



# UTPL

La Universidad Católica de Loja

Vicerrectorado de Modalidad Abierta y a Distancia

## Econometría Básica

Guía didáctica





**Facultad Ciencias Económicas y Empresariales**

## Econometría Básica

### Guía didáctica

Carrera	PAO Nivel
Economía	IV

#### Autor:

Oswaldo Francisco Ochoa Ordóñez



**Universidad Técnica Particular de Loja**

## **Econometría Básica**

### **Guía didáctica**

Oswaldo Francisco Ochoa Ordóñez

### **Diagramación y diseño digital**

Ediloja Cía. Ltda.

Marcelino Champagnat s/n y París

edilocialtda@ediloja.com.ec

[www.ediloja.com.ec](http://www.ediloja.com.ec)

**ISBN digital** - 978-9942-39-903-8

**Año de edición:** octubre, 2023

**Edición:** primera edición reestructurada en febrero 2025 (con un cambio del 5%)

Loja-Ecuador



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 licenciante. *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

<b>1. Datos de información .....</b>	<b>9</b>
1.1 Presentación de la asignatura.....	9
1.2 Competencias genéricas de la UTPL.....	9
1.3 Competencias del perfil profesional .....	9
1.4 Problemática que aborda la asignatura .....	10
<b>2. Metodología de aprendizaje .....</b>	<b>11</b>
<b>3. Orientaciones didácticas por resultados de aprendizaje.....</b>	<b>13</b>
<b>Primer bimestre .....</b>	<b>13</b>
<b>Resultados de aprendizaje 1 y 2: .....</b>	<b>13</b>
<b>Contenidos, recursos y actividades de aprendizaje recomendadas.....</b>	<b>13</b>
<b>Semana 1 .....</b>	<b>14</b>
Unidad 1. Análisis de regresión con datos de corte transversal.....	14
1.1 ¿Qué es la econometría? .....	14
1.2 Pasos en un análisis económico empírico.....	15
1.3 Estructura de los datos .....	15
1.4 Causalidad y la noción ceteris paribus .....	16
Actividades de aprendizaje recomendadas .....	17
Autoevaluación 1 .....	18
<b>Contenidos, recursos y actividades de aprendizaje recomendadas.....</b>	<b>20</b>
<b>Semana 2 .....</b>	<b>20</b>
Unidad 2. Regresión lineal con un regresor .....	20
2.1 El modelo de regresión lineal .....	20
2.2 Estimación de los coeficientes del modelo de regresión lineal .....	23
2.3 Medida de ajuste.....	27
Actividades de aprendizaje recomendadas .....	29
<b>Contenidos, recursos y actividades de aprendizaje recomendadas.....</b>	<b>30</b>
<b>Semana 3 .....</b>	<b>30</b>
Unidad 2. Regresión lineal con un regresor .....	30



2.4 Los supuestos de mínimos cuadrados.....	30
2.5 Distribución muestral de los estimadores de MCO .....	30
2.6 Conclusión.....	31
Actividades de aprendizaje recomendadas .....	32
<b>Contenidos, recursos y actividades de aprendizaje recomendadas.....</b>	<b>33</b>
<b>Semana 4.....</b>	<b>33</b>
Unidad 2. Regresión lineal con un regresor.....	33
2.7 Unidades de medición y forma funcional.....	33
2.8 Valores esperados y varianzas de los estimadores de MCO .....	35
2.9 Regresión a través del origen .....	37
Actividades de aprendizaje recomendadas .....	37
<b>Contenidos, recursos y actividades de aprendizaje recomendadas.....</b>	<b>38</b>
<b>Semana 5.....</b>	<b>38</b>
Unidad 2. Regresión lineal con un regresor.....	38
2.10 Contraste de hipótesis acerca de uno de los coeficientes de regresión .....	38
2.11 Intervalos de confianza para coeficiente de regresión.....	40
Actividades de aprendizaje recomendadas .....	41
<b>Contenidos, recursos y actividades de aprendizaje recomendadas.....</b>	<b>42</b>
<b>Semana 6.....</b>	<b>42</b>
Unidad 2. Regresión lineal con un regresor.....	42
2.12 Heterocedasticidad y homocedasticidad.....	42
2.13 Fundamentos teóricos de mínimos cuadrados ordinarios .....	44
2.14 La utilización del estadístico t en regresión para muestras pequeñas .....	46
Actividades de aprendizaje recomendadas .....	46
Autoevaluación 2.....	47
<b>Resultado de aprendizaje 3: .....</b>	<b>49</b>
<b>Contenidos, recursos y actividades de aprendizaje recomendadas.....</b>	<b>49</b>



**Semana 7** ..... 49

    Unidad 3. Análisis de regresión múltiple: estimación ..... 49

        3.1 Motivación para la regresión múltiple ..... 49

        3.2 Mecánica e interpretación de los mínimos cuadrados ordinarios..... 50

        3.3 Valor esperado de los estimadores de MCO ..... 53

        3.4 Varianza de los estimadores de MCO..... 55

        3.5 Eficiencia de MCO: el teorema de GAUSS – MARKOV ..... 56

    Actividades de aprendizaje recomendadas ..... 57

    Autoevaluación 3..... 58

**Resultados de aprendizaje 1 a 3:**..... 60

**Contenidos, recursos y actividades de aprendizaje recomendadas**..... 60

**Semana 8** ..... 60

    Actividades finales del bimestre ..... 60

    Actividades de aprendizaje recomendadas ..... 60

**Segundo bimestre**..... 62

**Resultado de aprendizaje 3:** ..... 62

**Contenidos, recursos y actividades de aprendizaje recomendadas**..... 62

**Semana 9** ..... 62

    Unidad 4. Análisis de regresión múltiple: inferencia ..... 62

        4.1 Distribución de muestreo de los estimadores de MCO ..... 62

        4.2 Prueba de hipótesis para un solo parámetro poblacional: la prueba t  
            ..... 64

        4.3 Intervalos de confianza..... 64

    Actividades de aprendizaje recomendadas ..... 65

**Contenidos, recursos y actividades de aprendizaje recomendadas**..... 66

**Semana 10** ..... 66

    Unidad 4. Análisis de regresión múltiple: inferencia ..... 66

        4.4 Prueba de hipótesis de una sola combinación lineal de los parámetros  
            ..... 66



4.5 Pruebas de restricciones lineales múltiples: la prueba F ..... 68

4.6 Informe de resultados..... 69

Actividades de aprendizaje recomendadas ..... 70

**Contenidos, recursos y actividades de aprendizaje recomendadas..... 71**

**Semana 11 ..... 71**

Unidad 4. Análisis de regresión múltiple: inferencia ..... 71

4.7 Consistencia..... 71

4.8 Normalidad asintótica e inferencia con muestras grandes ..... 73

4.9 Eficiencia asintótica de MCO ..... 76

Actividades de aprendizaje recomendadas ..... 76

**Contenidos, recursos y actividades de aprendizaje recomendadas..... 77**

**Semana 12..... 77**

Unidad 4. Análisis de regresión múltiple: inferencia ..... 77

4.10 Efectos del escalamiento de datos sobre los estadísticos de MCO 77

4.11 Más acerca de la forma funcional ..... 78

4.12 Más sobre la bondad de ajuste y selección de los regresores ..... 80

Actividades de aprendizaje recomendadas ..... 82

Autoevaluación 4..... 83

**Contenidos, recursos y actividades de aprendizaje recomendadas..... 85**

**Semana 13..... 85**

Unidad 5. Análisis de regresión múltiple con información cualitativa:  
variables binarias (o dummies)..... 85

5.1 Descripción de la información cualitativa ..... 85

5.2 Una sola variable binaria independiente..... 86

5.3 Uso de las variables binarias en categorías múltiples ..... 87

Actividades de aprendizaje recomendadas ..... 89

**Contenidos, recursos y actividades de aprendizaje recomendadas..... 90**

**Semana 14..... 90**



Unidad 5. Análisis de regresión múltiple con información cualitativa: variables binarias (o dummies) .....	90
5.4 Interacción en las que intervienen variables binarias .....	90
5.5 Más acerca del análisis de políticas y evaluación de programas .....	91
Actividades de aprendizaje recomendadas .....	92
Autoevaluación 5 .....	93
<b>Contenidos, recursos y actividades de aprendizaje recomendadas .....</b>	<b>94</b>
<b>Semana 15 .....</b>	<b>94</b>
Unidad 6. Multicolinealidad .....	94
6.1 Naturaleza de la multicolinealidad .....	94
6.2 Consecuencias prácticas de la multicolinealidad .....	94
6.3 Detección de la multicolinealidad .....	95
Actividades de aprendizaje recomendadas .....	102
Autoevaluación 6 .....	103
<b>Contenidos, recursos y actividades de aprendizaje recomendadas .....</b>	<b>105</b>
<b>Semana 16 .....</b>	<b>105</b>
Actividades finales del bimestre .....	105
Actividades de aprendizaje recomendadas .....	106
<b>4. Autoevaluaciones .....</b>	<b>107</b>
<b>5. Referencias bibliográficas .....</b>	<b>113</b>
<b>6. Anexos .....</b>	<b>114</b>





## 1. Datos de información

### 1.1 Presentación de la asignatura



### 1.2 Competencias genéricas de la UTPL

- Capacidad de abstracción, análisis y síntesis.
- Capacidad de aplicar los conocimientos en la práctica.
- 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 para identificar, plantear y resolver problemas.
- Capacidad para tomar decisiones.
- Habilidades interpersonales.

### 1.3 Competencias del perfil profesional

Aportar a los procesos productivos de los sectores estratégicos mediante la investigación sobre las actividades económicas en un contexto regional y nacional.



## 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 libro, así como también vaya desarrollando otros ejercicios que se encuentran propuestos en el texto.

### Metodología de aprendizaje por descubrimiento

[La metodología de aprendizaje por descubrimiento](#) es un enfoque educativo que busca promover el aprendizaje activo y significativo a través de la exploración y el descubrimiento por parte del estudiante. En lugar de transmitir información de manera pasiva, el aprendizaje por descubrimiento involucra a los estudiantes en el proceso de búsqueda, análisis y comprensión de conceptos y principios por sí mismos.



A diferencia del aprendizaje tradicional, en el que el docente es el principal transmisor del conocimiento, en el aprendizaje por descubrimiento, el rol del educador es el de facilitador o guía. Los estudiantes son animados a formular preguntas, plantear hipótesis, realizar experimentos y descubrir soluciones o respuestas por sí mismos.

### **Metodología de aprendizaje basado en problemas**

[La metodología de Aprendizaje Basado en Problemas](#) (ABP) es un enfoque educativo que se centra en el estudiante y su participación activa en la resolución de problemas reales o situaciones complejas. No se centra únicamente en la transmisión de conocimientos teóricos, el ABP promueve el aprendizaje a través de la investigación y la solución de problemas prácticos que los estudiantes pueden encontrar en su vida profesional o en el mundo real.





### 3. Orientaciones didácticas por resultados de aprendizaje



#### Primer bimestre

##### Resultados de aprendizaje 1 y 2:

- Analiza la relación lineal entre dos variables y predice la dependencia de una con respecto a la otra.
- Comprende la utilidad de un análisis de regresión lineal simple.

Para lograr los resultados 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 haya logrado comprender cada uno de los elementos conceptuales, podrá entender a qué se refieren los modelos de regresión, su utilidad y estructura.

#### Contenidos, recursos y actividades de aprendizaje recomendadas

Recuerde revisar de manera paralela los contenidos con las actividades de aprendizaje recomendadas y actividades de aprendizaje evaluadas.





### Unidad 1. Análisis de regresión con datos de corte transversal

#### 1.1 ¿Qué es la econometría?

Iniciamos el estudio de la econometría, destacando algunos conceptos básicos, que nos permitan entender qué es la econometría:

Para Wooldridge, (2009), “La Econometría se basa en métodos estadísticos para estimar relaciones económicas, probar teorías económicas y evaluar e implementar políticas públicas y de negocios”.

En esa misma línea, Goldberger, (1964) menciona que “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”.

Para Stock & Watson, (2012), “La econometría es la ciencia y el arte de utilizar la teoría económica y las técnicas estadísticas para analizar los datos económicos”.

En definitiva, la econometría se basa en la explicación cuantitativa de los fenómenos económicos, financieros y sociales de un país, a través del uso de las matemáticas, la estadística, la macro y microeconomía; en fin, construir modelos económicos que pueden ser probados y estimados utilizando datos observados. Estos modelos pueden incluir variables económicas, sociales y demográficas, y permiten estudiar cómo estas variables se relacionan entre sí y cómo influyen en los resultados económicos. Además, los modelos econométricos también son utilizados en las ciencias sociales, políticas y sociológicas.



## 1.2 Pasos en un análisis económico empírico

Los modelos econométricos son utilizados para probar alguna teoría económica o para expresar alguna relación que tiene una importancia para la toma de decisiones, ya sea de negocios o para el tema de políticas. Es por esto es por lo que debemos seguir una serie de pasos generales que nos permitan estructurar un análisis económico o de cualquier índole en forma empírica.

En la siguiente infografía se indican los pasos generales que permiten estructurar un modelo de regresión.

### [Análisis económico empírico](#)

## 1.3 Estructura de los datos

**Tabla 1**

*Tipos de información*

### Datos de corte transversal

- Contiene una muestra de datos sea de: individuos, empresas, hogares, cantones, provincias, países, etc, tomada en un punto específico del tiempo.
- Datos muy utilizados en la economía (microeconomía aplicada). Por ejemplo: economía laboral, economía urbana, finanzas públicas entre otras.
- Algunas tablas del **texto básico** que representan esta información son: tabla 1.1.
- A nivel país están los datos de la encuesta **ENEMDU**
- Datos también utilizados en las ciencias sociales.

### Datos de series de tiempo

- Es información de una o varias variables a lo largo del tiempo.
- Información de serie temporal puede ser: por horas, diarias, semanal, quincenal, mensual, bimestral, trimestral, semestral, anual, etc.
- Ejemplo de variables tenemos: PIB, inflación, tasas de interés, precios de las acciones, tasa anual de homicidios, etc.



### Datos de corte transversal

- Son un poco más difíciles de analizar, ya que no se puede mantener la independencia de las observaciones a lo largo del tiempo, por ende pueden estar altamente relacionadas con sus valores pasados.
- alguna tabla con este tipo de información es la tabla 1.3.
- El **Banco Mundial** posee datos de series temporales de muchas variables de carácter macroeconómico.

### Datos de panel o longitudinales

- Es una combinación de una serie de tiempo por cada unidad de una base de datos de corte transversal.
- Por ejemplo, la variación del salario promedio de los empleados públicos en cada provincia desde el 2010 al 2020. Estos datos incluyen individuos o provincias como información de corte transversal y también un periodo temporal de datos.
- Las tablas 1.4 y 1.5 son ejemplos de este tipo de información de panel.
- Como datos de panel tenemos los datos de la **CEPALSTAT**, que posee información de los países de América Latina.

---

Nota. Adaptado de Introducción a la Econometría (p. 5), por Wooldridge, J., 2010, Cengage Learning. México.

La tabla 1, muestra los distintos tipos de información para estructurar un modelo de regresión, además se citan algunos ejemplos de tipo de información que se adapta a estos tipos de información.

## 1.4 Causalidad y la noción ceteris paribus

La causalidad en econometría se refiere a la relación de causa y efecto entre variables económicas, mientras que la noción de “ceteris paribus” se utiliza para analizar el efecto de una variable específica manteniendo constantes todas las demás variables relevantes. Estos conceptos son fundamentales en la investigación económica y en la estimación de modelos econométricos.



Para aplicar el concepto de “ceteris paribus” en econometría, se utilizan técnicas estadísticas y modelos econométricos. Estos modelos intentan controlar otras variables que podrían afectar el resultado y evaluar cómo una variable en particular influye en otra. Para entender mejor estos conceptos, revise los ejemplos 1.3, 1.4 y 1.5, del libro de Wooldridge pág. 13.

Para entender mejor los temas de esta semana, se recomienda, revisar el capítulo 1 del libro de Wooldridge, de la página 1 a la 19.



## Actividades de aprendizaje recomendadas

Es momento de aplicar sus conocimientos a través de las actividades que se han planteado a continuación:

1. Revise las orientaciones desarrolladas por el docente en el aula virtual y participe en la tutoría permanente.

**Procedimiento:** para cada una de las semanas, su docente tutor realizará orientaciones sobre el tema para acercar con mayor claridad los temas expuestos tanto en la guía virtualizada como en el libro. Allí su docente tutor ubicará ejemplos demostrativos para cada uno de los temas que se desarrollan en la semana.

2. Revise los conceptos básicos y metodología econométrica propuesta en el libro de Wooldridge.

**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. Por favor, complete la actividad en un cuaderno o documento Word.*

3. Revise los ejercicios del 2.3 al 2.5 de la página 32 del libro de Wooldridge.



**Procedimiento:** en esta actividad usted entenderá cómo es la interpretación básica de los resultados de un modelo de regresión simple.

4. Desarrolle la autoevaluación 1 y las actividades recomendadas en la guía virtualizada.

**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.

La autoevaluación no es obligatoria, sin embargo, le ayudará a reforzar sus conocimientos y lo prepara para la evaluación bimestral. ¡Vamos a trabajar!



### Autoevaluación 1

De acuerdo a la información analizada, lea comprensivamente cada uno de los ítems, cuando la pregunta sea dicotómica escriba la respuesta (v/f) y en las preguntas de opción múltiple elija la respuesta correcta.

1. ¿Qué es la econometría?
  - a. Es la obtención de datos cualitativos y cuantitativos.
  - b. Es la obtención del valor medio de las variables independientes.
  - c. Es el uso de métodos estadísticos para estimar relaciones económicas.
2. ¿Qué son los datos no experimentales?
  - a. Datos secundarios cuantitativos.
  - b. Datos obtenidos de forma controlada.
  - c. Datos cualitativos secundarios.



3. El primer paso de un análisis económico empírico es:
- a. Formular la pregunta de interés.
  - b. Calcular las variables.
  - c. Analizar los resultados del modelo.
4. ( ) ¿Una de las variables del modelo de comportamiento delictivo es la probabilidad de ser atrapado?
5. ( ) ¿Cuándo se precisa el modelo teórico hay que transformarlo a un modelo matemático?
6. Los datos de corte transversal son:
- a. Datos que se generan en un mismo periodo de tiempo, pero en diversos individuos.
  - b. Datos en distintos periodos de tiempo.
  - c. Datos que combinan series de tiempo y corte transversal.
7. Un ejemplo de información de corte transversal es:
- a. PIB per cápita desde 1970 a 2022.
  - b. El VAB de las 24 provincias en el año 2022.
  - c. La inflación mensual desde 2020 a 2022.
8. La información de datos de panel es:
- a. Información temporal.
  - b. Información de una serie de tiempo por cada unidad de una base de corte transversal.
  - c. Información de corte transversal.
9. Se tiene datos de la oferta monetaria desde 1980 hasta el 2010, ¿qué tipo de información es?
- a. Serie temporal.
  - b. Panel.
  - c. Corte transversal.



10. El concepto "ceteris paribus" significa:

- a. Dejar hacer dejar pasar.
- b. Si todos los demás factores relevantes permanecen constantes.
- c. Relación simple.

[Ir al solucionario](#)

## Contenidos, recursos y actividades de aprendizaje recomendadas



### Semana 2

## Unidad 2. Regresión lineal con un regresor

### 2.1 El modelo de regresión lineal

Es interesante que podamos conocer ¿cómo el endurecimiento de infracciones de tránsito sobre conductores ebrios puede disminuir la tasa de mortalidad de los accidentes de tránsito? Con esta interrogante, lo que se quiere conocer son los efectos desconocidos del cambio de una variable (X), siendo esta el endurecimiento de multas, sobre una variable (Y).

Está la mortalidad por accidentes de tránsito. Lo que se busca es establecer una relación lineal entre dos variables, como es el caso de Y y X. Entonces, el problema econométrico consiste en estimar esta pendiente, es decir, el efecto unitario de la variable X sobre la variable Y, utilizando un conjunto de datos para cada una de las variables involucradas. Revisemos un ejemplo que nos permita entender mejor esta relación.

Queremos conocer cuál sería el efecto de la inflación sobre el Producto Interno Bruto (PIB) en Ecuador, es decir, qué sucede con el PIB, ante una variación de la inflación, supongamos de un 1 %. Podemos expresar esto en forma de relación matemática de la siguiente forma:



La letra griega  $\beta_{inflaci3n}$  distingue el efecto de la variaci3n en la inflaci3n de otros posibles efectos. Por tanto.

$$\beta_{inflaci3n} = \frac{\text{variaci3n del PIB}}{\text{variaci3n de la inflaci3n}} = \frac{\Delta PIB}{\Delta inflaci3n} \quad (1)$$

Donde la letra griega  $\Delta$  (delta) significa "variaci3n de". Es decir,  $\beta_{inflaci3n}$  es la variaci3n en el PIB que provoca la variaci3n en la inflaci3n dividida por la variaci3n en la inflaci3n. Si conoci3ramos el valor de  $\beta_{inflaci3n}$ , ser3a posible conocer c3mo una variaci3n en la inflaci3n, afectar3a al PIB. Si reorganizamos la ecuaci3n (1), tendr3amos ahora:

$$\Delta PIB = \beta_{inflaci3n} * \Delta inflaci3n \quad (2)$$

Supongamos que  $\beta_{inflaci3n} = -0.6$ . Entonces, un aumento del 1 % en la inflaci3n, dar3a lugar a una variaci3n en el PIB esperado de  $(-0.6) * (1) = -0.6$ , es decir, se podr3a esperar que el PIB disminuya en 0.6 miles de millones, como consecuencia del aumento de la inflaci3n en un 1 %.

La ecuaci3n (1) es la definici3n de la pendiente de una l3nea recta que relaciona al PIB con la inflaci3n. Esta recta se puede escribir como:

$$PIB = \beta_0 + \beta_{inflaci3n} \cdot inflaci3n, \quad (3)$$

Donde  $\beta_0$  es el intercepto o constante de esta recta y  $\beta_{inflaci3n}$  es la pendiente. De acuerdo con la ecuaci3n (3). Conociendo  $\beta_0$  y  $\beta_1$ , tambi3n podr3amos predecir el promedio del PIB de cualquier a3o, dado el valor de la inflaci3n de cualquier a3o.

Pero entendemos que el PIB no solo puede verse influenciado por la inflaci3n, sino que pueden existir otras variables o factores que lo afecten, que una vez determinados podr3an incluirse en la ecuaci3n (3). Por ahora tendr3amos:

$$PIB = \beta_0 + \beta_{inflaci3n} \cdot inflaci3n + \text{otros factores o variables} \quad (4)$$

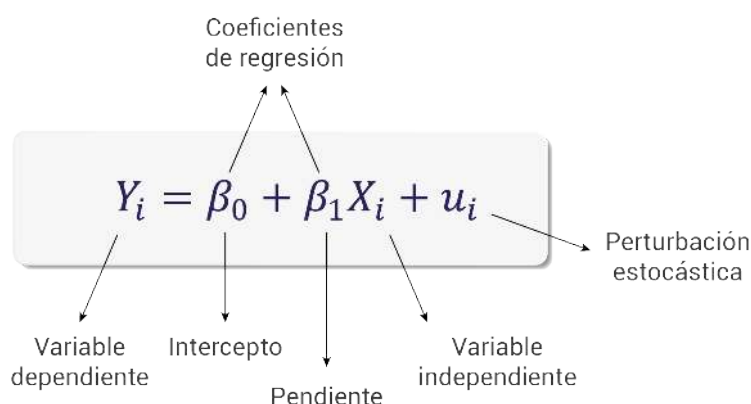


Con base en la teoría económica, podríamos decir que factores o variables como: inversión extranjera directa, oferta monetaria, deuda externa, entre otras, podrían ser incluidas como variables que expliquen cambios en el PIB.

Entonces, la ecuación (4), se puede expresar de forma general de la siguiente forma, dependiendo de las variables que escojamos para establecer una relación funcional.

**Figura 1**

*Estructura de un modelo de regresión simple*



*Nota. Adaptado de Guía didáctica de Econometría Básica (p. 2), por Ochoa, O., 2021, México: McGraw Hill, CC BY 4.0.*

La ecuación de la figura 1, es el modelo de regresión lineal con regresor único, la cual indica los componentes de esta ecuación. A esta ecuación también se la conoce como la **Función de Regresión Poblacional (FRP)**.

**El término** , se conoce como término de error, recoge el efecto de otros factores o variables sobre la variable dependiente, pero que no fueron incluidas en el modelo de regresión.

Con base a lo anterior, la regresión lineal es un método estadístico utilizado para modelar la relación entre una variable dependiente y una o más variables independientes. Se basa en la suposición de que **existe una relación lineal** entre las variables involucradas y lo que se busca es encontrar la mejor línea recta que se ajuste a un conjunto de datos, la cual se conoce como línea de

regresión o línea de mejor ajuste. El objetivo de la regresión lineal es minimizar la diferencia entre los valores reales observados y los valores predichos por el modelo lineal de regresión.

## 2.2 Estimación de los coeficientes del modelo de regresión lineal

Al estructurar un modelo de regresión a través de las dos variables seleccionadas, lo que sigue a continuación es calcular los parámetros de la ecuación, es decir, calcular los valores del intercepto y la pendiente poblacional desconocidas de la recta de regresión, pero como no conocemos a la población, debemos tomar una muestra de esa población y obtener los datos de las variables seleccionadas. Para ejemplificar esto tomemos datos del PIB y la inflación del Ecuador desde el año 1970 al 2019, ver [anexo 1](#).

Antes es importante conocer los estadísticos descriptivos de las variables antes mencionadas. Para todos los procesos estadísticos y econométricos vamos a utilizar el programa STATA en cualquiera de sus versiones a partir de la décima versión.

A continuación, lo invito a examinar la infografía

### [Programa STATA](#)

La cual proporciona una visión detallada sobre cómo utilizar el programa STATA en nuestras investigaciones.

### **El estimador de Mínimos Cuadrados Ordinarios (MCO)**

El estimador MCO elige los coeficientes de regresión de tal forma que la recta de regresión estimada se encuentre lo más cercana posible a los datos observados, y la cercanía está medida por la suma de los errores al cuadrado que se cometen con la predicción de Y dado X.



El estimador MCO extiende esta idea al modelo de regresión lineal. Sean  $b_0$  y  $b_1$  algunos de los estimadores de  $b_0$  y  $b_1$ . La recta de regresión basada en esos estimadores es:  $b_0 + b_1$ , por lo que el valor de  $Y_i$  previsto mediante esta recta es:

$$b_0 + b_1 X_i \quad (5)$$

Entonces el error cometido en la predicción de la observación de  $Y_i$  es:

$$Y_i - b_0 - b_1 X_i \quad (6)$$

La suma de estos errores al cuadrado para este error es:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{s_{XY}}{s_X^2} \quad (7)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad (8)$$

El valor de predicción MCO  $\hat{Y}_i$  y los residuos  $\hat{u}_i$  son:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i, \quad i = 1, \dots, n \quad (9)$$

$$\hat{u}_i = Y_i - \hat{Y}_i, \quad i = 1, \dots, n \quad (10)$$

Aplicando los MCO a la regresión con las variables del PIB y la inflación, por medio de STATA, tenemos lo siguiente:



**Figura 2***Regresión simple por MCO en STATA*

```
. reg PIBUSapreciosactuales Inflaciónpreciosalconsumidor
```

Source	SS	df	MS	Number of obs	=	52
Model	1.6587e+22	1	1.6587e+22	F(1, 50)	=	18.40
Residual	4.5067e+22	50	9.0133e+20	Prob > F	=	0.0001
				R-squared	=	0.2690
				Adj R-squared	=	0.2544
Total	6.1653e+22	51	1.2089e+21	Root MSE	=	3.0e+10

PIBUSapreciosactuales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Inflaciónpreciosalconsumidor	-8.44e+08	1.97e+08	-4.29	0.000	-1.24e+09	-4.49e+08
_cons	5.59e+10	5.76e+09	9.71	0.000	4.43e+10	6.75e+10

*Nota. Ochoa, O., 2023.*

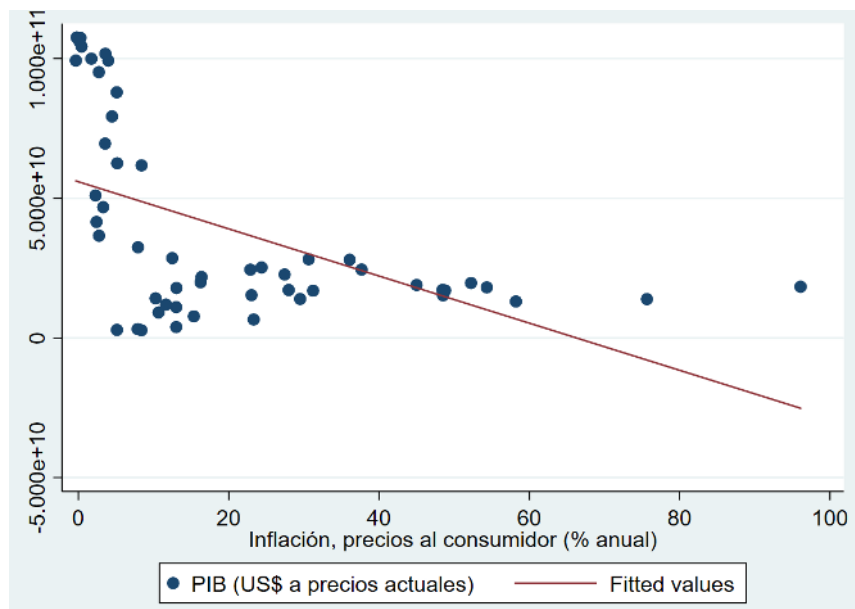
Los comandos utilizados para hacer esta regresión son los que están resaltados en celeste en la figura anterior. Por ende, los resultados de la estimación son:

$$\hat{PIB} = 5.59_{e+10} - 8.44_{e+08} \cdot \text{inflación} + u_i \quad (11)$$

El “sombrero”, encima de la palabra PIB, indica, que lo que se estima es el valor esperado del PIB, ante variaciones en la inflación. En la figura 3 se representa la recta de regresión MCO, ajustada a los datos del diagrama de dispersión de las variables utilizadas.

**Figura 3**

*Recta de regresión para los datos del PIB e inflación*



*Nota. Ochoa, O., 2023.*

Como se puede observar en la ecuación (4), el valor negativo de la pendiente indica una relación indirecta, ya que una variación en la inflación provoca un cambio negativo promedio en el PIB de \$ -8.44e+8, es decir, a mayor inflación, menor PIB. Lo que se está estimando aquí es un cambio promedio del PIB a lo largo del periodo temporal, pero con estos datos también se puede estimar un valor puntual, por ejemplo:

Queremos saber cuál es el valor promedio del PIB en el año 2015, ante un cambio de un 1 % en la inflación. Lo que debemos hacer es reemplazar en la ecuación estimada el valor de X, es decir, de la variable inflación por el valor de X en el año 2015 = 3.96 %.

$$\begin{aligned} \hat{PIB} &= 5.59_{e+10} - 8.44_{e+08} \cdot (3.96) + u_i & (12) \\ \hat{PIB} &= 5.26_{e+10} \end{aligned}$$

Se supone que al utilizar la MCO, los estimadores ( $\beta_0$  y  $\beta_1$ ) tienen propiedades teóricas deseables, es decir, son insesgados e inconsistentes.

## 2.3 Medida de ajuste

Ahora debemos preguntarnos que tan bien describe esta regresión a las variables utilizadas. ¿Qué tanto explica la variable X a la variable Y?, ¿Están las observaciones cerca de la recta o de regresión o muy dispersas?, esto lo sabremos a través del coeficiente de determinación  $R^2$  y de los errores estándar respectivamente, las cuales en conjunto miden la bondad de ajuste de la recta de regresión MCO a los datos.

El  $R^2$  es un valor que oscila entre 0 y 1, y mide la proporción de la varianza de  $Y_i$  explicada por  $X_i$ , mientras que el error estándar mide la distancia que habitualmente separa a  $Y_i$  de su valor esperado.

Matemáticamente, el  $R^2$  puede escribirse como el cociente entre la suma explicada (de cuadrados) y la suma total (de cuadrados). La Suma Explicada (SE) es la suma de las desviaciones al cuadrado de los valores de predicción de  $Y_i, \hat{Y}_i$ , respecto de su media, y la Suma Total (ST) es la suma de los cuadrados de las desviaciones de  $Y_i$  respecto de su media:

### Cálculo de $R^2$

$$SE = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \quad (12)$$

$$ST = \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (13)$$

$$SR = \sum_{i=1}^n \hat{u}_i^2 \quad (14)$$

$$R^2 = \frac{SE}{ST} \quad (15)$$

$$R^2 = 1 - \frac{SR}{ST} \quad (16)$$

Cualquiera de las dos fórmulas nos permite calcular el coeficiente de determinación.



### Cálculo del Error Estándar de la Regresión (ESR)

Este error es un estimador de la desviación típica del error de regresión  $u_i$ . Las unidades de  $u_i$  e  $Y_i$  son las mismas, por lo que el ESR es una medida de la dispersión de las observaciones en torno a la recta de regresión, medida en las unidades de la variable dependiente.

$$ESR = S_{\hat{u}} \quad \text{donde} \quad S_{\hat{u}}^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{u}_i^2 = \frac{SR}{n-2} \quad (18)$$

Para discernir si los resultados de mi regresión son confiables, los errores estándar deben ser menores a la tercera parte del valor de los parámetros B0 y B1.

En el programa STATA estos valores en el cuadro de resultados se presentan así:

**Figura 4**  
*Valores del R2 y los errores estándar*

```
. reg PIBUSapreciosactuales Inflaciónpreciosalconsumidor
```

Source	SS	df	MS	Number of obs	=	52
Model	1.6587e+22	1	1.6587e+22	F(1, 50)	=	18.40
Residual	4.5067e+22	50	9.0133e+20	Prob > F	=	0.0001
				R-squared	=	0.2690
				Adj R-squared	=	0.2544
Total	6.1653e+22	51	1.2089e+21	Root MSE	=	3.0e+10

PIBUSapreciosactuales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Inflaciónpreciosalconsumidor	-8.44e+08	1.97e+08	-4.29	0.000	-1.24e+09	-4.49e+08
_cons	5.59e+10	5.76e+09	9.71	0.000	4.43e+10	6.75e+10

Nota. Ochoa, O., 2023.

Para entender los temas revisados lo invito a revisar el capítulo 1 del libro de Wooldridge y el capítulo 1 del libro de Meza.



## Actividades de aprendizaje recomendadas

Es hora de reforzar los conocimientos adquiridos resolviendo las siguientes actividades:

1. Revise los ejercicios del 2.8 al 2.11 propuestos en el libro de Wooldridge, en la página 39.

**Procedimiento:** estos ejercicios le permitirán conocer cómo se interpretan los coeficientes de regresión y la interpretación del coeficiente de determinación  $R^2$ .

2. Revise las orientaciones desarrolladas por el docente en el aula virtual y participe en la tutoría permanente.

**Procedimiento:** para cada una de las semanas, su docente tutor realizará orientaciones sobre el tema para acercar con mayor claridad los temas expuestos tanto en la guía virtualizada como en el libro. Allí su docente tutor ubicará ejemplos demostrativos para cada uno de los temas que se desarrollan en la semana.

3. Reflexione sobre los temas de la unidad 2 trabajados en la guía didáctica y en los temas introductorios del capítulo 1 del libro de Wooldridge y capítulo 1 del libro de Meza.

**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. Por favor, complete las actividades en un cuaderno o documento Word.*





## Unidad 2. Regresión lineal con un regresor

### 2.4 Los supuestos de mínimos cuadrados

Los supuestos que se describen en la siguiente infografía son los más aplicados a la metodología de Mínimos Cuadrados Ordinarios (MCO).

#### [Los supuestos mínimos cuadrados](#)

Estos supuestos son esenciales para asegurar que los estimadores de mínimos cuadrados sean insesgados y eficientes, y para realizar inferencias válidas en el marco de la econometría. Sin embargo, es importante tener en cuenta que estos supuestos pueden no ser necesarios o cumplirse estrictamente en todos los casos de aplicación empírica.

### 2.5 Distribución muestral de los estimadores de MCO

Debido a que los estimadores MCO  $\hat{\beta}_0, \hat{\beta}_1$  se calculan a partir de una muestra seleccionada aleatoriamente, los estimadores en sí mismos son variables aleatorias con una distribución de probabilidad “la distribución muestral” que describe los valores que podrían tomar en las diferentes muestras aleatorias posibles.

Esta sección presenta estas distribuciones muestrales. En muestras pequeñas, estas distribuciones son complicadas, pero en muestras grandes, son aproximadamente normales por el teorema central del límite.

*La distribución muestral de los estimadores MCO.*

La distribución muestral de  $\underline{y}$ , la cual es calculada a través de una muestra de valores aleatorios. La probabilidad de estos valores es recogida por su distribución muestral, cuya propiedad más importante es:  $E(\underline{y}) = \mu_y$ ,



Siendo  $\underline{y}$  un estimador insesgado de  $\underline{u}_y$ , por lo que si la muestra (n) es más grande, esta distribución tiende a aproximarse a su distribución normal según el teorema del límite central.

La distribución de  $\hat{\beta}_0, \hat{\beta}_1$ , esto se transfiere a los estimadores de MCO, donde  $B_0$  es la constante o intercepto y  $B_1$  la pendiente. Estos también, al provenir de una muestra, la probabilidad de los diferentes valores se resume en sus distribuciones muestrales.

Los valores de  $B_0$  y  $B_1$  bajo los supuestos de MCO son:

$$E(\hat{\beta}_0) = \beta_0 \quad y \quad E(\hat{\beta}_1) = \beta_1 \quad (19)$$

Es decir,  $\hat{\beta}_0$  y  $\hat{\beta}_1$  son estimadores insesgados de  $B_0$  y  $B_1$ .

## 2.6 Conclusión

En este capítulo se ha abordado el uso de los Mínimos Cuadrados Ordinarios (MCO) para estimar el término independiente y la pendiente de una línea de regresión poblacional mediante una muestra de observaciones de una variable dependiente, Y, y un único regresor, X. La utilización de MCO para trazar una línea recta en un diagrama de dispersión tiene varias ventajas si se cumplen los supuestos de mínimos cuadrados.

Los estimadores MCO de la pendiente y el término independiente son insesgados, consistentes y su distribución muestral tiene una varianza inversamente proporcional al tamaño de la muestra, n. Además, si n es grande, la distribución muestral del estimador MCO se aproxima a una distribución normal.

Estas propiedades se cumplen bajo tres supuestos de mínimos cuadrados. El primer supuesto establece que el término de error del modelo de regresión lineal tiene una media condicional igual a cero dado el regresor X, lo que implica que el estimador MCO es insesgado. El segundo supuesto establece que las observaciones (Xi, Yi) son independientes e idénticamente distribuidas,



como en el caso del muestreo aleatorio simple, lo que permite obtener una fórmula para la varianza de la distribución muestral del estimador MCO. El tercer supuesto señala que los valores atípicos grandes son poco probables y que las variables  $X$  y  $Y$  tienen momentos de cuarto orden finitos (curtosis finita). Esto es importante porque los valores atípicos extremos pueden afectar la fiabilidad de los estimadores MCO.

Para entender los temas revisados, lo invito a que revisen adicionalmente a lo visto en el libro de Wooldridge, y el capítulo 1 del libro de Meza.



### Actividades de aprendizaje recomendadas

Reforcemos el aprendizaje resolviendo las siguientes actividades.

1. Reflexione sobre los temas de la unidad 2 y trabajados en la guía virtualizada y en los temas introductorios del capítulo 1 libro de Wooldridge y capítulo 1 del libro de Meza.

**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.

2. Realice los ejercicios del 4 al 6 de la página 61 del libro de Wooldridge.

**Procedimiento:** estos ejercicios le permitirán comprender la interpretación de resultados de un modelo de regresión simple y entender los supuestos de la metodología econométrica.

3. Revise las orientaciones desarrolladas por el docente en el aula virtual y participe en la tutoría permanente.

**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. Por favor, complete las actividades en un cuaderno o documento Word.

## Contenidos, recursos y actividades de aprendizaje recomendadas



### Semana 4

## Unidad 2. Regresión lineal con un regresor

### 2.7 Unidades de medición y forma funcional

Aquí revisaremos dos temas importantes:

#### 1. Efectos de los cambios de unidades de medición sobre los estadísticos de MCO

En el ejemplo que hemos venido trabajando sobre las variables PIB e inflación, la primera está medida en millones de dólares y la segunda en porcentaje. Hay que entender que, si transformamos los millones a miles de millones, los resultados de las estimaciones cambian de una forma esperada, por ejemplo, si el PIB del año 2017 es de: 104.300.000.000 millones y lo expresamos en miles de millones, entonces quedaría como: 104.300.000 miles de millones, por lo que ante estas transformaciones hay que tener cuidado a la hora de interpretar resultados. Es importante entender que el cambio de unidades de medición no afecta a las propiedades de los estimadores, ni tampoco afecta al coeficiente de determinación ( $R^2$ ), ya que el valor que expresa este coeficiente está expresado en porcentaje o en decimales.

#### 2. Incorporación de no linealidades en la regresión simple

La teoría econométrica manifiesta que la linealidad debe mantenerse en los parámetros, es decir, en las betas, pero esta linealidad puede estar o no en las variables del modelo. Para que un modelo sea considerado como lineal



en sus variables, estas no pueden estar elevadas a un exponente mayor a uno, ni multiplicadas o divididas para otra variable. Para que el modelo sea lineal, esta debe darse en todas las variables.

Es frecuente encontrar en las ciencias sociales ecuaciones de regresión, cuya variable dependiente es logarítmica.

Supongamos que tenemos el modelo de que el salario está en función de los años de educación.

$$\text{wage} = \beta_0 + \beta_1 \text{edu} + u \quad (20)$$

Donde el salario se mide en dólares por mes y la educación en años. Al interpretar esta ecuación, diríamos que, ante un incremento de un año de educación, el salario por mes en promedio aumentaría en X dólares. Pero si alguna variable deja de ser lineal, como se muestra en la siguiente ecuación:

$$\log(\text{wage}) = \beta_0 + \beta_1 \text{edu} + u \quad (21)$$

La interpretación ahora cambia, ya que al estar la variable Y en logaritmos, su unidad de medición pasó de dólares a porcentaje, por lo que la interpretación cambia y ahora nos diría que, ante un incremento de un año en educación, el salario se incrementa en promedio en X porcentaje. Ahora  $B_1$  se multiplica por 100, para obtener el cambio porcentual de Wage por un año más de educación.

El momento en que una de las variables del modelo deja de ser lineal, ya podemos decir, que el modelo cambia de forma funcional de lineal a no lineal en las variables. Aquí un resumen de algunas formas funcionales.



**Tabla 2**  
Resumen de formas funcionales

Modelo	Variable dependiente	Variable independiente	Interpretación de $\beta_1$
Nivel-nivel	$y$	$x$	$\Delta y = \beta_1 \Delta x$
Nivel-log	$y$	$\log(x)$	$\Delta y = \frac{\beta_1}{100\%}$
Log-nivel	$\log(y)$	$x$	$\% \Delta y = 100 \beta_1$
Log-log	$\log(y)$	$\log(x)$	$\% \Delta y = \beta_1 \%$

Nota. Adaptado de *Introducción a la Econometría* (p. 46), por Wooldridge, J., 2010, Cengage Learning. México.

## 2.8 Valores esperados y varianzas de los estimadores de MCO

Revisaremos las propiedades estadísticas de MCO, a través de las distribuciones de los  $\hat{\beta}_0$  y  $\hat{\beta}_1$  que resultan de diversas muestras aleatorias. Estas propiedades o supuestos complementan lo visto en el punto 2.4.

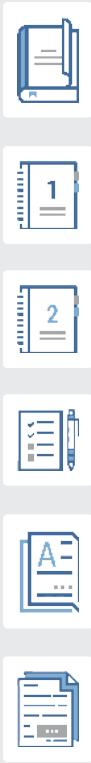
### Inesgamiento de los estimadores de MCO

Este inesgamiento se da cuando:

- Hay linealidad en los parámetros.
- Muestreo aleatorio.
- Variación muestral de la variable explicativa.
- Media condicional cero.

### Varianza de los estimadores de MCO

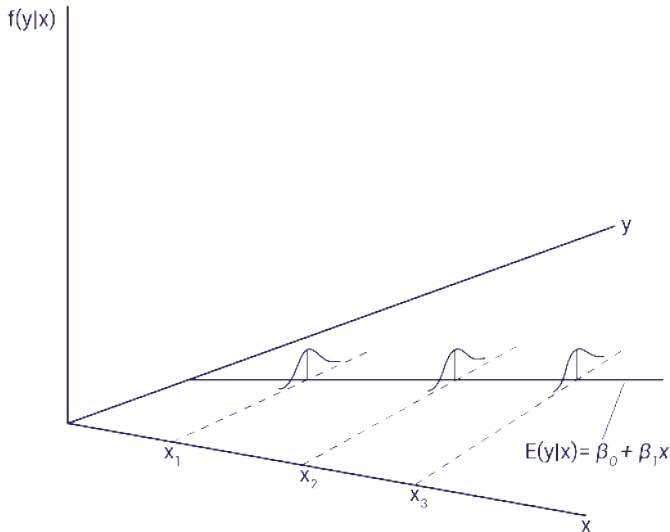
En cambio, los supuestos o propiedades ligados a la varianza o desviación estándar de los estimadores de MCO es la homocedasticidad.



$$\text{Var}\left(\frac{u}{x}\right) = \sigma^2 \quad (22)$$

**Figura 5**

*Modelo de regresión bajo homocedasticidad*



*Nota. Adaptado de Introducción a la Econometría (p. 54), por Wooldridge, J., 2010, Cengage Learning. México, CC BY 4.0.*

Esto significa que el valor esperado de Y dado X, es lineal en X, pero la varianza de Y dada X es constante.

### **Estimación de la varianza del error**

Conociendo la desviación  $\sigma^2$  se puede estimar la  $\text{Var}(\hat{\beta}_1)$  y  $\text{Var}(\hat{\beta}_0)$ . Es importante diferenciar entre los errores y los residuales, los errores provienen de la ecuación poblacional:  $Y_i = \beta_0 + \beta_1 X + u_i$ , mientras que los residuos provienen de los parámetros estimados y por ende se puede calcular estos residuos, pero no se pueden obtener los errores poblacionales.



## 2.9 Regresión a través del origen

Este tipo de regresión carece del intercepto o constante, es decir, que la recta de regresión pasa a través del punto  $X=0$ . Para estimar este modelo también se emplea la metodología de MCO. Adopta la siguiente forma funcional:

$$\tilde{y} = \tilde{\beta}_1 x \quad (23)$$

En STATA se realiza de la siguiente forma, observe la parte resaltada de los resultados, donde se indica los comandos para realizar este tipo de regresión:

**Figura 6**

*Modelo de regresión a través del origen*

. reg PIBUSapreciosactuales Inflaciónpreciosalconsumidor, noconstant						
Source	SS	df	MS	Number of obs	=	52
Model	1.0094e+22	1	1.0094e+22	F(1, 51)	=	3.96
Residual	1.2997e+23	51	2.5485e+21	Prob > F	=	0.0519
Total	1.4006e+23	52	2.6936e+21	R-squared	=	0.0721
				Adj R-squared	=	0.0539
				Root MSE	=	5.0e+10

PIBUSapreciosactuales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Inflaciónpreciosalconsumidor	4.76e+08	2.39e+08	1.99	0.052	-4168952	9.56e+08

*Nota. Ochoa, O., 2023.*

Como se puede observar en la figura 6, aquí no aparece el término de la constante.

Para entender los temas revisados lo invito a que revise el capítulo 2 del libro de Wooldridge.



### Actividades de aprendizaje recomendadas

Es momento de aplicar sus conocimientos a través de las actividades que se han planteado a continuación:

1. Reflexione sobre los temas trabajados en la guía virtualizada y en el texto de Wooldridge y del libro de Meza.

**Procedimiento:** para que comprenda los temas desarrollados, le aconsejo que revise con conceptos principales y revise los ejercicios resueltos.

2. Realice los ejercicios: 4, 5, y 7, apartado 1.14 del libro de Meza.

**Procedimiento:** para que comprenda los temas desarrollados, le aconsejo que realice los ejercicios que se plantean en cada tema revisado.

3. Revise las orientaciones desarrolladas por el docente en el aula virtual y participe en la tutoría permanente.

**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. Por favor, complete las actividades en un cuaderno o documento Word.*

## Contenidos, recursos y actividades de aprendizaje recomendadas



### Semana 5

## Unidad 2. Regresión lineal con un regresor

### 2.10 Contraste de hipótesis acerca de uno de los coeficientes de regresión

Una vez que estimamos los resultados de cualquier modelo de regresión simple, tenemos que determinar si esos resultados son confiables o no. En nuestro ejemplo del PIB e inflación, cómo podemos confiar en que un cambio de un 1 % en la inflación, genera que el PIB caiga en promedio en  $-8.44 \times 10^8$  millones de dólares; cómo saber si este resultado es significativo estadísticamente y además que la variable inflación es adecuada para explicar los cambios sobre el PIB.



Lo que buscamos aquí es contrastar la pendiente y el intercepto, por eso comenzamos con un contraste bilateral ( $B_1$ ) y un contraste unilateral ( $B_0$ ) de los parámetros antes mencionados. Para esto establecemos las siguientes hipótesis:

- $H_0$ :  $B_1 = 0$  hipótesis nula.
- $H_1$ :  $B_1 \neq 0$  hipótesis alterna.

Con la ayuda del STATA, obtenemos los Errores Estándar (ES), y estos nos sirven para calcular los valores del estadístico t, para cada beta. La fórmula es:

$$t = \frac{\hat{\beta}_1 - \beta_0}{ES(\hat{\beta}_1)} \tag{24}$$

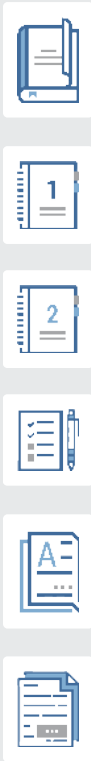
En resumen, es dividir el valor del parámetro obtenido para su error estándar.

El STATA, adicionalmente, nos ayuda a obtener el estadístico t y la probabilidad del t (p-valor). Estos dos criterios pueden ser utilizados para contrastar la hipótesis nula, y cada uno de ellos tiene una regla general, que se describe en el siguiente cuadro:

**Tabla 3**  
*Criterios de decisión para aceptar o rechazar la  $H_0$*

Estadísticos utilizados	Criterios de decisión para contrastar la $H_0$	
	Estadístico calculado	P-valor del estadístico
<i>t - student</i>		
<i>F - Fisher</i>	<i>tc &gt; tt = rechazar la <math>H_0</math>, si la regla no se cumple NO</i>	<i>si el p-valor (t), es menor al nivel de significancia (1%, 5% y 10%), se puede</i>
<i>Chi-cuadrado (X2)</i>	<i>podemos rechazar la <math>H_0</math> y nos toca aceptarla.</i>	<i>rechazar la <math>H_0</math> y aceptar la alterna (<math>H_1</math>), caso contrario debemos aceptar la <math>H_0</math></i>
<i>Z - normal</i>		

Nota. Adaptado de *Introducción a la Econometría* (p. 156), por Wooldridge, J., 2010, Cengage Learning. México.



Para nuestro ejemplo tenemos los siguientes resultados:

**Figura 7**  
*Uso de los valores t y p-valor para contrastar la H0*

```
. reg PIBUSapreciosactuales Inflaciónpreciosalconsumidor
```

Source	SS	df	MS	Number of obs	=	52
Model	1.6587e+22	1	1.6587e+22	F(1, 50)	=	18.40
Residual	4.5067e+22	50	9.0133e+20	Prob > F	=	0.0001
Total	6.1653e+22	51	1.2089e+21	R-squared	=	0.2690
				Adj R-squared	=	0.2544
				Root MSE	=	3.0e+10

PIBUSapreciosactuales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Inflaciónpreciosalconsumidor	-8.44e+08	1.97e+08	-4.29	0.000	-1.24e+09 -4.49e+08
_cons	5.59e+10	5.76e+09	9.71	0.000	4.43e+10 6.75e+10

Nota. Ochoa, O., 2023.

Si utilizamos el criterio del p-valor ( $p > |t|$ ), vemos que, para ambas betas, el valor de su probabilidad es menor al nivel de significancia del 5%. Al rechazar  $H_0$ , estamos concluyendo al menos en el parámetro  $B_1$ , que la variable Inflación es una variable adecuada para explicar cambios sobre el PIB, y que su resultado es estadísticamente significativo. Dejo al lector emitir la conclusión del parámetro  $B_0$ .

2.11 Intervalos de confianza para coeficiente de regresión

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,  $\hat{\beta}_2$  de  $\beta_2$  para esto buscamos dos números,  $\sigma$  y  $\alpha$  este último situado entre 0 y 1.

Para esto utilizamos la siguiente ecuación tanto para  $\hat{\beta}_0$  y  $\hat{\beta}_1$ :

**Figura 8**

*Ecuación para intervalos de confianza para coeficiente de regresión*

The diagram shows the equation 
$$\Pr(\hat{\beta}_2 - \delta \leq \beta_2 \leq \hat{\beta}_2 + \delta) = 1 - \alpha$$
 inside a light gray rounded rectangle. Four arrows point from labels to parts of the equation: 'Límite inferior' points to  $\hat{\beta}_2 - \delta$ , 'Límite' points to  $\hat{\beta}_2$ , 'Nivel de significancia' points to  $\alpha$ , and 'Coeficiente de confianza' points to  $1 - \alpha$ .

*Nota. Adaptado de Econometría (p. 132), por Gujarati, D., 2010, McGrawHill Interamericana editores S.A, CC BY 4.0.*

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. 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.



### Actividades de aprendizaje recomendadas

Es hora de reforzar los conocimientos adquiridos resolviendo las siguientes actividades:

1. Desarrolle una autoevaluación de conocimientos de la unidad 2 y las actividades recomendadas en la guía

**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.

2. Revise el punto 1.3.4, de la página 33, sobre el cálculo de los intervalos de confianza del libro de Meza.

**Procedimiento:** Revise cómo estos intervalos se calculan y cómo sus resultados permiten el contraste de los resultados del modelo.

*Nota. Por favor, complete las actividades en un cuaderno o documento Word.*

## Contenidos, recursos y actividades de aprendizaje recomendadas



### Semana 6

## Unidad 2. Regresión lineal con un regresor

### 2.12 Heterocedasticidad y homocedasticidad

En este apartado revisaremos de forma general los conceptos de heterocedasticidad y homocedasticidad, ya que estos temas serán ampliados en la materia de Modelos Econométricos. Cabe resaltar que la heterocedasticidad se produce cuando no se cumple uno de los supuestos de la metodología de MCO.



La heterocedasticidad y la homocedasticidad son términos que se utilizan para describir las características de la varianza de los errores en un modelo de regresión.

La **homocedasticidad**, también conocida como homogeneidad de varianza, se refiere a la situación en la que la varianza de los errores es constante en todos los niveles de las variables independientes. En otras palabras, la dispersión de los errores es la misma en todo el rango de los valores de las variables predictoras. Esto implica que la variabilidad alrededor de la línea de regresión es constante.



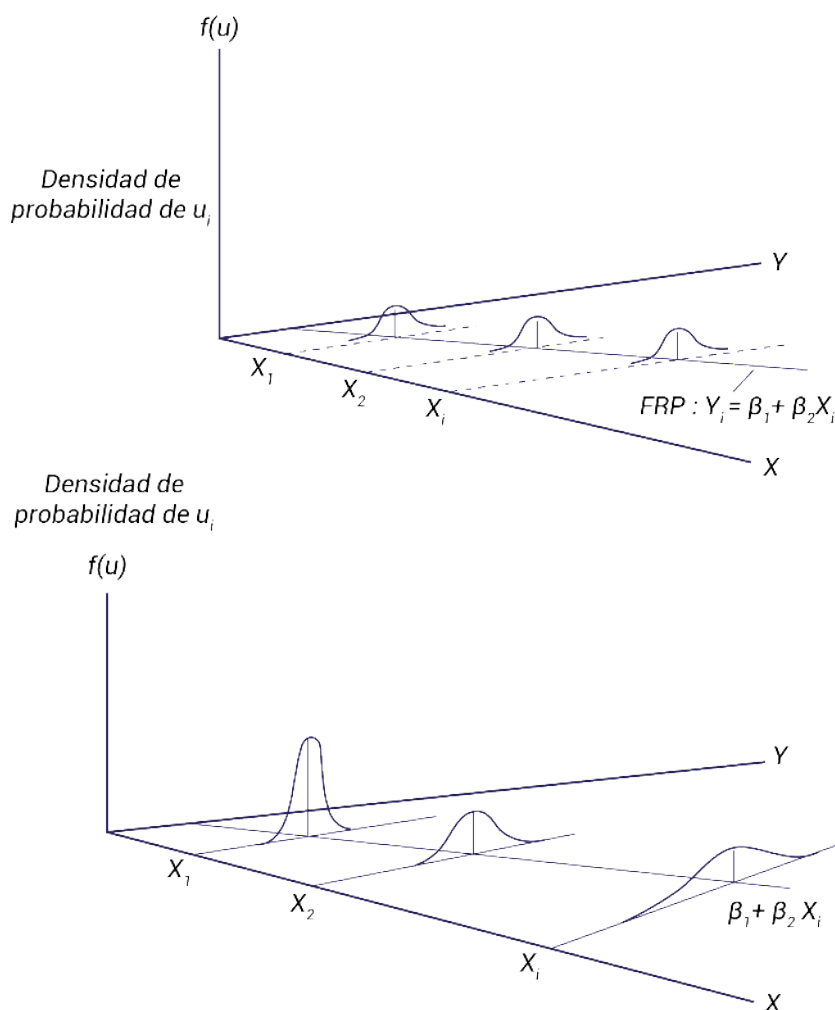
Por otro lado, la **heterocedasticidad** se presenta cuando la varianza de los errores no es constante en todos los niveles de las variables independientes. En este caso, la dispersión de los errores cambia a medida que varían los valores de las variables predictoras, cuyas implicaciones son importantes en el análisis de regresión. Puede conducir a estimaciones ineficientes de los coeficientes de regresión, lo que afecta la precisión de las inferencias y las pruebas de hipótesis. Además, los intervalos de confianza y los p-valores pueden estar sesgados si se ignoran los efectos de la heterocedasticidad.

Este tema será revisado con mayor detalle en la asignatura de Modelos Económicos, donde se conocerá su naturaleza, cómo se puede detectar este problema y las posibles soluciones.



**Figura 9**

*Homocedasticidad y heterocedasticidad*



*Nota. Adaptado de Econometría (p. 65), por Gujarati, D., 2010, McGrawHill Interamericana editores S.A, CC BY 4.0.*

## 2.13 Fundamentos teóricos de mínimos cuadrados ordinarios

El estimador de MCO en la regresión lineal tiene varias propiedades favorables. Es insesgado, lo que significa que, en promedio, no está sesgado y tiende a estimar correctamente el valor poblacional. También es consistente, lo que

implica que, a medida que el tamaño de la muestra aumenta, el estimador se acerca al valor verdadero. La varianza del estimador MCO es inversamente proporcional al tamaño muestral, lo que significa que, a medida que aumenta el tamaño de la muestra, la precisión de la estimación mejora. Cuando el tamaño muestral es grande, el estimador MCO sigue una distribución muestral normal. Además, si se cumplen ciertos supuestos y los errores son homocedásticos, el estimador MCO es más eficiente que otros estimadores lineales incondicionalmente insesgados en términos de tener la menor varianza.

### *Estimadores lineales condicionalmente insesgados y teorema de Gauss-Markov.*

Si se cumple todo lo antes mencionado, el estimador de MCO sería el estimador lineal condicionalmente insesgado óptimo, es decir, es ELIO, es lo que manifiesta el teorema de Gauss-Markov. Lo que busca el teorema es que, bajo un conjunto de condiciones conocidas, el estimador de MCO  $\beta_1$  tiene la menor varianza condicional, dados  $X_1, \dots, X_n$ , de todos los estimadores lineales condicionalmente insesgados de  $B_1$ . Este teorema tiene dos limitaciones importantes:

- Las condiciones o supuestos podrían no cumplirse en la práctica, por ende, los estimadores dejan de ser ELIO.
- Que, a pesar de cumplir con los supuestos del teorema, existen otros posibles estimadores que pueden ser no lineales, pero sí insesgados, lo que los vuelve más eficientes que MCO.

Pero existe una metodología diferente a los MCO, cuando los errores son heterocedásticos que permita obtener un estimador con menor varianza, este método se denomina Mínimos Cuadrados Ponderados (MCP), y lo que hace es ponderar la  $i$ -ésima observación por la inversa de la raíz cuadrada de la varianza condicional de  $u_i$  dado  $X_i$ . Debido a esto, los errores de esta regresión son homocedásticos.



## 2.14 La utilización del estadístico t en regresión para muestras pequeñas

Si se cumplen los supuestos de los MCO, y el estadístico t válido con homocedasticidad presenta una distribución t de student. El estadístico t para contrastar que  $B_1=0$  es  $\tilde{t} = \frac{\hat{\beta}_1 - \beta_0}{\tilde{\sigma}\hat{\beta}_1}$ , fórmula similar a lo visto en temas anteriores con respecto a determinar la significatividad de un estimador. Aunque en la práctica es difícil cumplir todos estos supuestos y sobre todo que los errores sean homocedásticos y estén normalmente distribuidos.



### Actividades de aprendizaje recomendadas

Es momento de aplicar sus conocimientos a través de las actividades que se han planteado a continuación:

1. Reflexione sobre los temas trabajados en la Guía virtualizada y el capítulo 1, punto 1.9, del texto de Meza y el capítulo 3 de Wooldridge.

**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.

2. Revise el punto 1.3.5 del planteo de hipótesis y el cálculo del estadístico t, para los coeficientes de regresión, para determinar su significancia individual, en la página 36 del texto de Meza.

**Procedimiento:** Para cada una de las semanas, su docente tutor realizará orientaciones sobre el tema para acercar con mayor claridad los temas expuestos tanto en la guía virtualizada como en los libros. Allí su docente tutor ubicará ejemplos demostrativos para cada uno de los temas que se desarrollan en la semana.

3. Desarrolle la autoevaluación de esta segunda unidad y las actividades recomendadas en la guía virtualizada.



**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.

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

De acuerdo a la información analizada, lea comprensivamente cada uno de los ítems. Cuando la pregunta sea dicotómica, escriba la respuesta (V/F), y en las preguntas de opción múltiple, elija la respuesta correcta.

1. Una regresión simple es aquella:
  - a. Regresión con un único regresor.
  - b. Es el promedio de una variable independiente.
  - c. Es el planteo de hipótesis sobre una variable.
2. La variable dependiente es aquella:
  - a. Variable que no depende de otra variable.
  - b. Es la variable regresora.
  - c. Es la variable regresada.
3. En el modelo de regresión con la función  $Y = B_0 + B_1X + u$ , el  $B_1$  representa:
  - a. Es el residuo de la regresión.
  - b. Es la pendiente.
  - c. Es el intercepto o constante.
4. Una medida de bondad de ajuste del modelo es:
  - a. Los residuos.



- b. Coeficiente de correlación.
- c. Coeficiente de determinación.

5. ( ) ¿Los errores estándar son una medida de ajuste del modelo?

6. Uno de los supuestos de mínimos cuadrados es:

- a. Dispersión de los errores.
- b. Los datos atípicos elevados son improbables.
- c. Netamente, para la reconstrucción de la infraestructura afectada.

7. Contrastar la hipótesis nula se lo puede hacer a través de:

- a. El valor del t estadístico.
- b. Con los errores.
- c. Con el coeficiente de determinación.

8. ( ) El cambio en las unidades de medición altera las propiedades de los estimadores de MCO.

9. ( ) ¿Una forma funcional de un modelo de regresión distinta a la lineal puede ser la logarítmica?

10. ( ) Puede existir o no linealidad en los parámetros.

[Ir al solucionario](#)



### Resultado de aprendizaje 3:

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

Este resultado de aprendizaje nos permite conocer sobre los modelos de regresión múltiple, en los cuales existe más de una variable independiente en la construcción de un modelo econométrico.

#### Contenidos, recursos y actividades de aprendizaje recomendadas



#### Semana 7

### Unidad 3. Análisis de regresión múltiple: estimación

Esta unidad es tomada del capítulo 3, del libro de Wooldridge.

#### 3.1 Motivación para la regresión múltiple

No es muy común que dentro del análisis econométrico se den modelos de regresión simple; por lo tanto, es necesario ampliar este modelo a modelos de regresión con varias variables explicativas. En la construcción de estos modelos podemos encontrarnos ahora que las variables independientes puedan estar relacionadas o que ahora se puedan incorporar otras formas funcionales distintas a las que vimos en semanas anteriores. Al igual que los modelos de regresión simple, estos modelos se calculan con la metodología de MCO.

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

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u \quad (30)$$

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



Para nuestro ejemplo del PIB, podemos añadir una variable independiente que puede ser las exportaciones. Con esto pasamos de un modelo de regresión simple a uno múltiple, cuya función es la siguiente:

$$PIB = \beta_0 + \beta_1 X_1 + \beta_2 Exportaciones_2 + u \tag{31}$$

Donde ahora el  $B_1$  mide el cambio en el PIB respecto a la inflación, manteniendo constante a las exportaciones; mientras que  $B_2$  mide el cambio en el PIB respecto de las exportaciones, manteniendo constante la inflación.

En la siguiente tabla podemos ver algunas terminologías para expresar a la variable dependiente e independiente.

**Tabla 4**  
*Terminología de la regresión múltiple*

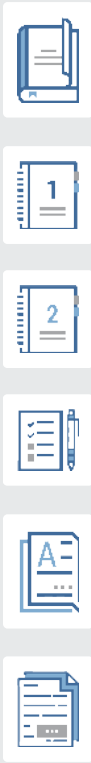
Y	x1, x2, .....xk
Variable dependiente	Variables independientes
Variable explicada	Variables explicativas
Variable de respuesta	Variables de control
Variable predicha	Variables predictoras
Regresando	Regresores

Nota. Adaptado de *Introducción a la Econometría* (p. 72), por Wooldridge, J., 2010, Cengage Learning. México.

### 3.2 Mecánica e interpretación de los mínimos cuadrados ordinarios

Partimos de la estimación de un modelo con dos variables independientes. La ecuación se muestra a continuación:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u \tag{32}$$



Donde para obtener los estimadores de las betas, usamos los MCO, que minimizan la suma de los residuales al cuadrado.

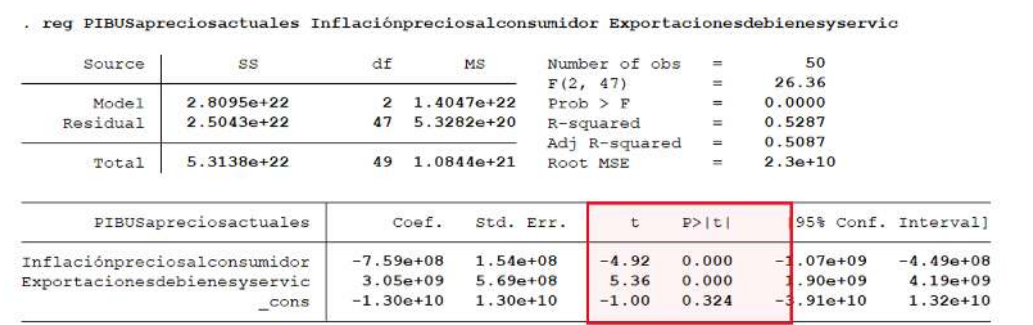
$$\sum_{i=1}^n \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} \right)^2$$

Este problema de minimización se resuelve empleando el cálculo multivariante:

$$\begin{aligned} \sum_{i=1}^n \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \dots - \hat{\beta}_k X_{ik} \right) &= 0 \\ \sum_{i=1}^n X_{i1} \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \dots - \hat{\beta}_k X_{ik} \right) &= 0 \\ \sum_{i=1}^n X_{i2} \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \dots - \hat{\beta}_k X_{ik} \right) &= 0 \\ \sum_{i=1}^n X_{ik} \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \dots - \hat{\beta}_k X_{ik} \right) &= 0 \end{aligned} \tag{34}$$

A estas ecuaciones se las conoce como condiciones de primer orden de MCO. Con la ayuda del STATA, procederemos a calcular nuestro modelo de regresión múltiple del PIB en función de la inflación y las exportaciones.

**Figura 10**  
*Modelo de regresión múltiple*



Nota. Ochoa, O., 2023.

La interpretación de esta regresión sería que ante un incremento en la inflación de 1 %, el PIB en promedio disminuiría en 759.000.000 millones de dólares, manteniendo constante las demás variables. En cambio, cuando las

exportaciones de bienes y servicios aumentan en 1 %, el PIB aumenta en promedio en 3050.000.000 millones de dólares, manteniendo constante las otras variables. Recuerde que las pendientes miden los cambios promedios de la variable dependiente ante cambios en las variables independientes.

A veces es necesario conocer el cambio en las variables independientes y no solo el cambio individual de una variable, por lo que tenemos:

$$\Delta \hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2 + \dots + \hat{\beta}_k \Delta x_k \quad (35)$$

Por lo que para nuestro ejercicio tendríamos que el PIB disminuye en promedio ante las variaciones de las dos variables en 729.000.000 millones de dólares.

$$\Delta PIB = -7.59e08 \Delta \text{Inflación} + 3.05e09 \Delta \text{Exportaciones} = -7.29e08 \quad (36)$$

Una vez realizada la regresión podemos obtener los valores predichos o estimados de la variable dependiente, en este caso del PIB estimado por cada año.

En la ecuación estimada se debe reemplazar en cada variable independiente por el valor correspondiente a cada una de ellas, con respecto al año que deseamos estimar, por ejemplo, quiero saber cuál es el valor del PIB estimado en el año 2013.

$$PIB_{2013} = -7.59e08 \cdot \text{Inflación} + 3.05e09 \cdot \text{Exportaciones} + u_i \quad (37)$$

$$PIB_{2013} = -7.59e08 \cdot \text{Inflación}(2.72) + 3.05e09 \cdot \text{Exportaciones}(28.6) + u_i$$

$$PIB_{2013} = -7.59e08 \cdot \text{Inflación}(2.72) + 3.05e09 \cdot \text{Exportaciones}(28.6) + u_i$$

$$PIB_{2013} = 8.7 \text{ miles de millones de dólares}$$

En cambio, que los residuos (u) de la regresión de forma matemática se obtienen de la siguiente forma:

$$\hat{u} = y_{\text{original}} - y_{\text{estimada}} \quad (38)$$

En el programa STATA, tanto los valores estimados de Y (PIB) y de los residuos (u), se obtienen de la siguiente forma:



```
.predict Yestimado, xb
(2 missing values generated)
.predict RESIDUOS, res
(2 missing values generated)
```

**Figura 11**

*Valores estimados del PIB y residuos de la regresión*

Year	Dolarización	Yestimado	RESIDUOS
1970	0	1.19e+10	-9.04e+09
1971	0	1.29e+10	-1.01e+10
1972	0	1.89e+10	-1.57e+10
1973	0	3.16e+10	-2.77e+10
1974	0	3.84e+10	-3.18e+10
1975	0	2.43e+10	-1.66e+10
1976	0	2.78e+10	-1.87e+10
1977	0	2.68e+10	-1.58e+10
1978	0	2.31e+10	-1.12e+10
1979	0	3.38e+10	-1.96e+10
1980	0	2.94e+10	-1.15e+10
1981	0	1.91e+10	2.68e+09

*Nota. Ochoa, O., 2023.*

Una vez interpretada las estimaciones, también debemos interpretar el coeficiente de determinación ( $R^2$ ), que en nuestra regresión es de: R-squared= 0.5287 o 53 %, lo que nos dice que el 53 % de los cambios promedios que se dan en el PIB, están siendo explicados por la inflación y las exportaciones.

### 3.3 Valor esperado de los estimadores de MCO

Aquí revisaremos cuatro supuestos adicionales a los ya revisados para los MCO, pero que se aplican a modelos de regresión múltiple.

**Figura 12**

*Supuestos de MCO para modelos de regresión lineal múltiple (RLM)*

**Supuesto 1**

Lineal en los parámetros



**Supuesto 2**

Muestreo aleatorio



**Supuesto 3**

No hay colinealidad perfecta



**Supuesto 4**

Media condicional cero



*Nota. Adaptado de Introducción a la Econometría (pp. 83-86), por Jeffrey Wooldridge, 2010, Cengage Learning. México, CC BY 4.0.*

Los supuestos anteriores nos permiten entender más el término insesgamiento, el cual manifiesta que los estimadores de Mínimos Cuadrados Ordinarios (MCO) son insesgados, eso significa que, en promedio, estos estimadores no están sistemáticamente sesgados hacia arriba o hacia abajo al estimar los parámetros de un modelo estadístico.

En la regresión lineal, los estimadores MCO se utilizan para estimar los coeficientes de regresión que describen la relación entre una variable dependiente y una o más variables independientes. Estos estimadores buscan minimizar la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos por el modelo. La insesgadez de los estimadores MCO implica que, en promedio, el valor esperado de los estimadores es igual al valor verdadero de los parámetros que se están estimando.

## Inclusión de variables irrelevantes en un modelo de regresión

En los modelos podemos tener la inclusión de una variable irrelevante o de sobreestimar el modelo en el análisis de regresión múltiple. Esto significa que una (o más) de las variables independientes está incluida en el modelo, aun cuando en la población no tiene ningún efecto parcial sobre  $Y$ , es decir, su coeficiente poblacional es cero. Esto no afecta el insesgamiento de MCO, pero sí puede afectar la varianza de los estimadores de MCO.

## Sesgo de la variable omitida

Podemos encontrarnos con un caso contrario, en el que se excluye una variable importante del modelo, esto generalmente sucede cuando se tiene más de una variable no significativa para el modelo, y decide eliminarse todas ellas de una sola. Cuando se realiza esto puede que entre las variables eliminadas esté una que es importante para explicar a  $Y$ , que pudo volverse significativa al eliminar de forma progresiva cada una de las variables independientes no significativas.

Tanto la inclusión de variables no relevantes y la eliminación de una variable importante pueden generar un error que se lo conoce como error o sesgo de especificación del modelo.



Para entender mejor los temas revisados aquí, revise el capítulo 3 del libro de Wooldridge hasta el tema 3.4.

## 3.4 Varianza de los estimadores de MCO

Con la varianza de los estimadores, se conocerá la tendencia central de las betas, además de tener una medida de dispersión en su distribución de muestreo. Aquí hacemos referencia al supuesto de la homocedasticidad, que si lo cumple tendremos una propiedad importante de eficiencia.

Haciendo referencia a nuestro ejemplo del PIB:



$$PIB = \beta_0 + \beta_1 \text{Inflación} + \beta_2 \text{Exportaciones} + u_i \quad (39)$$

Lo que se busca es que la varianza del error no observado ( $u$ ), no dependa de las variables independientes. Es decir.

$$\text{Var}(u \mid \text{Inflación}, \text{Exportaciones}) = \sigma^2 \quad (40)$$

Ahora, si esta varianza deja de ser constante y cambia de acuerdo con alguna de las dos variables explicativas, entonces tendremos el problema de la heterocedasticidad. Entonces, el cálculo de la varianza de las betas estimadas sería con base a la siguiente fórmula:

$$\text{Var}(\hat{\beta}_j) = \frac{\delta^2}{\text{STC}_j(1-R_j^2)} \quad (41)$$

Una varianza grande significa estimadores menos precisos y esto se traduce en intervalos de confianza grandes y pruebas de hipótesis menos exactas.

La varianza, al tener dentro de su cálculo a la suma total de cuadrados, debemos revisar la varianza del error, la variación muestral total en las variables independientes, y las relaciones independientes entre las variables independientes, tal vez por el problema de la multicolinealidad, tema que lo revisaremos en profundidad en las últimas semanas del segundo bimestre.

### 3.5 Eficiencia de MCO: el teorema de GAUSS – MARKOV

Este teorema hace referencia al uso del método de MCO, en lugar de otros métodos de estimación. Uno de ellos es el supuesto relacionado con el insesgamiento, que a más de cumplir con los otros supuestos que vimos en la unidad anterior, los estimadores a través de MCO, serían los Mejores Estimadores Lineales Insesgados (MELI) o también los Estimadores Lineales Insesgados Óptimos (ELIO), que es otra forma de referirlos.



Al hablar de linealidad, nos referimos a la linealidad en los parámetros, y el término mejor, hace referencia a la menor varianza que presenten los estimadores. Claro está que si no se cumple alguno de los supuestos, este teorema ya no será válido.



## Actividades de aprendizaje recomendadas

Es momento de aplicar sus conocimientos a través de las actividades que se han planteado a continuación:

1. Realice los ejercicios del 3.2 al 3.4 del libro de Wooldridge, página 106.

**Procedimiento:** para cada una de las semanas, su docente tutor realizará orientaciones sobre el tema para acercar con mayor claridad los temas expuestos tanto en la guía virtualizada como en el libro. Allí su docente tutor ubicará ejemplos demostrativos para cada uno de los temas que se desarrollan en la semana. Aquí aprenderá a interpretar los resultados de un modelo de regresión múltiple.

2. Revise las orientaciones desarrolladas por el docente en el aula virtual y participe en la tutoría permanente.

**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. Por favor, complete las actividades en un cuaderno o documento Word.*

3. Desarrolle la autoevaluación de esta tercera unidad y las actividades recomendadas en la guía virtualizada.

**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.



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

De acuerdo a la información analizada, lea comprensivamente cada uno de los ítems. Cuando la pregunta sea dicotómica, escriba la respuesta (V/F), y en las preguntas de opción múltiple, elija la respuesta correcta.

1. Un modelo de regresión múltiple es aquel:
  - a. Que tiene más de dos variables dependientes.
  - b. Que tiene más de una variable independiente.
  - c. Que tiene múltiples ecuaciones o funciones.
2. El método de mínimos cuadrados ordinarios se basa en:
  - a. La minimización de los residuales al cuadrado.
  - b. Calcular el coeficiente de correlación.
  - c. Calcular los valores del estadístico t.
3. ( ) ¿Las estimaciones  $B_0$  y  $B_1$  se interpretan como efectos parciales?
4. ( ) ¿El efecto “ceteris paribus” hace referencia a la forma funcional de un modelo?
5. La regresión a través del origen es aquella que:
  - a. No tiene el intercepto.
  - b. No tiene la pendiente.
  - c. Carece de los residuos.
6. Incluir una variable irrelevante produce:
  - a. Sobre especificación del modelo.



- b. Otra forma funcional.
  - c. Modelo de regresión a través del origen.
7. ( ) Cuando habla de homocedasticidad nos dice que los errores tienen la misma varianza.
8. ( ) ¿La varianza de  $B_j$  depende de tres factores?
9. Un análisis de error de especificación hace referencia a:
- a. Sesgo por omisión de variable.
  - b. Problemas de heterocedasticidad.
  - c. Un modelo lineal.
10. Colinealidad hace referencia a:
- a. Error de especificación.
  - b. Variable omitida.
  - c. Relación entre variables independientes.

[Ir al solucionario](#)



## Resultados de aprendizaje 1 a 3:

- Analiza la relación lineal entre dos variables y predice la dependencia de una con respecto a la otra.
- Comprende la utilidad de un análisis de regresión lineal simple.
- Analiza la relación lineal entre una variable dependiente en función de otras denominadas independientes.

### Contenidos, recursos y actividades de aprendizaje recomendadas



## Semana 8

### Actividades finales del bimestre

Estimado estudiante, en esta semana se va a realizar un repaso de las unidades revisadas en el primer bimestre.

- **Unidad 1.** La naturaleza de la econometría.
- **Unidad 2.** Regresión lineal con un regreso único.
- **Unidad 3.** Análisis de regresión múltiple: estimación.



### Actividades de aprendizaje recomendadas

Es hora de reforzar los conocimientos adquiridos resolviendo las siguientes actividades:

1. Revise las orientaciones desarrolladas por el docente en el aula virtual y participe en la tutoría permanente.

**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.

2. Desarrolle la evaluación presencial.



**Procedimiento:** debe revisar los contenidos de las tres unidades estudiadas a lo largo de este primer bimestre.





## Segundo bimestre

### Resultado de aprendizaje 3:

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

Este resultado de aprendizaje nos permite conocer sobre los modelos de regresión múltiple, en los cuales existe más de una variable independiente en la construcción de un modelo econométrico.

### Contenidos, recursos y actividades de aprendizaje recomendadas

Recuerde revisar de manera paralela los contenidos con las actividades de aprendizaje recomendadas y actividades de aprendizaje evaluadas.



## Semana 9

### Unidad 4. Análisis de regresión múltiple: inferencia

#### 4.1 Distribución de muestreo de los estimadores de MCO

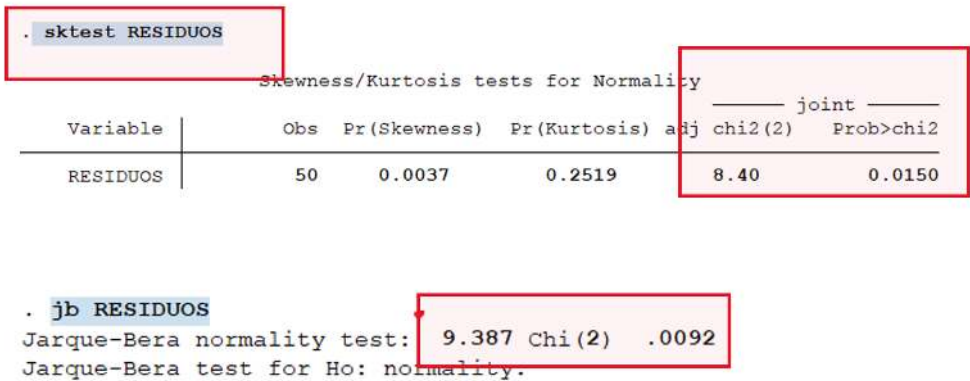
Continuamos con el libro de Wooldridge a partir de la página 118.

En este tema hacemos referencia al supuesto de la **normalidad**, de los residuos, ya que, al condicionar los valores de las variables independientes en la muestra, es claro que la distribución muestral de los estimadores de MCO, depende de la distribución de los errores. Aquí partimos del supuesto de que los errores no observables de las betas poblacionales están normalmente distribuidos. Para comprobar este supuesto dentro de nuestro modelo, lo podemos hacer con base en algunas pruebas. Existen algunas pruebas para probar la normalidad de los residuos de una regresión, tal es el caso de la



prueba de **simetría y curtosis** y la prueba de Jarque-Bera. Apliquemos estas dos pruebas a nuestro ejercicio del PIB, y veamos si los residuos de la regresión cumplen con el supuesto de normalidad.

**Figura 13**  
*Pruebas de normalidad (simetría y curtosis y Jarque-Bera)*



*Nota. Ochoa, O., 2023.*

Como se ve en la figura anterior la prueba de simetría y curtosis a través del estadístico Chi-cuadrado ( $\chi^2$ ) y de su p-valor podemos comprobar si los residuos cumplen el supuesto de normalidad, para esto las hipótesis planteadas son:

- $H_0$ : Normalidad de los residuos.
- $H_1$ : No normalidad de los residuos.

Si utilizamos el criterio de decisión basado en el p-valor del  $\chi^2$ , entonces tendremos lo siguiente:

$\Pr(\chi^2) < \text{nivel de significancia (5\%)}$ , entonces podemos rechazar  $H_0$ .

La probabilidad del  $\chi^2$  es de 1.5% y al ser menor al 5%, podemos rechazar  $H_0$  y aceptar  $H_1$ , por lo que nuestro modelo NO presenta normalidad en los errores. Lo mismo sucede con la prueba de Jarque-Bera, ya que su valor de

Chi2 es de 0.92%, menor al 5% por lo tanto también se rechaza  $H_0$  y se acepta  $H_1$ . Si no hay normalidad NO pasa nada, no significa que el modelo no sirva, solo que no cumple con este supuesto.

## 4.2 Prueba de hipótesis para un solo parámetro poblacional: la prueba t

Al igual que en las regresiones simples, la evaluación de la significatividad o de la relación lineal de cada una de las variables independientes con la dependiente se comprueba a través del valor del estadístico t.

Para probar esto a través de hipótesis, estas se plantean de la siguiente forma:

- $H_0: B_1=0$ : no hay relación lineal de  $X_1$  con respecto a la variable  $Y$ , o la variable  $X$  no es estadísticamente significativa.
- $H_1: B_1 \neq 0$ : si hay relación lineal de  $X_1$  con respecto a la variable  $Y$ , o la variable  $X$  si es estadísticamente significativa.

Lo que se debe contrastar es la  $H_0$ , y saber si se acepta o se rechaza esta hipótesis. Para esto utilizamos los criterios de decisión basada en los valores de t o del p-valor de t. En nuestra regresión los valores de  $B_0$ ,  $B_1$  y  $B_2$ , son estadísticamente significativos o tienen una relación lineal con el PIB, porque los valores del p-valor de t de cada una es menor al 5 %, y por ende se rechaza  $H_0$ , y se acepta la  $H_1$ , que manifiesta lo antes mencionado. Ver figura 14.

Para mayor entendimiento revisa los ejercicios del 4.2 al 4.4. Recuerda que el valor de t y de la construcción de sus intervalos está ligado a pruebas de hipótesis de una cola y de dos colas.

## 4.3 Intervalos de confianza

Al igual que lo hicimos en las unidades iniciales, la construcción de los intervalos de confianza para determinar la significatividad de una variable depende de los valores t, de los errores estándar y del nivel de significancia que se esté utilizando.



A los intervalos de confianza se les llama también estimaciones por intervalo, porque proporcionan un rango de valores posibles para el parámetro poblacional y no solo una estimación puntual. Las fórmulas para calcular estos intervalos son las siguientes:

$$\underline{\beta_j} = \hat{\beta_j} - c \cdot ee(\hat{\beta_j}) \tag{42}$$

$$\overline{\beta_j} = \hat{\beta_j} + c \cdot ee(\hat{\beta_j}) \tag{43}$$

A través de STATA estos intervalos de confianza se construyen con facilidad, tal como lo demuestra la siguiente figura:

**Figura 14**  
*Intervalos de confianza*

```
. reg PIBUSapreciosactuales Inflaciónpreciosalconsumidor Exportacionesdebienesyservicio
```

Source	SS	df	MS	Number of obs	=	50
Model	2.8095e+22	2	1.4047e+22	F(2, 47)	=	26.36
Residual	2.5043e+22	47	5.3282e+20	Prob > F	=	0.0000
Total	5.3138e+22	49	1.0844e+21	R-squared	=	0.5287
				Adj R-squared	=	0.5087
				Root MSE	=	2.3e+10

PIBUSapreciosactuales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Inflaciónpreciosalconsumidor	-7.59e+08	1.54e+08	-4.92	0.000	-1.07e+09	-4.49e+08
Exportacionesdebienesyservicio	3.05e+09	5.69e+08	5.36	0.000	1.90e+09	4.19e+09
_cons	-1.30e+10	1.30e+10	-1.00	0.324	-3.91e+10	1.32e+10

Nota. Ochoa, O., 2023.

### Actividades de aprendizaje recomendadas

Reforcemos el aprendizaje resolviendo las siguientes actividades.

1. Reflexione sobre los temas trabajados en la guía virtualizada y en el capítulo 4 del libro de Wooldridge.

**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.

2. Revise los ejercicios del 4.4 al 4.8 del texto de Wooldridge.

**Procedimiento:** analice los ejercicios que muestran, a través de los errores estándar y valores  $t$ , la significatividad de los resultados de la regresión.

3. Revise las orientaciones desarrolladas por el docente en el aula virtual y participe en la tutoría permanente.

**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. Por favor, complete las actividades en un cuaderno o documento Word.

## Contenidos, recursos y actividades de aprendizaje recomendadas



### Semana 10

## Unidad 4. Análisis de regresión múltiple: inferencia

### 4.4 Prueba de hipótesis de una sola combinación lineal de los parámetros

En la práctica, a veces se requiere probar hipótesis en las que interviene más de un parámetro poblacional; por ende, se busca probar en una sola hipótesis más de una Beta ( $\beta_i$ ).



Generalmente, dentro de los modelos de regresión múltiple se puede probar si las estimaciones de dos variables son iguales, ya que suelen existir variables consideradas como similares, por ejemplo:

$$\text{Consumo} = \beta_0 + \beta_1 \text{Ingreso} + \beta_2 \text{Ahorro} + \beta_3 \text{Riqueza} + u \quad (44)$$

Donde, al parecer, las variables ingreso y riqueza pueden ser consideradas similares, por ende, se debe comprobar esto a través del planteo de hipótesis, haciendo uso del estadístico F y de su p-valor. Para entender mejor este tema, realicemos un ejercicio con nuestro modelo del PIB. Lo que haremos es que al modelo del PIB, que está en función de la inflación y las exportaciones, le vamos a incorporar la variable balanza comercial, la cual se estructura con base en las exportaciones e importaciones.

Balanza comercial = exportaciones – importaciones.

Ahora vamos a estimar el siguiente modelo:

$$PIB = \beta_0 + \beta_1 \text{Inflación} + \beta_2 \text{Exportaciones} + \beta_3 \text{Balanza Comercial} + u_i \quad (45)$$

Y para determinar si hay un posible igual de los estimadores de las variables exportaciones y balanza comercial, para esto vamos a probar la hipótesis de que dos coeficientes de la regresión son iguales, y las hipótesis planteadas son las siguientes:

- $H_0: B_2 = B_3 = 0$ . Los coeficientes de regresión son iguales o no hay diferencia.
- $H_1: B_2 \neq B_3 \neq 0$ . Los coeficientes de regresión no son iguales o existe diferencia.

Procedemos a realizarlo, es STATA. Para ello, le invito a revisar la presentación interactiva:

[Prueba de hipótesis](#)



En la presentación interactiva anterior se describe todo el proceso a seguir en STATA. Para realizar esta prueba, y como se puede observar, nos fijamos en los valores del estadístico F y su p-valor, por lo que con base al p-valor podemos rechazar la  $H_0$ , y aceptar que no hay una igualdad en los coeficientes de regresión de  $B_2$  y  $B_3$ .

Para un mejor entendimiento de la parte matemática y demostrativa de la construcción de las hipótesis para probar esto, revise el ejemplo de la página 141 del libro de Wooldridge.

#### 4.5 Pruebas de restricciones lineales múltiples: la prueba F

Esta prueba busca determinar si un grupo de variables no tiene efecto sobre la variable dependiente, es decir, vamos a probar si las variables independientes en conjunto tienen o no una relación lineal con la variable dependiente o si estas en conjunto son estadísticamente significativas para explicar cambios sobre la variable dependiente. Para ejemplificar esto, volvamos a nuestro modelo del PIB, el cual está siendo explicado por la inflación y las exportaciones, y a través del planteo de hipótesis y del estadístico F o su p-valor, podemos contrastar la hipótesis nula. Las hipótesis establecidas son:

- $H_0: B_1 = B_2 = 0$  Las variables en conjunto **no** tienen una relación lineal con la variable dependiente o estas en conjunto **no** son estadísticamente significativas para explicar a la variable dependiente.
- $H_1: B_1 \neq B_2 \neq 0$ : Las variables en conjunto **si** tienen una relación lineal con la variable dependiente o estas en conjunto **sí** son estadísticamente significativas para explicar a la variable dependiente.



**Figura 15**

*Hipótesis conjunta de la regresión*

. reg PIBUSapreciosactuales Inflaciónpreciosalconsumidor Exportacionesdebienesyservicio						
Source	SS	df	MS	Number of obs = 50		
Model	2.8095e+22	2	1.4047e+22	F(2, 47)	=	26.36
Residual	2.5043e+22	47	5.3282e+20	Prob > F	=	0.0000
				R-squared	=	0.5287
				Adj R-squared	=	0.5087
Total	5.3138e+22	49	1.0844e+21	Root MSE	=	2.3e+10
PIBUSapreciosactuales		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Inflaciónpreciosalconsumidor		-7.59e+08	1.54e+08	-4.92	0.000	-1.07e+09 -4.49e+08
Exportacionesdebienesyservicio		3.05e+09	5.69e+08	5.36	0.000	1.90e+09 4.19e+09
_cons		-1.30e+10	1.30e+10	-1.00	0.324	-3.91e+10 1.32e+10

*Nota. Ochoa, O., 2023.*

Como se puede observar el valor del p-valor de F, es del 0%, lo que es menor al 5%, lo que nos permite rechazar la H<sub>0</sub>, y aceptar que las variables en conjunto explican a al PIB, o que tanto la inflación como las exportaciones tienen una relación lineal conjunta con respecto al PIB.

Revise del punto 4.5 todos los procesos matemáticos y demostrativos con respecto a este tema.

## 4.6 Informe de resultados

Es importante que al realizar un modelo de regresión múltiple se plasmen los resultados obtenidos y se analicen los mismos para determinar la viabilidad y confiabilidad del modelo estimado. Generalmente en modelos de regresión simple o múltiple deberíamos analizar los siguientes resultados obtenidos en una regresión:

- Determinar la relación de las variables independientes con respecto a la variable dependiente, es decir, si estas relaciones son directas o positivas o inversas o negativas; esto lo podemos determinar a través de los signos de los coeficientes beta en la ecuación de regresión obtenida.
- Luego se debe interpretar las estimaciones de las betas, para conocer cómo las variaciones en las variables independientes afectan en promedio



a la variable dependiente. Estas variaciones son los valores de los coeficientes beta.

- Analizar la relación lineal o la significatividad de las variables independientes con respecto a la variable dependiente, de forma individual con base al estadístico  $t$  o de forma conjunta con base al estadístico  $F$ .
- También se debe interpretar el coeficiente de determinación  $R^2$ , para conocer el ajuste del modelo o que tanto explican las variables independientes a la variable dependiente.  $R^2$  alto generalmente mayor al 80 %, indica un buen ajuste de la variable independiente y un  $R^2$  bajo menor al 50 %, un ajuste bajo de la variable independiente con respecto a la dependiente.
- Por último, es importante probar si los residuos de nuestra regresión cumplen con el supuesto de la normalidad de los residuos a través de las distintas pruebas revisadas.



## Actividades de aprendizaje recomendadas

Es momento de aplicar sus conocimientos a través de las actividades que se han planteado a continuación:

1. Reflexione sobre los temas trabajados en la guía virtualizada y en el capítulo 4 del libro de Wooldridge y el capítulo 1 del texto de Meza a partir de la página 51, sobre significancia individual y pruebas de hipótesis.

**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. Por favor, complete las actividades en un cuaderno o documento Word.*

2. Revise los ejercicios 4.9 y 4.10 del texto de Wooldridge.

**Procedimiento:** para entender el tema de inferencia tanto a nivel individual y global, revise los ejercicios propuestos en los cuales se detallan las hipótesis planteadas y los estadísticos utilizados para el contraste de la hipótesis nula.

3. Tome la tabla denominada TWO YEAR. RAW de la página 141 del libro de Wooldridge, y realice en STATA el modelo planteado y valide sus resultados con los del libro.

**Procedimiento:** Cargue la base de datos en STATA, estime el modelo planteado en la página 141 y compare sus resultados con los plasmados en el libro.

4. Desarrolle la autoevaluación de conocimientos al culminar la unidad 4 y las actividades recomendadas en la guía didáctica.

**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. Por favor, complete las actividades en un cuaderno o documento Word.*

## Contenidos, recursos y actividades de aprendizaje recomendadas



### Semana 11

## Unidad 4. Análisis de regresión múltiple: inferencia

### 4.7 Consistencia

Con base en lo que hemos revisado en los temas anteriores, es difícil que los modelos de regresión puedan presentar estimadores insesgados todo el tiempo, algo que no debe quitarnos, el sueño. Pero existen muchos economistas que manifiestan que algo más importante que el insesgamiento, es la consistencia, la cual es un requisito mínimo para un estimador.



En estadística, la consistencia se refiere a una propiedad deseable de los estimadores, que implica que, a medida que el tamaño de la muestra aumenta indefinidamente, el estimador converge al valor real del parámetro que se está estimando. En el contexto de modelos de regresión, la consistencia implica que, a medida que se tienen más datos disponibles, el estimador utilizado para calcular los coeficientes de regresión se acerca cada vez más a los verdaderos valores de los coeficientes. Esto es fundamental para tomar decisiones informadas basadas en los resultados del modelo de regresión y para realizar inferencias válidas sobre la relación entre las variables. Revise la siguiente figura que nos ayuda a entender mejor este tema.

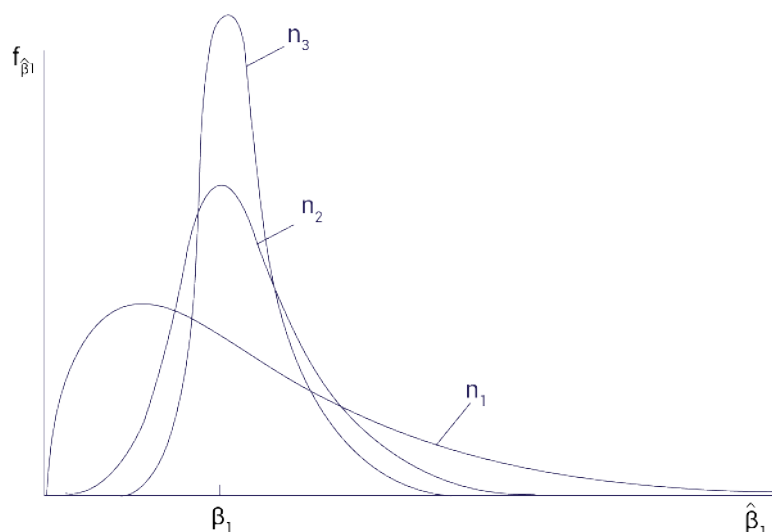
La pérdida de la consistencia o inconsistencia se da cuando existe una correlación entre los residuos ( $u$ ) y cualquiera de las variables independientes ( $X_i$ ), lo que ocasiona la inconsistencia de los estimadores. Por temas prácticos, puede considerarse que la inconsistencia es lo mismo que el sesgo. La diferencia radica en que la inconsistencia se basa en términos de la varianza poblacional de  $X_1$  y de la covarianza poblacional de  $X_1$  y  $X_2$ . En cambio, el sesgo se basa en la muestra, porque se condiciona a los valores de  $X_1$  y  $X_2$  de esta muestra.



## Figura 16

Distribuciones de muestreo de  $\beta_1$ , para distintos tamaños de muestra

Distribuciones de muestreo de  $\hat{\beta}_1$ , para muestras de tamaños  $n_1 < n_2 < n_3$



Nota. Adaptado de *Introducción a la Econometría* (p. 170), por Jeffrey Wooldridge, 2010, Cengage Learning. México, CC BY 4.0.

La figura muestra las distintas distribuciones de muestreo de una variable, ante distintos tamaños de la muestra. Para esto revise el teorema 5.1 de la página 169.

## 4.8 Normalidad asintótica e inferencia con muestras grandes

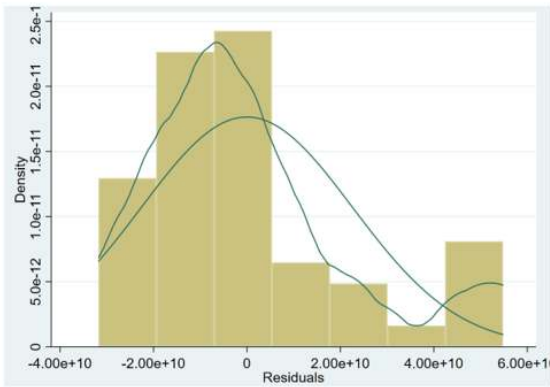
Se parte del supuesto de que, en los modelos clásicos, sus distribuciones de muestreo son normales, lo que es importante para obtener las distribuciones  $t$  y  $F$ , que generalmente son aplicadas para determinar la confiabilidad de un modelo.

Si los errores no presentan una distribución normal, tampoco lo presentarán las betas estimadas en la regresión, lo que afectaría a los estadísticos  $t$  y  $F$ . Esta normalidad no afecta la insesgadez de MCO, ni tampoco afecta a las conclusiones de que MCO, es el mejor estimador lineal insesgado, claro esto bajo los supuestos de Gauss-Markov.

Para entender el tema de la normalidad en modelos de regresión múltiple, apliquemos las pruebas ya conocidas, que aprendimos en los modelos de regresión con un único regresor. Estimemos nuevamente nuestro modelo del PIB, comprobemos si los residuos cumplen la condición de estar normalmente distribuidos.



**Figura 17**  
*Pruebas de normalidad para modelos de regresión múltiple*



```
. histogram RESIDUO, normal kdensity
(bin=7, start=-3.175e+10, width=1.236e+10)

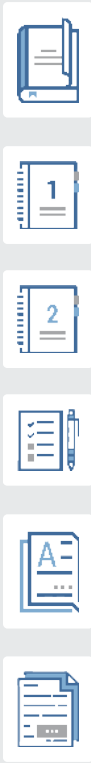
. sktest RESIDUO
```

Skewness/Kurtosis tests for Normality					
Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	Prob>chi2
RESIDUO	50	0.0037	0.2519	8.40	0.0150

```
. jbb RESIDUO
Jarque-Bera normality test:  9.387 Chi(2)  .0092
Jarque-Bera test for Ho: normality:
```

*Nota. Ochoa, O., 2023.*

Como se puede observar en la figura 17, los p-valores del estadístico Chi2, son menores al 5%, lo que nos permite rechazar la H0 de normalidad y aceptar la H1 de que los residuos de la regresión no cumplen con el supuesto de NORMALIDAD, gráficamente también es observable esta situación.



## 4.9 Eficiencia asintótica de MCO

Sabemos que bajo los supuestos de Gauss-Markov, los estimadores de MCO, son los mejores estimadores lineales (MELI) o ELIO. Aquí a lo que hace referencia es que los estimadores de MCO, tienen las menores varianzas asintóticas.

Debe revisar las partes demostrativas de los supuestos para que entienda cómo funciona cada uno de ellos.



### Actividades de aprendizaje recomendadas

Es hora de reforzar los conocimientos adquiridos resolviendo las siguientes actividades:

1. Revise los ejercicios del 5.1 al 5.3 del texto de Wooldridge.

**Procedimiento:** para cada una de las semanas, su docente tutor realizará orientaciones sobre el tema para acercar con mayor claridad los temas expuestos tanto en la guía virtualizada como en el libro. Allí su docente tutor ubicará ejemplos demostrativos para cada uno de los temas que se desarrollan en la semana.

2. Revise las orientaciones desarrolladas por el docente en el aula virtual y participe en la tutoría permanente.

**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. Por favor, complete la actividad en un cuaderno o documento Word.*

3. Desarrolle en STATA el ejercicio C1 de la página 183 del libro de Wooldridge.



**Procedimiento:** para esto, tome los datos de la tabla: WAGE 1. RAW, súbalos a STATA y desarrolle la regresión sugerida.

## Contenidos, recursos y actividades de aprendizaje recomendadas



### Semana 12

## Unidad 4. Análisis de regresión múltiple: inferencia

### 4.10 Efectos del escalamiento de datos sobre los estadísticos de MCO

Aquí se vuelve a revisar un tema que vimos en modelos de regresión simple, en los cuales la transformación de las unidades de medida de las variables podría representar un problema al momento de la interpretación de los resultados de los estimadores. Esto también se lo conoce como reescalar las variables, técnica utilizada muchas veces para maquillar los datos, es decir, suponga que tiene datos como el siguiente: 2300.000.000 millones de dólares. Esta cantidad puede ser maquillada para que ahora se vea así: 2.3 miles de millones de dólares.

Cuando se reescalan las variables, tanto los coeficientes como los errores estándar, los intervalos de confianza, los estadísticos t y F se modifican de una manera tal que se preservan todos los efectos medidos y los resultados de las pruebas. El tema es que esta “transformación de datos”, debe ser analizada con cuidado al momento de las interpretaciones de los resultados. Como puede observar en la tabla 6.1 del texto de Wooldridge, una de las variables independientes como es el peso del recién nacido está medido inicialmente en onzas, pero luego estos datos de peso se transforman a libras, por ende, los resultados de las estimaciones cambian, pero esto no altera las propiedades de eficiencia del modelo. Revise el ejemplo planteado en esta parte del libro de Wooldridge el cual detalla qué sucede con las variables reescaladas.



## 4.11 Más acerca de la forma funcional

En los modelos de regresión simple o con una única variable independiente, vimos que los modelos de regresión se expresaban en forma lineal, traduciendo esta linealidad tanto en las variables, como en los parámetros (Bs). Pero se conoce que muchas teorías económicas no trabajan con modelos lineales en las variables, sino que adoptan otras formas funcionales que se relacionan directamente a las variables que dejan de ser lineales.

Se dice que una variable es lineal cuando esta variable no está ni multiplicada, ni dividida por otra variable o tampoco elevada a un exponente distinto de uno. Cuando las variables dejan de ser lineales, podemos decir que los modelos cambian de forma funcional.

La forma funcional que puede adoptar un modelo de regresión simple es fácil de identificar al graficar la relación entre las dos variables, pero en un modelo de regresión múltiple esto ya no se puede realizar.

Realicemos algunos ejemplos de formas funcionales distintas a la lineal en el caso del modelo del PIB.

### Modelo lineal

$$PIB = \beta_0 + \beta_1 \text{Inflación} + \beta_2 \text{Exportaciones} + u_i \quad (46)$$

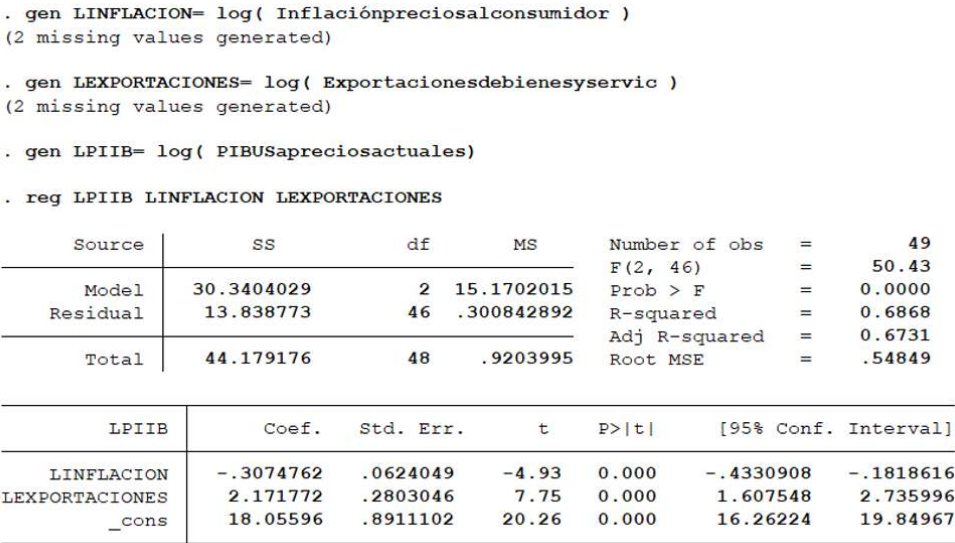
### Modelo logarítmico

$$LPIB = \beta_0 + \beta_1 L\text{Inflación} + \beta_2 L\text{Exportaciones} + u_i \quad (47)$$

Lo que estamos haciendo en el modelo logarítmico es que, a todas las variables del modelo, le sacamos el logarítmico y, una vez realizado esto, estimamos el modelo y analizamos. En STATA se procede de la siguiente forma:



**Figura 18**  
*Modelo de regresión múltiple logarítmico*



Nota. Ochoa, O., 2023.

Para trabajar en este modelo tuvimos que generar nuevas variables que son los logaritmos de las variables PIB, inflación y exportaciones. Una vez realizado esto estimamos el modelo y obtuvimos los resultados de la figura 18.

Cuando las variables se transforman a logaritmos, las unidades de las variables cambian porcentajes, con base a esto, le dejo de tarea que interprete los resultados de los estimadores. 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.  $e^{\beta_1 + \beta_2 X}$  o  $\ln Y = \beta_1 + \beta_2 X$
- Exponencial.  $Y = \beta_1 e^{\beta_2 X}$  o  $\ln Y = \beta_1 + \beta_2 X$

#### 4.12 Más sobre la bondad de ajuste y selección de los regresores

Sabemos que el coeficiente de determinación  $R^2$ , es una medida de bondad de ajuste del modelo, que nos indica que tanto la o las variables independientes explican a la variable dependiente. Sin embargo, a medida que se agregan más variables predictoras al modelo, el  $R^2$  siempre aumenta, incluso si las variables adicionales no son estadísticamente significativas o no aportan información relevante al modelo.

En modelos de regresión múltiple aparece el coeficiente de determinación ajustado ( $\underline{R}_2$ ). Es aquí donde entra en juego el coeficiente de determinación ajustado. Este coeficiente penaliza el número de variables predictoras en el modelo y ajusta el  $R^2$  para tener en cuenta la cantidad de información adicional que aportan estas variables. En otras palabras, el  $\underline{R}_2$  tiene en cuenta tanto la calidad del ajuste como la complejidad del modelo. Realicemos un ejemplo:

Vamos a tomar el modelo del PIB y añadirle algunas variables para ver cómo se comporta el  $\underline{R}_2$ .



**Figura 19**

*Modelo original del PIB en función de la inflación y exportaciones*

```
. reg PIBUSapreciosactuales Inflaciónpreciosalconsumidor Exportacionesdebienesyservic
```

Source	SS	df	MS	Number of obs	=	50
Model	2.8095e+22	2	1.4047e+22	F(2, 47)	=	26.36
Residual	2.5043e+22	47	5.3282e+20	Prob > F	=	0.0000
				R-squared	=	0.5287
				Adj R-squared	=	0.5087
Total	5.3138e+22	49	1.0844e+21	Root MSE	=	2.3e+10

PIBUSapreciosactuales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Inflaciónpreciosalconsumidor	-7.59e+08	1.54e+08	-4.92	0.000	-1.07e+09	-4.49e+08
Exportacionesdebienesyservic	3.05e+09	5.69e+08	5.36	0.000	1.90e+09	4.19e+09
_cons	-1.30e+10	1.30e+10	-1.00	0.324	-3.91e+10	1.32e+10

Nota. Ochoa, O., 2023.

En esta Figura vemos el modelo múltiple inicial y sus valores de  $R^2$  y  $\underline{R}_2$  a este modelo ahora le vamos a incorporar algunas variables sean o no significativas y ver como se comporta el  $\underline{R}_2$ .

**Figura 20**

*Modelo del PIB con más variables independientes*

```
. reg PIBUSapreciosactuales Inflaciónpreciosalconsumidor Exportacionesdebienesyservic Ahorroint  
> nesyservic Inversiónextranjeradirectaen
```

Source	SS	df	MS	Number of obs	=	50
Model	3.4906e+22	5	6.9811e+21	F(5, 44)	=	16.85
Residual	1.8232e+22	44	4.1436e+20	Prob > F	=	0.0000
				R-squared	=	0.6569
				Adj R-squared	=	0.6179
Total	5.3138e+22	49	1.0844e+21	Root MSE	=	2.0e+10

PIBUSapreciosactuales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Inflaciónpreciosalconsumidor	-6.23e+08	1.42e+08	-4.39	0.000	-9.10e+08	-3.37e+08
Exportacionesdebienesyservic	-1.52e+09	1.41e+09	-1.08	0.288	-4.37e+09	1.33e+09
AhorrointernobrutodelPIB	4.80e+09	1.33e+09	3.62	0.001	2.13e+09	7.48e+09
Importacionesdebienesyservic	3.77e+09	1.34e+09	2.81	0.007	1.07e+09	6.47e+09
Inversiónextranjeradirectaen	9.38e+08	3.15e+09	0.30	0.767	-5.41e+09	7.29e+09
_cons	-1.11e+11	3.02e+10	-3.66	0.001	-1.72e+11	-4.99e+10

Nota. Ochoa, O., 2023.

Como se puede observar en la figura 20, los coeficientes  $R^2$  y  $\underline{R}_2$ , ahora son menores a los del modelo original, por tal motivo, el hecho de aumentar variables al modelo no implica que estos coeficientes aumenten en gran medida, a existido un aumento de valor, pero el  $\underline{R}_2$  es aquel que restringe que este valor crezca demasiado.

Para un mejor entendimiento de esta parte, revise los temas adicionales que son complementarios a los que hemos revisado aquí.



### Actividades de aprendizaje recomendadas

Reforcemos el aprendizaje resolviendo las siguientes actividades.

1. Reflexione sobre los temas trabajados en la guía virtualizada y en el capítulo 6 del libro de Wooldridge.

**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.

2. Revise los ejercicios del 6.1 al 6.4 del texto de Wooldridge.

**Procedimiento:** para cada una de las semanas, su docente tutor realizará orientaciones sobre el tema para acercar con mayor claridad los temas expuestos tanto en la guía virtualizada como en el libro. Allí su docente tutor ubicará ejemplos demostrativos para cada uno de los temas que se desarrollan en la semana.

3. Realice el ejercicio 7 de la página 219 del libro de Wooldridge.

**Procedimiento:** para este ejercicio deberá explicar las razones de cuál modelo con base en sus resultados es más apropiado.

*Nota. Por favor, complete las actividades en un cuaderno o documento Word.*



4. Desarrolle la autoevaluación de esta cuarta unidad y las actividades recomendadas en la guía virtualizada.

**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.

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

De acuerdo a la información analizada, lea comprensivamente cada uno de los ítems. Cuando la pregunta sea dicotómica, escriba la respuesta (V/F), y en las preguntas de opción múltiple, elija la respuesta correcta.

1. ( ) ¿Uno de los supuestos de los MCO es la normalidad de los residuos?
2. ( ) ¿Si al modelo clásico de regresión lineal le añadimos el supuesto de heterocedasticidad, tendremos un modelo clásico de regresión lineal normal?
3. ( ) En la regresión múltiple, los residuos no están correlacionados con  $\bar{Y}$ .
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. ( ) ¿El coeficiente de determinación en la regresión múltiple tiene los límites entre 0 y 0.5?
6. ( ) ¿El valor medio de Y estimado es igual al valor medio de Y observado?
7. ( ) ¿El  $R^2$  es una función creciente por el número de variables?
8. ( ) ¿Un modelo polinomial puede ser considerado con un modelo de regresión múltiple?
9. Los coeficientes de correlación entre regresoras y regresores se denomina:
- a. Parámetros.
  - b. Coeficientes de correlación simple.
  - c. Variables.
10. A medida que aumenta el número de regresoras,  $R^2$ :
- a. Aumenta mínimamente.
  - b. Aumenta considerablemente.
  - c. Disminuye considerablemente.

[Ir al solucionario](#)





## Unidad 5. Análisis de regresión múltiple con información cualitativa: variables binarias (o dummies)

### 5.1 Descripción de la información cualitativa

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.

$$\begin{aligned}\text{Consumo} &= f(\text{Ingreso}, \text{raza}) \\ C &= f(Y, \underline{Rz}, Re())\end{aligned}\quad (48)$$

Este modelo ha pasado de ser un modelo de regresión simple a uno 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 = casado, 0 = no casado.

Para la tabla de datos con la que hemos venido trabajando a lo largo de la guía, le vamos a incorporar la variable **dolarización**, con valores de 1 a los años a partir del año 2000 y con 0 a los valores antes del año 2000.

**Figura 21**

*Variable dummy de dolarización*

Inflaciónpr-r	Inversiónex-n	PIBUSaprec-s	Exportacio-c	Ahorrointe-B	Balanzacom-e	Deudaexter-I	Year	Dolarización
22.886133	1.8519303	2.443e+10	21.286903	16.487301	-3.337464	72.317825	1995	0
24.373089	1.9807826	2.523e+10	22.273285	18.868682	.33086669	72.05078	1996	0
30.64298	2.5706578	2.816e+10	21.534926	18.361611	-1.9473339	67.645568	1997	0
36.098433	3.1090816	2.798e+10	17.891951	16.384452	-7.6112062	69.050623	1998	0
52.242351	3.3005905	1.965e+10	26.377542	23.182295	3.5538325	104.15804	1999	0
96.094114	-1.12788994	1.833e+10	32.127535	26.069092	4.7903713	100.4136	2000	1
37.678021	2.2010849	2.447e+10	23.222747	18.04963	-4.299563	76.630638	2001	1
12.484019	2.7435725	2.855e+10	21.492374	17.310577	-6.3918019	71.721568	2002	1
7.92941	2.6871311	3.243e+10	22.5984	17.544723	-2.0448306	64.534085	2003	1
2.7421813	2.2872413	3.659e+10	24.554349	18.642463	-1.5565432	59.343903	2004	1
2.4077698	1.1887461	4.151e+10	27.618174	20.77362	-8.8634815	51.454492	2005	1
3.2987485	.57995085	4.680e+10	30.333075	23.41618	.95636635	45.326894	2006	1
2.2761848	.38008425	5.101e+10	31.931768	23.981225	1.2763975	43.043696	2007	1

Nota. Ochoa, O., 2023.

## 5.2 Una sola variable binaria independiente

Modelos en los cuales la variable dependiente está expresada en forma cuantitativa, y todas las variables independientes en forma cualitativa o como variables dicótomas, la cual adopta la siguiente forma para el caso del ejercicio del PIB:

$$Y = \beta_0 + \beta_1 D_1 + u_i \quad (49)$$

$$PIB = \beta_0 + \beta_1 D_1 (\text{Dolarización}) + u_i$$

Ahora procedemos a realizar la regresión del PIB en función de una variable independiente *dummy* (dolarización).

**Figura 22***Regresión simple con variable dummy*

```
. reg PIBUSapreciosactuales Dolarización
```

Source	SS	df	MS	Number of obs	=	52
Model	3.9411e+22	1	3.9411e+22	F(1, 50)	=	88.59
Residual	2.2242e+22	50	4.4485e+20	Prob > F	=	0.0000
				R-squared	=	0.6392
				Adj R-squared	=	0.6320
Total	6.1653e+22	51	1.2089e+21	Root MSE	=	2.1e+10

PIBUSaprec~s	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Dolarización	5.57e+10	5.92e+09	9.41	0.000	4.38e+10	6.76e+10
_cons	1.53e+10	3.85e+09	3.96	0.000	7.52e+09	2.30e+10

*Nota. Ochoa, O., 2023.*

La interpretación de esta regresión con variable dummy es que después de adoptar la dolarización, el PIB creció en promedio en 5.57e10 dólares más que en los años en los cuales aún existía el sucre. Apparently el proceso de dolarización incentivó el crecimiento del PIB. Además, esta variable **Dolarización**, es estadísticamente significativa y el R2 es del 64%, lo que indica que explica en gran medida los cambios sobre el PIB. Revise los ejemplos del 7.1 al 7.3 del texto de Wooldridge para un mejor entendimiento de este tipo de modelos.

### 5.3 Uso de las variables binarias en categorías múltiples

En la economía no es muy frecuente que se den modelos como los que acabamos de revisar, ya que estos se presentan en áreas como la psicología, sociología, educación e investigación de mercados. La mayor parte de la investigación económica incluye dentro de sus modelos variables explicativas, cuantitativas y cualitativas, ya que permite controlar estadísticamente los efectos de regresoras cuantitativas y cualitativas. Su forma funcional general sería:

$$Y = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 X_i + u_i \quad (50)$$

Para ejemplificar esto tomemos la base de datos del libro de Wooldridge (wage.1), y realicemos un modelo de regresión cuyas variables independientes pueden ser cualitativas y cuantitativas. El modelo que vamos a estimar es el siguiente:

$$\text{Wage} = \beta_0 + \beta_1 D_{1\text{female}} + \beta_2 D_{2\text{married}} + \beta_3 X_{\text{educ}} + \beta_4 X_{\text{exper}} + u_i \tag{51}$$

Los resultados en STATA son los siguientes:

**Figura 23**  
*Modelo de regresión con múltiples variables dummy*

. reg wage educ exper female married						
Source	SS	df	MS	Number of obs	=	526
Model	2261.25908	4	565.31477	F(4, 521)	=	60.12
Residual	4899.15523	521	9.40336896	Prob > F	=	0.0000
				R-squared	=	0.3158
				Adj R-squared	=	0.3105
Total	7160.41431	525	13.6388844	Root MSE	=	3.0665

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.5833233	.051656	11.29	0.000	.4818437	.6848029
exper	.0556664	.0110553	5.04	0.000	.0339479	.0773849
female	-2.067101	.2722077	-7.59	0.000	-2.601861	-1.532342
married	.6602419	.2968513	2.22	0.027	.0770693	1.243414
_cons	-1.790662	.7512121	-2.38	0.017	-3.266439	-.3148853

Nota. Ochoa, O., 2023.

Como se puede observar todas las variables independientes tanto cualitativas como cuantitativas son estadísticamente significativas, el  $R^2$  es del 32%, es decir, que los cambios en el salario (wage), están siendo explicados en un 32% por las variables independientes.

Ahora la interpretación de los estimadores sería que, si la *educación* aumenta en un año, el salario mensual aumentaría en promedio en \$0.58 por hora, manteniendo las demás variables constantes. Respecto a la *experiencia* tenemos que, si esta variable aumenta en un año, el salario mensual en promedio aumentaría en \$0.055 por hora. Por el contrario, la variable *female* nos dice que las mujeres ganan en promedio menos de \$2.06 por hora que los

hombres. Finalmente, el coeficiente de la variable *married* indica que las personas casadas ganan en promedio \$0.66 más por hora que las personas no casadas.

Para un mejor entendimiento de estos temas, revise los temas complementarios estipulados en el punto 7.3.



### Actividades de aprendizaje recomendadas

Es momento de aplicar sus conocimientos a través de las actividades que se han planteado a continuación:

1. Reflexione sobre los temas trabajados en la guía virtualizada y en el capítulo 7 del texto de Wooldridge.

**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.

2. Revise los ejercicios del 7.1 al 7.4 del texto de Wooldridge, sobre modelos con variables *dummy*.

**Procedimiento:** para cada una de las semanas, su docente tutor realizará orientaciones sobre el tema para acercar con mayor claridad los temas expuestos tanto en la guía virtualizada como en el libro. Allí su docente tutor ubicará ejemplos demostrativos para cada uno de los temas que se desarrollan en la semana.

3. Realice el ejercicio 1 de la página 258 del libro de Wooldridge.

**Procedimiento:** para esto, analice los resultados que plantea el ejercicio y conteste los literales propuestos.

*Nota. Por favor, complete las actividades en un cuaderno o documento Word.*





## Unidad 5. Análisis de regresión múltiple con información cualitativa: variables binarias (o dummies)

### 5.4 Interacción en las que intervienen variables binarias

Como lo hemos visto en los temas antes revisados, las variables *dummy* expresan la diferencia que tiene cada una con respecto a sus categorías, pero de forma individual, es decir, la diferencia entre casado o no, o la diferencia entre hombre o mujer, pero estas se interpretan individualmente. Ahora lo que queremos conocer es qué sucede si un individuo cumple dos condiciones de forma simultánea, por ejemplo: hombre-casado o mujer- soltera. Sería apropiado conocer los resultados de estas interacciones, para esto la función general se presenta de la siguiente forma:

$$Y = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 (D_{1i} \cdot D_{2i}) + u_i \quad (52)$$

Adaptando esto al ejercicio anterior, tendríamos la siguiente función de regresión:

$$Wage = \beta_0 - \beta_1 D_{1female} + \beta_2 D_{2married} - \beta_3 X_{educ} + \beta_4 X_{exper} + \beta_1 (D_{1female} * D_{2married}) + u_i \quad (53)$$

Lo que queremos conocer a través de esta regresión es saber cómo se comporta el salario de una **mujer-casada**, respecto a los hombres no casados, que sería la categoría de comparación. En STATA tendríamos lo siguiente:



**Figura 24**  
*Interacción de variables dummy*

```

. gen female_married=( female* married)

. reg wage female married female_married educ exper

```

Source	SS	df	MS	Number of obs	=	526
Model	2483.87804	5	496.775608	F(5, 520)	=	55.24
Residual	4676.53627	520	8.99333898	Prob > F	=	0.0000
				R-squared	=	0.3469
				Adj R-squared	=	0.3406
Total	7160.41431	525	13.6388844	Root MSE	=	2.9989

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
female	-.4237263	.4242263	-1.00	0.318	-1.257134	.4096817
married	2.079997	.4070731	5.11	0.000	1.280287	2.879707
female_married	-2.707349	.5441563	-4.98	0.000	-3.776364	-1.638334
educ	.5781078	.0505281	11.44	0.000	.4788435	.6773721
exper	.053222	.0108228	4.92	0.000	.0319603	.0744837
_cons	-2.655184	.7549211	-3.52	0.000	-4.138254	-1.172114

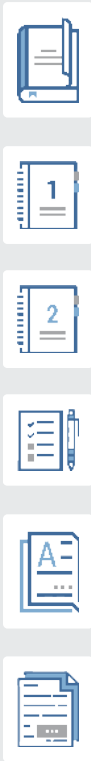
Nota. Ochoa, O., 2023.

Este tipo de modelo de interacción entre variables dummy, nos indica que al menos en la variable (female\_married), su interpretación manifiesta que una **mujer\_casada**, gana en promedio \$2.70 menos por hora con respecto a los **hombres\_no casados**, que vendría a ser la categoría de comparación de esta interacción entre variables. Por ende, si queremos conocer alguna otra diferencia entre los salarios por hora debemos definir la variable interacción con base a los que queremos conocer.

Para un mejor entendimiento de este puede apoyarse del texto de Gujarati en el capítulo 9 que también habla del tema de los modelos de regresión con variables dicótomas.

### 5.5 Más acerca del análisis de políticas y evaluación de programas

Este tema hace referencia a que los modelos con variables *dummy*, generalmente son utilizados para análisis de programas enfocados a políticas públicas, donde existen grupos de tratamiento y de control, los cuales no pueden ser elegidos al azar, sino que deben cumplir con ciertas cualidades o



atributos que las convierten en variables cualitativas o *dummy*. Hay que tener en cuenta que estos tipos de modelos están relacionados con variables dependientes de carácter cualitativo, es decir, una variable dependiente *dummy*. Aquí entran los modelos de respuesta binaria, tema que será revisado con mayor profundidad en la materia de econometría avanzada.



## Actividades de aprendizaje recomendadas

Es hora de reforzar los conocimientos adquiridos resolviendo las siguientes actividades:

1. Revise los ejercicios del 7.8 al 7.11, del texto de Wooldridge, para entender los temas revisados en esta semana.

**Procedimiento:** para cada una de las semanas, su docente tutor realizará orientaciones sobre el tema para acercar con mayor claridad los temas expuestos tanto en la guía virtualizada como en el libro. Allí su docente tutor ubicará ejemplos demostrativos para cada uno de los temas que se desarrollan en la semana.

2. Realice los ejercicios C2 y C3 de la página 262 del libro de Wooldridge.

**Procedimiento:** para esto, utilice las bases de datos: WAGE 2. RAW y MLB1. RAW, corra los modelos y conteste los literales sugeridos.

3. Desarrolle la autoevaluación de esta quinta unidad y las actividades recomendadas en la guía virtualizada.

**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.



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**

De acuerdo a la información analizada, lea comprensivamente cada uno de los ítems, y responda verdadero o falso.

1. ( ) ¿Un modelo con variables dicótomas puede generar heterocedasticidad?
2. ( ) ¿Una variable dicótoma puede tomar valores de 0 o 1?
3. ( ) ¿Una variable dicótoma o dummy es una variable cuantitativa?
4. ( ) ¿La trampa de la variable dicótoma se da cuando se incluye igual número de variables que de categorías?
5. ( ) ¿A las variables dicótomas también se les llama categóricas?
6. ( ) ¿La variable años es una variable dicótoma?
7. ( ) ¿La variable estado civil puede representarse como una variable dicótoma?
8. ( ) ¿Un ejemplo de variable dicótoma es la tasa de crecimiento del PIB?
9. ( ) ¿La predicción de datos con variables dicótomas es menos precisa?
10. ( ) ¿Si quiero aplicar una variable dicótoma para los trimestres del año, debo aplicar 3 variables para no cometer la trampa de la variable dicótoma?





## Unidad 6. Multicolinealidad

### 6.1 Naturaleza de la multicolinealidad

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 ante la presencia de multicolinealidad, los estimadores dejan de ser confiables, lo que generaría que los resultados obtenidos no sean los más adecuados para explicar un fenómeno económico o financiero.

Los casos de multicolinealidad perfecta son casi inexistentes, por lo que en la mayor parte de los modelos de regresión existe multicolinealidad, pero imperfecta, la cual sí nos permite poder obtener los resultados de los coeficientes de regresión.

### 6.2 Consecuencias prácticas de la multicolinealidad

Las consecuencias que se pueden dar son las siguientes:

- Varianzas y covarianzas grandes que presentan problemas de estimación precisa.
- Intervalos de confianza amplios, lo que lleva a una aceptación más fácil de la Hipótesis Nula ( $H_0$ ).
- La razón  $t$  tiende a no ser significativa.



- Los estimadores de Mínimos Cuadrados Ordinarios (MCO) y sus errores pueden ser sensibles a pequeños cambios en la información.

### 6.3 Detección de la multicolinealidad

Para poder detectar si los resultados de una regresión presentan multicolinealidad, se debe tomar en cuenta lo siguiente:

- Cuando se tiene un coeficiente de determinación alto ( $R^2$ ) por encima del 80 % y pocas  $t$  significativas de los coeficientes de regresión, o las probabilidades de estos son mayores al 5 %, hay multicolinealidad.
- El valor de  $F$  es significativo.
- Alta correlación entre parejas de regresores (correlaciones de orden cero), es decir, coeficientes de correlación simple mayores al 80 %. Por ejemplo, la relación entre  $X_2$  y  $X_3$  se denomina  $r_{23}$ .
- Coeficientes de correlación parcial mayores al 80 %. Por ejemplo, si tenemos  $X_2$ ,  $X_3$ , y  $X_4$ , las relaciones que se dan son las siguientes:  $R_{12.34}$ ,  $R_{13.24}$ ,  $R_{14.23}$ .
- Realizar regresiones auxiliares, las mismas que consisten en regresar una variable explicativa en función de otra, por ejemplo,  $X_2 = f(X_3)$ , así de manera sucesiva dependiendo el número de variables independientes existentes en el modelo.
- A través del indicador de Tolerancia (TOL) y del Factor de Inflación de la Varianza (FIV).

Una vez que ya entendemos por qué se da el problema de la multicolinealidad, vamos a poner en práctica esto en el modelo múltiple del PIB.

$$\bar{PIB} = \beta_0 + \beta_1 \text{Inflación} + \beta_2 \text{Exportaciones} + u_i \quad (54)$$

A este modelo vamos a determinar si tiene problemas de multicolinealidad a través de STATA.



**Figura 25**  
*Detección de la multicolinealidad*

```
. reg PIBUSapreciosactuales Inflaciónpreciosalconsumidor Exportacionesdebienesyservic
```

Source	SS	df	MS	Number of obs	=	50
Model	2.8095e+22	2	1.4047e+22	F(2, 47)	=	26.36
Residual	2.5043e+22	47	5.3282e+20	Prob > F	=	0.0000
				R-squared	=	0.5287
				Adj R-squared	=	0.5087
Total	5.3138e+22	49	1.0844e+21	Root MSE	=	2.3e+10

PIBUSapreciosactuales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Inflaciónpreciosalconsumidor	-7.59e+08	1.54e+08	-4.92	0.000	-1.07e+09	-4.49e+08
Exportacionesdebienesyservic	3.05e+09	5.69e+08	5.36	0.000	1.90e+09	4.19e+09
_cons	-1.30e+10	1.30e+10	-1.00	0.324	-3.91e+10	1.32e+10

Nota. Ochoa, O., 2023.

Una de las reglas para saber si hay un problema de multicolinealidad es cuando tenemos un R2 superior al 80% y algunas variables no significativas, en los resultados del modelo del PIB, vemos que esto no se cumple, por ende, podríamos decir que aparentemente no hay problemas de multicolinealidad.

Utilicemos una prueba más formal y metodológicamente más fuerte, como lo es el Factor Inflador de la Varianza (FIV), el cual si presenta valores superiores a 10 podríamos decir que hay problemas de multicolinealidad y si también presenta una Tolerancia (TOL) cercana a cero. Veámoslo en STATA.

**Figura 26**  
*Prueba del factor inflacionario de la varianza (FIV)*

```
. estat vif
```

Variable	VIF	1/VIF
Exportacio~c	1.00	0.999987
Inflaciónp~r	1.00	0.999987
Mean VIF	1.00	

Nota. Ochoa, O., 2023.

Con base a esta prueba podemos ver que el VIF de las variables independientes, es de 1, es decir, no supera el valor de 10 y el valor del TOL ( $1/\text{VIF}$ ), está bastante alejado de cero, por lo que se puede concluir que no hay problemas de **multicolinealidad**.

En el caso de que exista un problema de multicolinealidad en nuestro modelo, existen algunas soluciones que son muy generales a la hora de corregir este problema, entre estas tenemos:

- Eliminar la variable o variables altamente colineales.
- Aumentar el tamaño de la muestra.
- Adoptar una forma funcional distinta.

Estimado estudiante, con esto hemos terminado de revisar los temas correspondientes a la materia de **Econometría Básica**, por lo que lo invito a revisar los temas planteados, los ejemplos sugeridos y a resolver los ejercicios sugeridos en algunas unidades. Recuerde que todo lo aprendido debe ser realizado en el programa STATA.

### Ejercicio práctico en STATA

Para recapitular los contenidos abordados este segundo bimestre, vamos a realizar un ejercicio práctico. La base de datos ([wage1](#)) es tomada del libro de Wooldridge, y está en formato (dta), para que se cargue directamente al STATA. El momento en que haga clic en el archivo de la base de datos, automáticamente se abrirá el programa STATA, y las variables se cargarán automáticamente en el programa.

Vamos a realizar un modelo en el cual el salario (*wage*), sea explicado por las variables: educación (*educ*), experiencia (*exper*), antigüedad (*tenure*), número de dependientes, estas como variables cuantitativas y además las variables femeninas (*female*) y casado (*married*), como variables *dummy*. La información de esta base de datos es de corte transversal, es decir, ha sido generada en un punto específico del tiempo.



Figura 27

Modelo de regresión múltiple del salario

```
. reg wage educ exper tenure numdep female married
```

Source	SS	df	MS	Number of obs	=	526
Model	2651.76052	6	441.960086	F(6, 519)	=	50.87
Residual	4508.65379	519	8.68719421	Prob > F	=	0.0000
				R-squared	=	0.3703
				Adj R-squared	=	0.3631
Total	7160.41431	525	13.6388844	Root MSE	=	2.9474

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.5752695	.0519569	11.07	0.000	.4731978	.6773413
exper	.0218868	.0122498	1.79	0.075	-.0021785	.045952
tenure	.138238	.0211235	6.54	0.000	.09674	.1797361
numdep	.1443296	.1084827	1.33	0.184	-.0687895	.3574488
female	-1.757463	.2665637	-6.59	0.000	-2.28114	-1.233787
married	.4655339	.2942895	1.58	0.114	-.1126111	1.043679
_cons	-2.000828	.777664	-2.57	0.010	-3.528584	-.473072

Nota. Ochoa, O., 2023.

Ahora realizaremos una interpretación de los resultados obtenidos a través de esta regresión múltiple:

- Existe una significatividad conjunta del modelo o las variables independientes en conjunto tiene una relación lineal con la variable dependiente salario (wage). Esto debido a que el p-valor de F, es menor al 5%, lo que nos permite concluir que las variables en su conjunto son adecuadas para explicar a la variable dependiente.
- El coeficiente de determinación ( $R^2 = 0.37$  o 37%), lo que manifiesta que las variables independientes explican en un 37% los cambios promedios que se puedan dar en la variable dependiente (wage). Este coeficiente es una medida de bondad de ajuste del modelo.

Si ahora revisamos la significatividad individual de los coeficientes de regresión a través de sus betas, tenemos lo siguiente:

- Las variables (numdep y married), son no significativas a nivel individual o no tienen ninguna relación lineal con la variable dependiente (wage), ya que

sus p-valores son superiores al 5 %, lo que no nos permite rechazar la hipótesis nula de **no significatividad o no relación lineal**.

- Mientras que las otras variables son estadísticamente significativas, es decir, estas variables son adecuadas para explicar cambios promedio en la variable dependiente (*wage*).

Si interpretamos los coeficientes de regresión individuales o betas, tendríamos lo siguiente:

- **B0:** es la constante o intercepto y que, ante valores de cero en las variables independientes, el salario promedio por hora disminuiría en \$2.0 dólares.
- **B1:** nos dice que, por cada año adicional de educación, nuestro salario aumenta en promedio en \$0.57 dólares por hora, manteniendo las otras variables constantes.
- **B2:** nos dice que, por cada año adicional de experiencia, nuestro salario aumenta en promedio en \$0.021 dólares por hora, manteniendo las otras variables constantes.
- **B3:** nos dice que, por cada año adicional de antigüedad, nuestro salario aumenta en promedio en \$0.14 dólares por hora, manteniendo las otras variables constantes.
- **B4 y B6:** no se interpretan porque son **no** significativas.
- **B5:** nos dice que el salario por hora de una mujer es menor en promedio en \$1.75 dólares, con respecto al de los hombres.

Ante estos resultados, lo mejor sería eliminar del modelo las variables **no** significativas (*numdep* y *married*), y volver a correr el modelo, obteniendo los siguientes resultados:



**Figura 28***Modelo de regresión múltiple del salario corregido*

```
. reg wage educ exper tenure female
```

Source	SS	df	MS	Number of obs	=	526
Model	2603.1066	4	650.77665	F(4, 521)	=	74.40
Residual	4557.30771	521	8.74723169	Prob > F	=	0.0000
				R-squared	=	0.3635
				Adj R-squared	=	0.3587
Total	7160.41431	525	13.6388844	Root MSE	=	2.9576

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.5715048	.0493373	11.58	0.000	.4745803	.6684293
exper	.0253959	.0115694	2.20	0.029	.0026674	.0481243
tenure	.1410051	.0211617	6.66	0.000	.0994323	.1825778
female	-1.810852	.2648252	-6.84	0.000	-2.331109	-1.290596
_cons	-1.567939	.7245511	-2.16	0.031	-2.99134	-.144538

*Nota. Ochoa, O., 2023.*

Como repaso, se le pide que realice una interpretación de estos resultados, como se hizo con el modelo anterior. Es importante que también probemos si el modelo cumple con el supuesto de la **normalidad**, ya que es parte de los supuestos que rigen a los MCO. Para esto debemos generar los residuos de la última regresión, aquella que se estimó sin las variables **no** significativas.



Figura 29

Cálculo de los residuos y prueba de normalidad

```
. predict RESIDUOS, resid
. sktest RESIDUOS
```

Skewness/Kurtosis tests for Normality						
Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	—— joint ——		
				adj	chi2(2)	Prob>chi2
RESIDUOS	526	0.0000	0.0000	.		0.0000

```
. jbr RESIDUOS
Jarque-Bera normality test: 715.6 Chi(2) 4.e-156
Jarque-Bera test for Ho: normality:
```

Nota. Ochoa, O., 2023.

Ante estos resultados podemos manifestar que los residuos de nuestra regresión NO cumplen con el supuesto de **normalidad**, ya que se rechaza la H0: normalidad y se acepta la H1: NO **normalidad**.

A pesar de que el modelo presente variables independientes significativas, un R2=37%, y una relación lineal conjunta con la variable dependiente, esto no deslinda que puedan existir problema de **multicolinealidad**. Para esto vamos a aplicar la prueba del FIV, para saber si nuestro modelo tiene problema de multicolinealidad, los resultados obtenidos son los siguientes:

**Figura 30**  
*Prueba para detectar el problema de la multicolinealidad*

. estat vif

Variable	VIF	1/VIF
exper	1.48	0.675755
tenure	1.40	0.712853
educ	1.12	0.892704
female	1.05	0.950138
Mean VIF	1.26	

*Nota. Ochoa, O., 2023.*

Como se puede observar, los valores FIV de cada variable están por debajo de 10, los valores del TOL (1/VIF), están alejados de cero, por ende, podemos concluir que el modelo no tiene problemas de **multicolinealidad**.

Con este ejercicio damos por terminado la revisión de los contenidos correspondientes al segundo bimestre.

*¡Éxitos!*



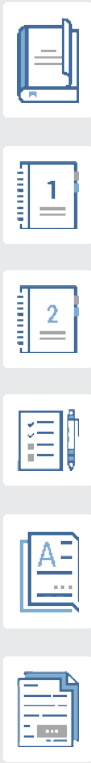
### Actividades de aprendizaje recomendadas

Reforcemos el aprendizaje resolviendo las siguientes actividades.

1. Revise el capítulo 10 del texto de Gujarati.

**Procedimiento:** para que comprenda el tema de la multicolinealidad, le aconsejo que realice un resumen donde detalle por qué se produce el problema de la multicolinealidad, los métodos para detectar este problema y cuáles serían las posibles soluciones.

2. Realice los ejercicios 10.3 y 10.27 del texto de Gujarati, el último hágalo en STATA.



**Procedimiento:** para cada una de las semanas, su docente tutor realizará orientaciones sobre el tema para acercar con mayor claridad los temas expuestos tanto en la guía virtualizada como en el libro. Allí su docente tutor ubicará ejemplos demostrativos para cada uno de los temas que se desarrollan en la semana.

*Nota. Por favor, complete las actividades en un cuaderno o documento Word.*

3. Desarrolle la autoevaluación de esta sexta unidad y las actividades recomendadas en la guía virtualizada.

**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.

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

De acuerdo a la información analizada, lea comprensivamente cada uno de los ítems, cuando la pregunta sea dicotómica escriba la respuesta (v/f) y en las preguntas de opción múltiple elija la respuesta 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. ¿Cuándo la multicolinealidad es leve o moderada se pueden presentar los siguientes problemas?
- a. Existe coeficientes de determinación muy elevados.
  - b. Existen  $t$  muy significativas.
  - c. Existe  $\chi^2$  muy elevados.
4. ¿Qué métodos se puede utilizar para la detección de multicolinealidad?
- a. Prueba de White.
  - b. Correlograma.
  - c. Factor inflador de la varianza.
5. ¿Cómo se puede solucionar la multicolinealidad?
- a. Mejorar el diseño muestral extrayendo la información máxima.
  - b. Eliminar datos sin importar el número de observaciones.
  - c. Utilizar un modelo econométrico lineal simple.
6. Existe multicolinealidad si en la matriz de correlaciones parciales, el coeficiente de correlación entre  $X_2$  y  $X_3$  es:
- a. 0.95.
  - b. 0.09.
  - c. 0.
7. El término MELI significa:
- a. Modelo de estimación lineal.



- b. Mejor estimador lineal insesgado.
- c. Método de estimación insesgada.

8. ¿En qué tipo de regresión se puede hacer el examen de correlaciones parciales?

- a. Regresión simple con una variable independiente.
- b. Regresión múltiple.
- c. Regresión simple con variable logarítmica.

9. ¿Podría eliminarse la multicolinealidad con?

- a. Incrementar las variables explicativas.
- b. Combinar información de corte transversal y serie temporal.
- c. Colocar una variable dicótoma.

10. Si la relación  $r_{23}$  es de 0.20, ¿existe multicolinealidad o no?

- a. Existe multicolinealidad.
- b. No existe multicolinealidad.
- c. Ninguno de los 2 casos anteriores.

[Ir al solucionario](#)

**Contenidos, recursos y actividades de aprendizaje recomendadas**



## Semana 16

### Actividades finales del bimestre

Estimado estudiante, en esta semana se va a realizar un repaso de las unidades estudiadas en el segundo bimestre.

- **Unidad 4.** Análisis de regresión múltiple: inferencia.
- **Unidad 5.** Análisis de regresión múltiple con información cualitativa: variables binarias (o *dummy*).



• **Unidad 6.** Multicolinealidad.



Debe revisar nuevamente los contenidos de las unidades del 4 al 6, para un reforzamiento de los temas y como preparación previa al examen presencial.



## Actividades de aprendizaje recomendadas

Es momento de aplicar sus conocimientos a través de las actividades que se han planteado a continuación:

1. Tome la tabla de datos 10.14, del libro de Gujarati, y replique todos los procesos que se realizaron en el ejercicio final propuesto en la semana 15.

**Procedimiento:** para este, cargue la base de datos 10.14, en STATA, y, replique los procesos realizados en el ejercicio final propuesto.

2. Revise las orientaciones desarrolladas por el docente en el aula virtual y participe en la tutoría permanente.

**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. Por favor, complete las actividades en un cuaderno o documento Word.*





## 4. Autoevaluaciones

### Autoevaluación 1

Pregunta	Respuesta	Retroalimentación
1	c	La econometría es usar la matemática y estadística para explicar teorías o relaciones económicas.
2	b	Son datos de individuos, empresas o segmentos de la economía.
3	a	La pregunta de interés nos ayuda a probar una teoría económica, la cual será expresada de forma cuantitativa a través de un modelo econométrico.
4	v	Esta variable constituye una variable independiente que explicará a la variable dependiente.
5	f	Se transforma a un modelo econométrico.
6	a	Es la información que se da en un punto específico del tiempo.
7	b	El VAB provincial corresponde a información de la producción recopilada en el año 2022.
8	b	Es una combinación de información de corte transversal y serie temporal.
9	a	Información que se genera en distintos periodos de tiempo es información de serie de tiempo.
10	b	Un modelo econométrico busca que la variable independiente explique a la variable dependiente mientras las otras variables permanecen constantes.

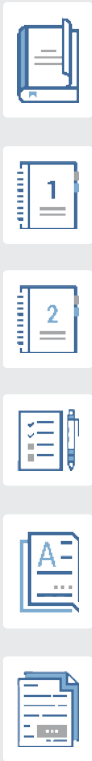
[Ir a la autoevaluación](#)



Autoevaluación 2

Pregunta	Respuesta	Retroalimentación
1	a	Una regresión lineal simple es aquella regresión que solo tiene una variable independiente.
2	c	La variable dependiente se la denomina también variable regresada, variable endógena, variable predictora.
3	b	Es la pendiente que mide los cambios que se dan en la variable dependiente.
4	c	Este estadístico nos dice qué tanto la variable independiente explica a la variable dependiente.
5	v	Los errores son una medida de dispersión de las observaciones en torno a la recta de regresión.
6	b	Los datos muy altos o bajos pueden generar estimaciones incorrectas.
7	c	La hipótesis puede ser contrastada con el valor de estadístico o con su p-valor.
8	f	Las propiedades no se ven alteradas, solo los valores de los coeficientes para su interpretación.
9	v	El logarítmico o semilogarítmico son otras formas funcionales de los modelos de regresión distinta a la lineal.
10	f	Siempre se debe mantener la linealidad en los parámetros.

Ir a la autoevaluación



### Autoevaluación 3

Pregunta	Respuesta	Retroalimentación
1	b	Un modelo múltiple es aquel que tiene más de una variable independiente.
2	a	Lo que busca es minimizar los errores para que las estimaciones sean más confiables.
3	v	Sí, porque miden los efectos individuales de cada variable independiente sobre la variable dependiente.
4	f	Este efecto se refiere a la interpretación de los estimadores de B0 y B1.
5	a	Este modelo en su ecuación carece del intercepto o constante.
6	a	Un modelo mal especificado puede ser por estar sobreestimado.
7	v	Cuando este supuesto no se cumple, hablamos de un problema de heterocedasticidad.
8	v	Depende de la $\sigma^2$ , $STC$ , $R^2$ .
9	a	El error de especificación se da por omitir una variable importante.
10	c	Al existir una alta colinealidad entre variables independientes, no deja explicar el aporte de cada una de forma individual sobre la variable dependiente.

[Ir a la autoevaluación](#)



## Autoevaluación 4

Pregunta	Respuesta	Retroalimentación
1	v	Este supuesto es importante y es parte de la metodología de MCO.
2	f	Se tiene un modelo de regresión lineal normal.
3	f	Sí, están correlacionados con los residuos.
4	c	Es una metodología usada para modelos de regresión al igual que la de MCO.
5	f	Este indicador tiene límites entre 0 y 1.
6	f	De esa forma, la estimación de los parámetros es eficiente.
7	f	El cambio en $R^2$ depende más de la relación que puedan ejercer las variables independientes sobre la dependiente.
8	v	Porque, además de sus variables lineales, también está la variable cuadrática.
9	c	Esto nos indica la asociación entre pares de variables.
10	a	El hecho de tener varias variables independientes hará variar mínimamente el coeficiente de determinación.

[Ir a la autoevaluación](#)



Autoevaluación 5

Pregunta	Respuesta	Retroalimentación
1	v	Es uno de los problemas de trabajar con variable dicótoma.
2	v	Con estos valores se vuelve cuantificable a la variable.
3	f	Es una variable cualitativa.
4	v	Sí, porque se produce una perfecta colinealidad.
5	v	Sí, porque asignan categorías a la variable cualitativa.
6	f	Es una variable de carácter cuantitativo.
7	v	Es una variable que puede tomar valores de 1 o 0 dependiendo de la categoría.
8	f	Porque es una variable numérica cuantificable.
9	f	Está casi al mismo nivel de las variables cuantitativas.
10	f	Se debe incluir (m-1) variables dicótomas por categorías.

Ir a la autoevaluación



## Autoevaluación 6

Pregunta	Respuesta	Retroalimentación
1	b	Alta relación entre variables independientes.
2	b	La variabilidad extrema entre datos puede producir este problema.
3	a	Un $R^2$ elevado es un síntoma de un problema de multicolinealidad.
4	c	El FIV, conjuntamente con la tolerancia, determina el problema de multicolinealidad.
5	a	Aumentar la muestra o mejorar la recolección de información pueden corregir el problema.
6	a	Valores superiores a 0.8 son considerados un problema.
7	b	A través de la metodología de MCO se pueden obtener los MELI.
8	b	Las correlaciones se dan entre pares de variables independientes.
9	b	Combinar la información disminuye la posible asociación entre variables.
10	b	Es un valor bajo de relación entre variables que no es preocupante.

[Ir a la autoevaluación](#)





## 5. Referencias bibliográficas

- Ochoa, O. (2023). Guía didáctica de Econometría Básica. Ediloja. UTPL
- Gujarati, D. & Porter, D. (2010). Econometría. México: McGraw Hill. 5.<sup>a</sup> Edición.
- MEZA CARVAJALINO, C. A. Econometría. 1. ed. Bogotá: Universidad de La Salle - Ediciones Unisalle, 2022. 339 p. Disponible en: <https://elibro.net/es/ereader/bibliotecautpl/222212?page=52>. Consultado en: 10 Dec 2024.
- Stock, J. & Watson, M. (2020). Introducción a la econometría. 4.<sup>a</sup> edición. Pearson Education. Madrid.
- Wooldridge, J. (2015). Introducción a la econometría. 5.<sup>a</sup> edición. Cengage Learning. México.





---

## 6. Anexos

---



## Anexo 1. Base de datos del PIB

IPC (% anual)	IED_ENC (% del PIB)	PIB (US\$)	XBS (% del PIB)	AIB (% del PIB)	BCBS (% del PIB)	DEA (% del INB)	Year
5.1	3.1	2862504170	9.4	16.0	-3.6	14.2	1970
8.4	5.9	2754220263	10.6	16.3	-6.1	15.9	1971
7.9	2.5	3185987235	12.4	20.1	-3.1	17.1	1972
13.0	1.3	3891755552	17.9	21.7	1.2	15.2	1973
23.3	1.2	6599259421	22.7	28.5	3.9	11.0	1974
15.4	1.2	7731677257	16.1	17.3	-3.3	12.9	1975
10.7	-0.2	9091924305	16.0	19.1	-1.9	13.6	1976
13.0	0.3	11026346590	16.3	22.7	-1.8	23.0	1977
11.7	0.4	11922502171	14.7	22.8	-3.8	35.1	1978
10.3	0.4	14175166008	17.9	22.5	-0.3	33.7	1979
13.0	0.4	17881514683	17.1	24.6	-0.7	35.4	1980
16.4	0.3	21810767209	14.6	25.1	-0.7	36.9	1981
16.3	0.2	19929853575	14.7	18.7	-3.1	41.2	1982
48.4	0.3	17152483214	15.6	20.9	0.2	47.3	1983
31.2	0.3	16912515183	17.4	20.5	1.7	53.1	1984
28.0	0.4	17149094590	19.7	22.6	3.6	54.6	1985
23.0	0.5	15314143988	15.6	21.7	-1.9	66.0	1986
29.5	0.9	13945431882	15.4	18.4	-7.0	81.9	1987
58.2	1.2	13051886552	19.0	21.5	-4.7	91.5	1988
75.6	1.2	13890828708	20.3	22.5	-4.5	90.8	1989
48.5	0.8	15239278100	22.8	25.0	0.9	86.7	1990
48.8	0.9	16988535268	23.7	24.2	2.2	78.9	1991
54.3	1.0	18094238119	24.1	26.1	3.5	71.5	1992
45.0	2.5	18938717359	20.0	18.0	-3.1	78.1	1993
27.4	2.5	22708673337	20.3	18.1	-3.2	83.2	1994
22.9	1.9	24432884442	21.3	16.5	-3.3	72.3	1995
24.4	2.0	25226393197	22.3	18.9	0.3	72.1	1996
30.6	2.6	28162053027	21.5	18.4	-1.9	67.6	1997
36.1	3.1	27981896948	17.9	16.4	-7.6	69.1	1998
52.2	3.3	19645272636	26.4	23.2	3.6	104.2	1999
96.1	-0.1	18327764882	32.1	26.1	4.8	100.4	2000
37.7	2.2	24468324000	23.2	18.0	-4.3	76.6	2001
12.5	2.7	28548945000	21.5	17.3	-6.4	71.7	2002
7.9	2.7	32432858000	22.6	17.5	-2.0	64.5	2003
2.7	2.3	36591661000	24.6	18.6	-1.6	59.3	2004
2.4	1.2	41507085000	27.6	20.8	-0.9	51.5	2005
3.3	0.6	46802044000	30.3	23.4	1.0	45.3	2006
2.3	0.4	51007777000	31.9	24.0	1.3	43.0	2007
8.4	1.7	61762635000	34.2	26.7	0.3	31.5	2008
5.2	0.5	62519686000	25.2	24.0	-1.6	21.8	2009
3.6	0.2	69555367000	27.9	23.5	-4.5	22.4	2010
4.5	0.8	79276664000	31.1	25.9	-2.2	20.7	2011
5.1	0.6	87924544000	30.2	26.4	-1.4	18.8	2012
2.7	0.8	95129659000	28.6	26.1	-2.3	20.3	2013
3.6	0.8	1.01726E+11	28.1	26.7	-1.6	24.9	2014
4.0	1.3	99290381000	21.3	24.1	-2.7	29.1	2015
1.7	0.8	99937696000	19.5	25.5	0.5	36.1	2016

IPC (% anual)	IED_ENC (% del PIB)	PIB (US\$)	XBS (% del PIB)	AIB (% del PIB)	BCBS (% del PIB)	DEA (% del INB)	Year
0.4	0.6	1.04296E+11	20.8	25.5	-0.8	40.4	2017
-0.2	1.3	1.07562E+11	22.6	25.6	-1.2	42.9	2018
0.3	0.9	1.07436E+11	23.4	25.0	0.0	49.5	2019