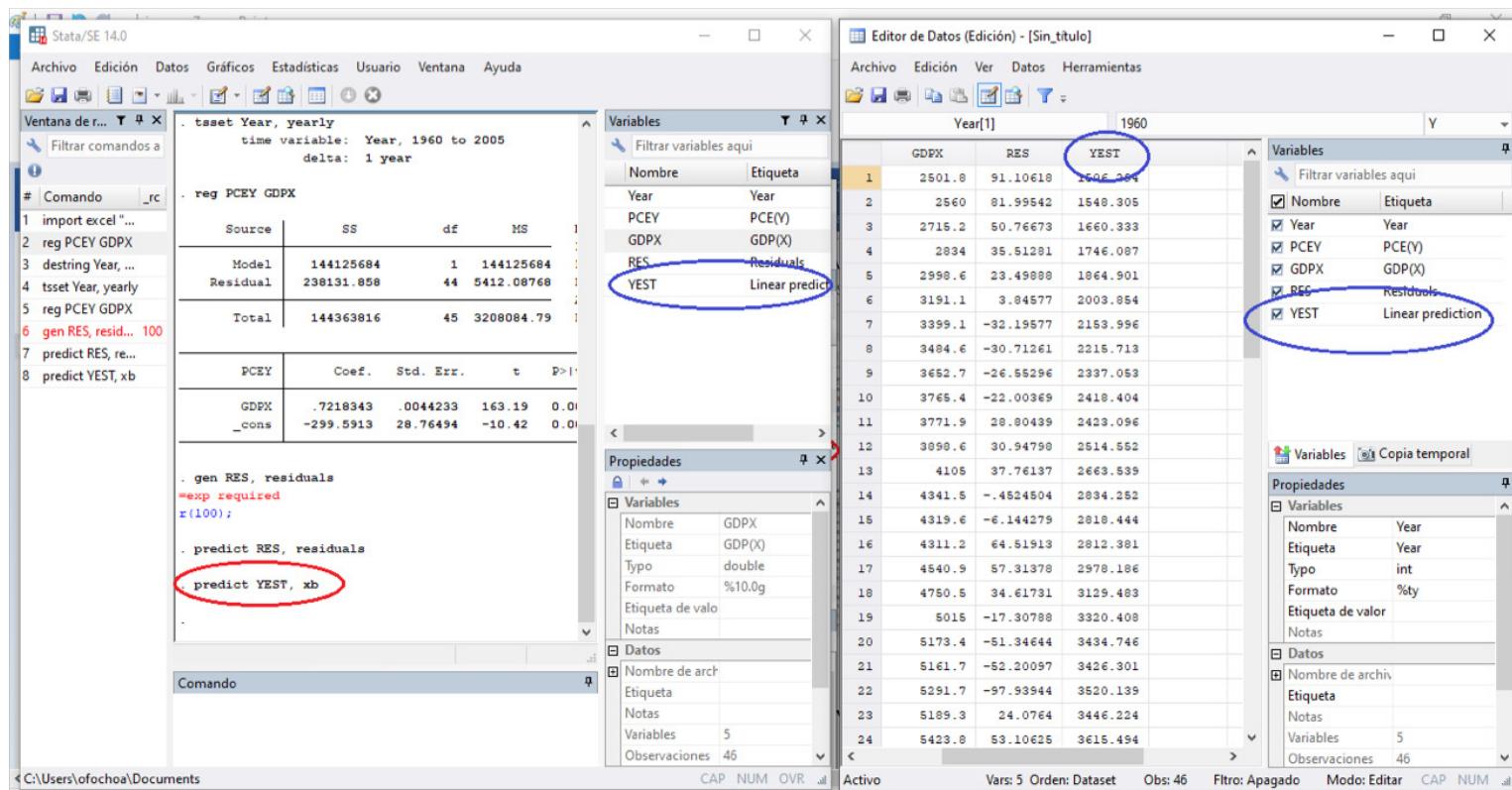
predict RES, residuals: con este comando lo que hacemos es crear una nueva variable con el nombre RES, la que generará los residuos de la regresión.

Figura 6.



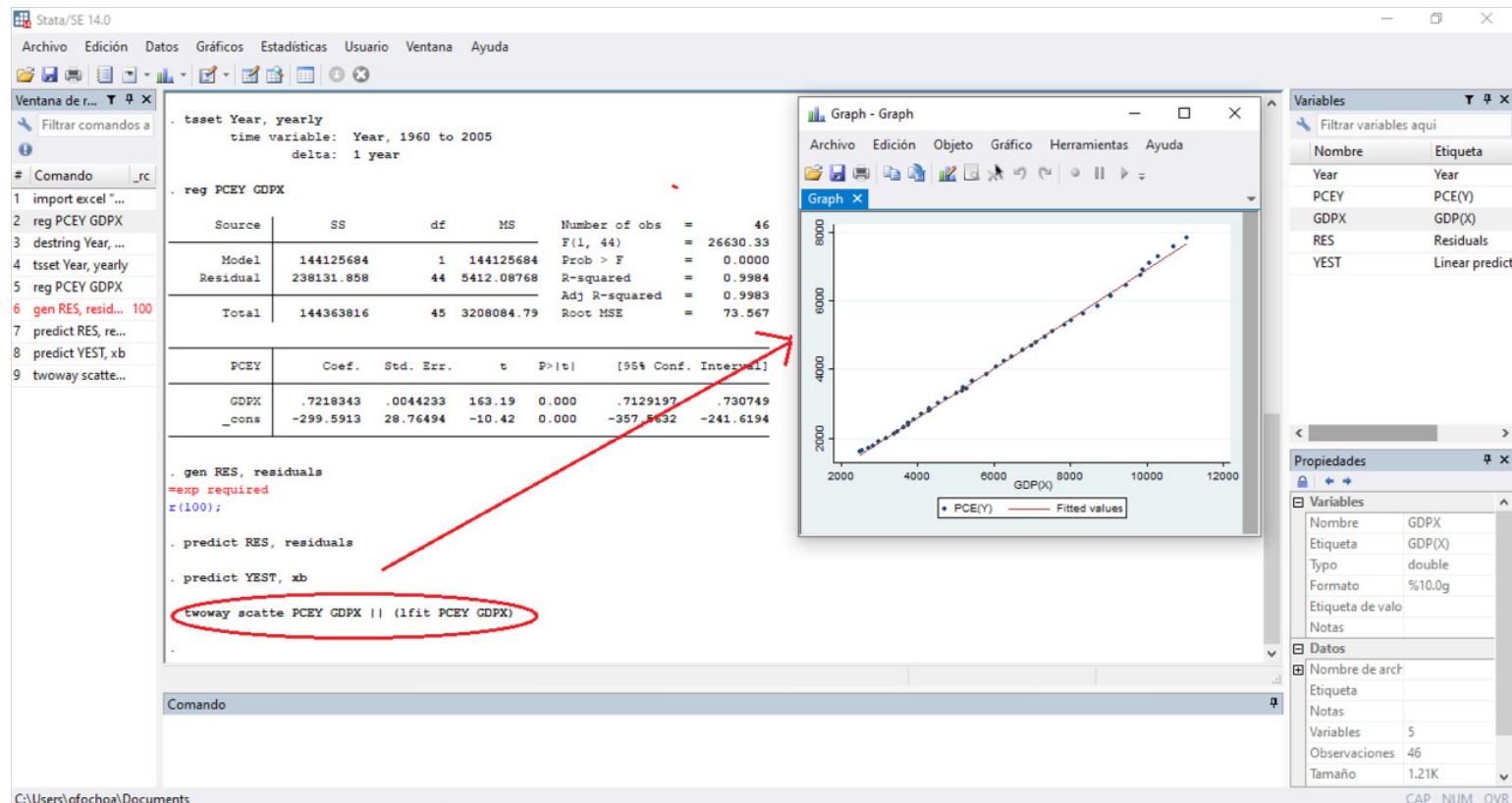
6. También podemos obtener todos los valores estimados para la variable dependiente, esto lo hacemos con el siguiente comando: **predict YEST, xb**

Figura 7.



7. Adicionalmente se puede graficar la línea de regresión que se ajusta a los datos de las variables, para esto utilizamos el siguiente comando: **twoway scatter PCEY GDPX || lfit scatter PCEY GDPX**. Esta gráfica es similar a la figura vista en el ejercicio práctico.

Figura 8.



La multicolinealidad

Revisemos el ejercicio 10.27 del texto básico, el cual proporciona cifras sobre importaciones, PIB e IPC de USA de 1975 a 2005 (tabla 10.13). Vamos a detectar y corregir la multicolinealidad. El modelo planteado es:

The screenshot shows the Stata SE 14.0 interface with the following components:

- Log Output Area:** Displays the log command and its output. The log command includes:

```
. clear  
. import excel "C:\Users\ofochoa\Desktop\Respaldos ofochoa\Documents\RESPALDOS ofochoa\Documents\Francisco\FRANCISCO\Modalidad ad Presencial\Econometria ECONOMIA\Tablas libro GUJARATI\Tablas Gujarati nueva versión\Table 10_13.xls", sheet("Table 10_1 2") firstrow  
gen LIPC=log( IPC)  
gen LPPIB=log( PIB )  
gen LIMPOR=log( Imports )  
reg LIMPOR LIPC LPPIB
```

The output of the regression command is:

Source	SS	df	MS
Model	17.2842186	2	8.64210929
Residual	.139293109	28	.004974754
Total	17.4235117	30	.580783723

Number of obs = 31
F(2, 28) = 1737.19
Prob > F = 0.0000
R-squared = 0.9920
Adj R-squared = 0.9914
Root MSE = .07053

LIMPOR	Coef.	Std. Err.	t	P> t	(95% Conf. Interval)
LIPC	-.8733684	.2848051	-3.07	0.005	-1.456765 -.2899715
LPPIB	1.850098	.1829115	10.11	0.000	1.475421 2.224775
_cons	1.409417	.2700739	5.22	0.000	.0561958 1.962630
- Command History:** Shows a list of previous commands run in the session.
- Variables View:** A tree view of variables in the current dataset, showing their names and types. Variables LIPC, LPPIB, and LIMPOR are highlighted with red boxes.
- Properties View:** A detailed view of the selected variable properties, including Name (LPIB), Type (float), Format (%9.0g), and Value Label.
- Datos View:** Summary statistics for the dataset, including Name of file, Labels, Notes, Variables (15), Observations (31), and Size (1.39K).

Aquí se presentan los resultados de esta regresión múltiple, y lo que vamos a hacer es detectar si hay un problema de multicolinealidad.

Nos damos cuenta de que tenemos un R^2 alto casi del 99%, esto podría ser una señal de multicolinealidad, a más de esto vamos a establecer una matriz de correlación que nos permitirá ver la relación entre las variables independientes, el comando es **corr LIPC LPIB**.

The screenshot shows the Stata/SE 14.0 interface with the following details:

- Command History:** The left pane displays a series of Stata commands and their outputs. A blue box highlights the command `corr LIPC LPIB obs=31`. A red box highlights the resulting correlation matrix table.
- Results of a Regression Model:** Below the command history, the output of a regression command is shown:

Source	SS	df	MS	Number of obs	=	31
Model	17.2842186	2	8.64210929	F(2, 28)	=	1737.19
Residual	.139293109	28	.004974754	Prob > F	=	0.0000
Total	17.4235117	30	.580783723	R-squared	=	0.9920
				Adj R-squared	=	0.9914
				Root MSE	=	.07053

LIMPOR	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
LIPC	-.8733684	.2848051	-3.07	0.005	-.1.456765 -.2899715
LPIB	1.850098	.1829115	10.11	0.000	1.475421 2.224775
_cons	1.409417	.2700739	5.22	0.000	.8561958 1.962638
- Variables View:** The right pane shows a list of variables with their names and labels. The variable `LPIB` is selected.
- Properties View:** Below the Variables view, the properties of the selected variable `LPIB` are displayed, including its type (float) and format (%9.0g).

Según la matriz de correlación, existe una alta relación entre las variables independientes, por lo que existe un problema de multicolinealidad.

Ahora vamos a aplicar otra prueba conocida como el Factor Inflador de la Varianza (FIV), esta permite ver cómo se incrementa la varianza a medida que las variables se relacionan más entre sí. Ahora lo haremos en STATA, utilizamos el comando **vif**, lo que nos da el siguiente resultado:

The screenshot shows the Stata SE 14.0 interface with the following details:

- Menu Bar:** Archivo, Edición, Datos, Gráficos, Estadísticas, Usuario, Ventana, Ayuda.
- Command History:** Shows a series of commands run in the command window, including regressions and correlation checks.
- Output Window:** Displays the results of a regression model and a VIF analysis.
- Variables View:** A list of variables in the dataset, with LPIB selected.
- Properties View:** Shows the properties of the selected variable LPIB, including its name, type (float), and format (%9.0g).
- Selected Data:** A table showing the VIF values for LIPC and LPIB, and the mean VIF.

Specific output from the VIF analysis:

Variable	VIF	1/VIF
LIPC	69.31	0.014427
LPIB	69.31	0.014427
Mean VIF	69.31	

El FIV nos indica que las variables tienen un valor de 69, lo que indica que hay un problema de multicolinealidad, ya que cualquier FIV superior a 10 es señal de multicolinealidad.

Una vez que detectamos el problema de MULTICOLINEALIDAD, debemos corregirlo; para esto revise el capítulo 10 de su libro básico y determine que métodos de corrección existen.