UNIDAD 3

Regresión múltiple

PAPIME
PE300520

Coordinador
Ing. Alejandro Augusto Pérez Pascual

Participantes
Lic. Claudia Jacqueline Bribiesca Silva,
Mtro. Ignacio Cruz López
Mtra. Hortensia Martínez Valdez
Dr. Jorge Pablo Rivas Díaz

Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional

Contenido

1. EVALUACIÓN DIAGNÓSTICA 3

2. PRESENTACIÓN DE LA UNIDAD 4

3. UNIDAD 3 - ACTIVIDAD 1 6

4. UNIDAD 3 - ACTIVIDAD 2 8

5. UNIDAD 3 - ACTIVIDAD 3 9

6. UNIDAD 3 - ACTIVIDAD 4 11

7. UNIDAD 3 - ACTIVIDAD 5 13

8. UNIDAD 3 - ACTIVIDAD 6 15

9. FORMULARIO 16

10. NOTAS DE APOYO 20

11. AUTOEVALUACIÓN 23

1. EVALUACIÓN DIAGNÓSTICA

Lee con atención y responde si los siguientes enunciados son verdaderos o falsos.

1. La econometría se puede entender como el análisis cuantitativo de los fenómenos económicos reales, basado en el análisis simultáneo de teoría y comportamiento de la información empírica.

   Verdadero Falso

2. Pueden ser consideradas como las cuatro etapas fundamentales en la elaboración de un modelo econométrico:

   1) La Especificación, 2) Estimación, 3) Validación y 4) Predicción.

   Verdadero Falso

3. En el contexto de un modelo de regresión, una predicción se puede entender como: la solución Y que el modelo proporciona para un valor dado (o conjunto de valores dados) de las variables explicativas X.

   Verdadero Falso

4. El Método de Mínimos Cuadrados Ordinario (MCO) sirve para determinar los estimadores cuyo error cuadrático es mínimo dentro de una relación económica dada.

   Verdadero Falso

5. La relación económica más sencilla que se puede establecer en un modelo econométrico es la ecuación lineal entre dos variables, siendo de la forma: $y = \beta_0 + \beta_1 x + u$.

   Verdadero Falso

6. El valor de r^2 dentro del contexto del modelo de regresión lineal simple, calculado por MCO indica el grado de ajuste de la regresión, donde valores más bajos o negativos indican una mejor capacidad predictiva del modelo de forma global.

   Verdadero Falso

2. PRESENTACIÓN DE LA UNIDAD

OBJETIVO PARTICULAR

Al finalizar el estudio de la unidad 3, el estudiante será capaz de aplicar y analizar los modelos económicos de análisis de regresión múltiple, así como interpretar las fluctuaciones de los coeficientes de correlación y regresión para hacer predicciones de variables económicas, apoyado en el uso de software especializado.

TEMARIO

3.1 Modelos con dos variables explicativas.
3.2 Pruebas de significancia de los coeficientes.
3.3 Interpretación de los coeficientes de regresión.
3.4 Correlación parcial y múltiple.
3.5 Predicción.
3.6 Análisis de varianza y pruebas de hipótesis.
3.7 Grados de libertad y R ajustada.
3.8 Pruebas de estabilidad.
3.9 Pruebas de LR y W.
3.10 Alcances y limitaciones: análisis de resultados.
3.11 Aplicaciones a la economía.

PRESENTACIÓN

En esta unidad se estudiarán los fundamentos de los modelos económicos de análisis de regresión múltiple, es decir, de los modelos de una ecuación de regresión con más de una variable explicativa, de manera que la variable dependiente es la única variable endógena de la ecuación.

El análisis de regresión múltiple ha sido sumamente difundido para la investigación y modelación de fenómenos y procesos económicos, formando parte central en la práctica profesional del economista, la curricular escolar y los libros de texto, ya que permite dotar nuevas herramientas a la economía, para adquirir, ampliar y consolidar conocimientos que implican más solidez para las conclusiones, razonamientos empíricos y las recomendaciones de política que se pueden obtener del procesamiento de datos reales.

La modelación micro y macroeconómica permite mostrar empíricamente los argumentos teóricos, a partir de la recopilación y tratamiento de los datos con los que se cuenta en un momento determinado, fortaleciendo la capacidad para validarlos y apoyando y facilitando su interpretación actual.

Modelos con múltiples variables explicativas.

Introducción

En la regresión multivariada lo que se busca es analizar la relación entre una variable dependiente Y y un conjunto de variables explicativas de carácter independiente $X_1, X_2, \dots, X_k$.

En economía es sumamente relevante dicho análisis ya que rara vez las relaciones económicas son de carácter simple, por lo general, existen múltiples causas que explican un mismo fenómeno o proceso económico.

Algunos ejemplos:

A nivel macroeconómico, por ejemplo, el crecimiento económico puede estar siendo explicado por el nivel de ingresos familiares, la cantidad de empresas, el desarrollo tecnológico alcanzado, y el tipo de cambio, por mencionar algunas variables. El desempeño de una reforma a la política fiscal puede ser determinado por la política de gasto, la magnitud de la recaudación, los ingresos corrientes provenientes de los impuestos sobre la renta, sobre el valor agregado, la evasión fiscal en el pago de impuestos u otras tantas variables.

A nivel microeconómico, el desempeño en las ventas de las empresas, por ejemplo, puede estar siendo determinado por el ingreso disponible del consumidor, los precios del producto ofertado, los precios de los bienes sustitutos y el ingreso disponible de los consumidores.

Para estimar un modelo económico que explique dichas relaciones con confiabilidad y precisión, se busca que los estimadores de MCO cuenten con un error estándar (ee) mínimo, un coeficiente de determinación R^2 alto y que cada variable integrante del modelo de regresión tenga un impacto justificado y relevante al momento de explicar el comportamiento de la variable dependiente.

La validez de un modelo de regresión múltiple puede ser afectada por diversas causas, desde los problemas de especificación, cuando no se integran las variables adecuadas en el modelo, los problemas de medición de las variables y los problemas de estimación del modelo, lo que requiere de someter al modelo de manera conjunta y a sus distintos estimadores a diversas pruebas de hipótesis, intervalos de confianza y en general a distintos procedimientos de cálculo del valor de estadísticos de prueba para aceptar o rechazar valor del de los estimadores y del modelo.

Se sigue la línea explicativa de la unidad anterior, sobre el modelo de mínimos cuadrados ordinarios (MCO) y el cálculo de los mejores estimadores linealmente insesgados (MELI) en los modelos de regresión lineal, por lo que es fundamental haber superado la unidad II del curso para abordar la actual.

Para comprender los temas de la unidad es necesario que retomes literatura básica y complementaria del curso y los aprendizajes de la unidad anterior. En lo particular se sugiere la lectura detallada de los siguientes contenidos:

- Gujarati, D (2010) Capítulo 7 Análisis de regresión múltiple: el problemas de estimación, y Capítulo 8 Análisis de regresión múltiple: el problema de la inferencia. En Econometría. México: McGraw-Hill.
- Maddala, G.S (1996) Capítulo 4 Regresión Múltiple. En Introducción a la Econometría. México: Prentice Hall.
- Wooldridge, Jeffrey M. (2015) Capítulo 3 Análisis de regresión múltiple: Estimación y Capítulo 4 Análisis de regresión múltiple: Inferencia. En Introducción a la Econometría. Quinta Edición. Cengage.

Posteriormente realiza las siguientes actividades siguiendo al pie de la letra las instrucciones.

3. UNIDAD 3 - ACTIVIDAD 1

Modelos con múltiples variables explicativas

En esta actividad se describirán los elementos básicos en la construcción del modelo de regresión múltiple como herramienta para el análisis econométrico.

Objetivo: Explicar los procedimientos para obtener los parámetros y coeficientes del modelo de regresión múltiple.

Instrucciones: considere solo uno de los siguientes textos:

- Gujarati, D (2010) Capítulo 7 Análisis de regresión múltiple: el problemas de estimación, y Capítulo 8 Análisis de regresión múltiple: el problema de la inferencia. En Econometría. México: McGraw-Hill.
- Maddala, G.S (1996) Capítulo 4 Regresión Múltiple. En Introducción a la Econometría. México: Prentice Hall.
- Wooldridge, Jeffrey M. (2015) Capítulo 3 Análisis de regresión múltiple: Estimación y Capítulo 4 Análisis de regresión múltiple: Inferencia. En Introducción a la Econometría. Quinta Edición. Cengage.

1. **Elabora un texto descriptivo** con la siguiente estructura:

   A. Introducción: redacta en un párrafo el tema que será descrito (Elementos para el análisis de regresión múltiple).
   B. Desarrollo: en esta sección responde las siguientes preguntas:

      i. ¿Qué es un modelo de regresión múltiple?
      ii. ¿Cómo se determinan los estimadores en el modelo de regresión múltiple?
      iii. ¿Cómo se interpretan los coeficientes de regresión?
      iv. ¿Cuáles son las pruebas de significancia de los coeficientes y cuáles son sus parámetros de decisión?
      v. ¿Qué representa y cómo se estiman los coeficientes de correlación parcial $r_{12.3}; r_{13.2}\ y\ r_{23.1}$ y la correlación múltiple $R_{1.23}$ en los modelos de regresión múltiple?

   C. Conclusión: con tus propias palabras explica para qué sirve la estimación de un modelo de regresión múltiple.

2. El texto debe tener la estructura de un trabajo académico: carátula, Introducción, desarrollo, conclusiones, citas, notas, referencias y fuentes de consulta en formato APA.

3. La extensión máxima del texto es de 10 cuartillas.

4. Revisa los criterios de evaluación para considerarlos en el desarrollo de tu actividad.

Instrumento de evaluación:

#	Criterio	Puntaje	Cumple	
			si	no
1	Aborda todos puntos y preguntas dispuestas en las instrucciones	2		
2	Explica cómo se integra el modelo de regresión múltiple y lo presenta en su forma matemática general	1.5		
3	Explica cómo se determina la confiabilidad o precisión de los estimadores del modelo de regresión múltiple	1.5		
4	Explica cómo se interpretan los coeficientes en los modelos de regresión múltiple de manera generalizada.	1.5		
5	Explica cuáles son las pruebas de significancia de los coeficientes y cuáles son sus parámetros de decisión	1.5		
6	Explica cómo se determinan y para qué sirven los coeficientes de correlación parcial ($r_{12.3}; r_{13.2}$ y $r_{23.1}$) y múltiple ($R_{1.23}$)	1		
7	Recupera información de fuentes adicionales	0.35		
	La actividad cumple el formato solicitado (máximo 12 diapositivas)	0.35		
8	Se envía la actividad en tiempo	0.30		
	Puntaje total	10		

4. UNIDAD 3 - ACTIVIDAD 2

Ejercicio de estimación de regresión múltiple con dos variables

Objetivo: Aplicar el procedimiento para obtener la función de regresión múltiple con dos variables regresoras.

Instrucciones:

1. Instala el programa de procesamiento econométrico E-Views, R, R Studio o en su defecto utiliza Excel.
2. Descarga el archivo titulado “U3_Ejercicio_A”, considera la tabla de consumo real, ingreso y riqueza real de para la economía de los Estados Unidos del año 1947 a 2000.
3. Con dicha información determina lo siguiente:

   a) Estima la función de consumo lineal usando los datos de ingreso y riqueza.

   b) ¿Qué indican los coeficientes estimados sobre las relaciones entre las variables y el gasto de consumo?

   c) Calcula e interpreta los valores de los coeficientes de determinación R^2 y de correlación múltiple R del modelo.

   d) Calcula e interpreta el valor de la R^2 ajustada.

4. Revisa los criterios de evaluación para considerarlos en el desarrollo de tu actividad.

Instrumento de evaluación:

#	Criterio	Puntaje	Cumple	
			si	no
1	Estima la función de consumo lineal	2		
2	Responde la pregunta b	2		
3	Obtiene e interpreta los valores del coeficiente de determinación R^2 y de correlación múltiple R del modelo	3		
4	Obtiene e interpreta el valor de la R^2 *ajustada*	3		
	Puntaje total	10		

5. UNIDAD 3 - ACTIVIDAD 3

Ejercicio de estimación de regresión múltiple

Objetivo: Aplicar el procedimiento para obtener la función de regresión múltiple e interpretar los parámetros y estimadores obtenidos de dicho proceso

Instrucciones:

1. Instala el programa de procesamiento econométrico E-Views, R, R Studio o en su defecto utiliza Excel.
2. Descarga el archivo titulado "U3_Ejercicio_B, considera la tabla que proporciona datos sobre los componentes de la Demanda Agregada para la economía de los Estados Unidos del año 2000 a 2019.
3. Elabora un texto descriptivo con la siguiente estructura:

   A. Introducción: redacta en un párrafo el tema que será descrito.

   B. Desarrollo: en esta sección desarrolla las siguientes tareas:

      i. Introduce los datos de la tabla directamente al programa EViews, Excel o mediante programación en R-Studio.

      ii. Agrupa todas las variables y grafica el comportamiento de las series y describe lo que se puede observar.

      iii. Especifica y estima el modelo de regresión múltiple representativo del modelo de la demanda agregada con economía abierta: $Y = C + I + G + (SBC)$
      donde SBC = Saldo de la Balanza Comercial (X – M)

      iv. Grafica por separado las relaciones de *Y* respecto del resto a cada una de sus variables explicativas *X* y describe lo que se puede observar.

      v. Calcula e interpreta los valores de los coeficientes de determinación R^2, de correlación múltiple R y la $R^2 ajustada$ del modelo

   C. Conclusión: con tus propias palabras explica lo que se puede entender del modelo de regresión múltiple desarrollado, donde se ofrezca una reflexión económica relevante sobre los hallazgos encontrados a la luz de la teoría económica básica.

5. Revisa los criterios de evaluación para considerarlos en el desarrollo de tu actividad.

Instrumento de evaluación:

#	Criterio	Puntaje	Cumple	
			si	no
1	Aborda todos puntos y preguntas dispuestas en las instrucciones	2		
2	Logró introducir y maneja las series de datos en EViews, R o Excel.	1		
3	Obtiene la función de regresión múltiple solicitada.	2		
4	Gráfica y describe las relaciones parciales de la regresión	1		
5	Obtiene e interpreta los valores de los coeficientes de determinación R^2, de correlación R y la $R^2\text{ajustada}$ del modelo	1		
6	Obtiene e interpreta los valores de los coeficientes β del modelo de regresión	1		
7	Ofrece una reflexión económica relevante sobre los hallazgos encontrados mediante el modelo a la luz de la teoría económica básica	2		
	Puntaje total	10		

Nota: Deberás guardar los resultados para completar las siguientes actividades.

6. UNIDAD 3 - ACTIVIDAD 4

Inferencia con modelos de regresión múltiple

En esta actividad se describirán los elementos básicos para la inferencia con modelos de regresión múltiple para el análisis econométrico.

Objetivo: Explicar los procedimientos y pruebas asociados a la inferencia con modelos de regresión múltiple.

Instrucciones: considera uno de los siguientes textos:

- Gujarati, D (2010) Capítulo 7 Análisis de regresión múltiple: el problemas de estimación, y Capítulo 8 Análisis de regresión múltiple: el problema de la inferencia. En Econometría. México: McGraw-Hill.
- Maddala, G.S (1996) Capítulo 4 Regresión Múltiple. En Introducción a la Econometría. México: Prentice Hall.
- Wooldridge, Jeffrey M. (2015) Capítulo 3 Análisis de regresión múltiple: Estimación y Capítulo 4 Análisis de regresión múltiple: Inferencia. En Introducción a la Econometría. Quinta Edición. Cengage.

1. Elabora un texto descriptivo con la siguiente estructura:

A. Introducción: redacta en un párrafo el tema que será descrito (Inferencia en el análisis de regresión múltiple).

B. Desarrollo: en esta sección responde las siguientes preguntas:

i. ¿En qué consiste la labor de inferencia asociada al modelo de regresión múltiple?

ii. ¿Qué es y cuál es la función del Análisis de Varianza y las Pruebas de Hipótesis en la labor de inferencia de los modelos de regresión múltiple?

iii. ¿Qué son, cómo se estiman y para qué sirve la $R^2 ajustada$ y los grados de libertad en el proceso de especificación e inferencia del modelo de regresión múltiple?

iv. ¿Cuál es la función del análisis de estabilidad en la capacidad de inferencia de los modelos de regresión múltiple?

v. ¿Cuáles son las principales pruebas de estabilidad y cómo se estiman?

C. Conclusión: con tus propias palabras explica para qué sirve en general la capacidad de inferencia de los modelos de regresión múltiple dentro del campo de estudio de la economía.

2. El texto debe tener la estructura de un trabajo académico: carátula, Introducción, desarrollo, conclusiones, citas, notas, referencias y fuentes de consulta en formato APA.

3. La extensión máxima del texto es de 10 cuartillas.

4. Revisa los criterios de evaluación para considerarlos en el desarrollo de tu actividad.

Instrumento de evaluación:

#	Criterio	Puntaje	Cumple	
			si	no
1	Aborda todos puntos y preguntas dispuestas en las instrucciones	2		
2	Explica en qué consiste la labor de inferencia	1.5		
3	Explica el papel del análisis de Varianza y las Pruebas de Hipótesis en el modelo de regresión múltiple	1.5		
4	Explica la función y cálculo de la R^2ajustada y los grados de libertad.	1.5		
5	Explica la función del análisis de estabilidad y las pruebas existentes para evaluarla	1.5		
6	Propone una reflexión clara acerca de la función de la inferencia para el análisis de los fenómenos económicos	1		
7	Recupera información de fuentes adicionales	0.35		
	La actividad cumple el formato solicitado (máximo 12 diapositivas)	0.35		
8	Se envía la actividad en tiempo	0.30		
	Puntaje total	10		

7. UNIDAD 3 - ACTIVIDAD 5

Ejercicio de inferencia de regresión múltiple

En esta actividad se aplicará el conocimiento adquirido para el análisis de inferencia en regresión múltiple.

Objetivo: Aplicar el conocimiento e instrumental tecnológico para el análisis de inferencia de modelos de regresión múltiple

Instrucciones:

1. Instala el programa de procesamiento econométrico EViews, R, R Studio o en su defecto utiliza Excel.
2. Descarga el archivo titulado “U3_Ejercicio_B”, considera la tabla que proporciona datos sobre los componentes de la Demanda Agregada para la economía de los Estados Unidos del año 2000 a 2019 y los resultados de estimación del modelo de regresión realizado en el ejercicio 3.
3. Elabora un texto con la siguiente estructura:

   A. Introducción: redacta en un párrafo el tema que será descrito (Análisis de inferencia y estabilidad de la economía norteamericana).
   B. Desarrollo: en esta sección desarrolla las siguientes tareas:

      i. Realiza una breve descripción de la regresión múltiple representativa del PIB ¿Qué es lo que observas?
      ii. Calcula y realiza el análisis de varianza.
      iii. Prueba la capacidad de predicción de la ecuación de regresión múltiple y argumenta si el modelo resulta útil para la inferencia.

   C. Conclusión: con tus propias palabras explica: 1) si puedes predecir el desempeño de la economía de los Estados Unidos a futuro, dados los resultados de estimación e inferencia del modelo; 2) argumenta cuál es la utilidad de elaborar análisis derivados de modelos de regresión múltiple, para el análisis de la dinámica económica
   D. Anexo: En un anexo presenta el abanico de pruebas de estabilidad en los modelos de regresión múltiple, enlista las pruebas, describe las generalidades de su cálculo y describe las condiciones en que se ocupa cada una de ellas. Prueba de análisis de varianza (Chow), y pruebas predictivas para estabilidad (Prueba LR y W)

4. Revisa los criterios de evaluación para considerarlos en el desarrollo de tu actividad.

Instrumento de evaluación:

#	Criterio	Puntaje	Cumple	
			si	no
1	Aborda todos puntos y preguntas dispuestas en las instrucciones	2		
2	Realiza un análisis de la regresión múltiple	1		
3	Realizó el cálculo y análisis de varianza y la $R^2 ajustada$	2		
4	Probó la capacidad predictiva y estabilidad del modelo	2		
5	Realizó un análisis sobre la posibilidad de inferir el desempeño de la economía norteamericana	2		
6	Presentó el anexo con la descripción metodológica de las pruebas de estabilidad solicitadas	1		
	Puntaje total	10		

8. UNIDAD 3 - ACTIVIDAD 6

Mapa mental

En esta actividad se busca realizar un ejercicio de reflexión acerca de la aplicación, alcances y limitaciones del modelo de regresión múltiple al análisis económico.

Objetivo: Integrar el instrumental del análisis, estimación e inferencia del modelo de regresión múltiple al análisis económico actual, contextualizándolo en sus alcances, limitaciones y aplicaciones para la práctica profesional del economista.

Instrucciones:

1. **Elabora un mapa mental** con la idea central del "*análisis de regresión múltiple*", de donde se desprendan tanto los conceptos de carácter matemático-econométrico, como los elementos de aplicación al análisis económico, integrando los siguientes elementos:

   A. Estimación e Inferencia de los modelos y sus conceptos auxiliares
   B. Alcances prácticos de los modelos de regresión múltiple
   C. Limitaciones de los modelos para el análisis económico
   D. Potencial de aplicación para el análisis económico
   E. Ejemplos

2. Puedes ocupar cualquier programa gratuito online para la construcción de mapas mentales: Bubble.us**,** MindMeister**,** Mindomo**,** Mind42**,** Wisemapping**,** Text 2 Mind Map**,** Litpen, u otros.
3. Revisa los criterios de evaluación para considerarlos en el desarrollo de tu actividad.

Instrumento de evaluación:

#	Criterio	Puntaje	Cumple	
			si	no
1	Aborda todos puntos dispuestos en las instrucciones	2		
2	Desglosa elementos en estimación e inferencia y sus conceptos auxiliares	1		
3	Desglosa elementos en alcances prácticos de los modelos de regresión múltiple	1		
4	Desglosa elementos en limitaciones de los modelos para el análisis económico	1		
5	Presenta cruces entre elementos de trasfondo matemático econométrico con los de aplicación económica dando noción de potencial de aplicación para el análisis económico	2		
6	Presenta ejemplos para favorecer los argumentos que presenta en el mapa mental	2		
7	Hace uso de bibliografía complementaria	1		
	Puntaje total	10		

9. FORMULARIO

Función de regresión poblacional: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$

Media condicional: $E(Y_i|X_{2i}, X_{3i}) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i}$

Función de regresión muestral: $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i$

Suma de cuadrados de los residuos (SCR):

$$\sum \hat{u}_i^2 = \sum \left(Y_i - \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}\right)^2$$

Coeficientes de regresión:

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3$$

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

$$\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

Varianza y error estándar de los estimadores:

$$var(\hat{\beta}_1) = \left[\frac{1}{n} + \frac{\bar{X}_2^2 \sum x_{3i}^2 + \bar{X}_3^2 \sum x_{2i}^2 - 2\bar{X}_2 \bar{X}_3 \sum x_{2i} x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i} x_{3i})^2}\right] \sigma^2$$

$$ee(\hat{\beta}_1) = +\sqrt{var(\hat{\beta}_1)}$$

$$var(\hat{\beta}_2) = \left(\frac{\sum x_{3i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}\right) \sigma^2 = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{23}^2)}$$

$$ee(\hat{\beta}_2) = +\sqrt{var(\hat{\beta}_2)}$$

$$var(\hat{\beta}_3) = \left(\frac{\sum x_{2i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}\right) \sigma^2 = \frac{\sigma^2}{\sum x_{3i}^2 (1 - r_{23}^2)}$$

$$ee(\hat{\beta}_3) = +\sqrt{var(\hat{\beta}_3)}$$

Donde r_{23} es el coeficiente de correlación muestral entre X_2 y X_3

$Coeficiente\ de\ determinación$: $r_{23}^2 = \frac{(\sum x_{2i} x_{3i})^2}{\sum x_{2i}^2 \sum x_{3i}^2}$

$Coeficiente\ de\ correlaxión\ muestral\ entre\ X_2\ y\ X_3$:

$$r_{23} = \sqrt{\frac{(\sum x_{2i} x_{3i})^2}{\sum x_{2i}^2 \sum x_{3i}^2}}$$

$Varianza\ de\ las\ perturbaciones\ poblacionales$: $\sigma^2 = \dfrac{\sum \hat{u}_i^2}{n-3}$

$Covarianza\ de\ los\ estimadores$: $cov(\hat{\beta}_2, \hat{\beta}_3) = \dfrac{-r_{23}\sigma^2}{(1-r_{23}^2)\sqrt{\sum x_{2i}^2}\sqrt{\sum x_{3i}^2}}$

$Coeficiente\ de\ determinación\ múltiple$:
$R^2 = \dfrac{SCE}{SCT} = \dfrac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{\sum y_i^2}$

$Coeficiente\ de\ correlación\ múltiple$: $R = \sqrt{R^2}$

$\bar{R}^2\ ajustada$: $\bar{R}^2 = 1 - \dfrac{\sum \hat{u}_i^2 / (n-k)}{\sum \hat{y}_i^2 / (n-1)}$

Coeficiente de correlación parcial entre Y y X_2, manteniendo X_3 constante:

$r_{12.3} = \dfrac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}$

Coeficiente de correlación parcial entre Y y X_3, manteniendo X_2 constante:

$r_{13.2} = \dfrac{r_{13} - r_{12}r_{23}}{\sqrt{(1-r_{12}^2)(1-r_{23}^2)}}$

Coeficiente de correlación parcial entre X_2 y X_3, manteniendo Y constante:

$r_{23.1} = \dfrac{r_{23} - r_{12}r_{13}}{\sqrt{(1-r_{12}^2)(1-r_{13}^2)}}$

Donde: $r_{12} = \dfrac{\sum y_i x_{2i}}{\sqrt{\sum y_i^2 \sum x_{2i}^2}}$ $r_{13} = \dfrac{\sum y_i x_{3i}}{\sqrt{\sum y_i^2 \sum x_{3i}^2}}$ $r_{23} = \dfrac{\sum x_{2i} x_{3i}}{\sqrt{\sum x_{2i}^2 \sum x_{3i}^2}}$

$Coeficiente\ de\ correlación\ múltiple\ R_{1.23}$:

$R_{1.23} = \sqrt{\dfrac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1-r_{23}^2}}$

$Coeficiente\ de\ determinación\ múltiple\ R_{1.23}^2$:

$R_{1.23}^2 = (R_{1.23})^2 = \dfrac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1-r_{23}^2}$

Donde:

$$r_{12}^2 = \left(\frac{\sum y_i x_{2i}}{\sqrt{\sum y_i^2 \sum x_{2i}^2}}\right)^2 \qquad r_{13}^2 = \left(\frac{\sum y_i x_{3i}}{\sqrt{\sum y_i^2 \sum x_{3i}^2}}\right)^2 \qquad r_{23}^2 = \left(\frac{\sum x_{2i} x_{3i}}{\sqrt{\sum x_{2i}^2 \sum x_{3i}^2}}\right)^2$$

Pruebas de significancia: $H_0: \beta_2 = \beta_3 = 0$

$$F = \frac{\left(\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}\right)/2}{\sum \hat{u}_i^2 /(n-3)} = \frac{SCE/gl}{SCR/gl} = \frac{SCE/(k-1)}{SCR/(n-k)}$$

Si $F > F_{\propto}(k-1, n-k)$, se rechaza $H_0: \beta_2 = \beta_3 = 0$; de lo contrario, no se rechaza.

Si el $valor\ p\ de\ F$ de bastante bajo, se puede rechazar $H_0: \beta_2 = \beta_3 = 0$

Tabla ANOVA:

Origen de la variación	SC	gl	SCM
Debido a la regresión SCE	$\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}$	2	$\frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{2}$
Debido a los residuos SCR	$\sum \hat{u}_i^2$	$n-3$	$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-3}$
Total	$\sum y_i^2$	$n-1$	

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

Si $F > F_{\propto}(k-1, n-k)$, se rechaza $H_0: \beta_2 = \beta_3 = 0$; de lo contrario, no se rechaza.

Tabla ANOVA en términos de R^2:

Origen de la variación	SC	gl	SCM
Debido a la regresión	$R^2\left(\sum y_i^2\right)$	2	$\frac{R^2(\sum y_i^2)}{2}$
Debido a los residuos	$(1-R^2)\left(\sum y_i^2\right)$	$n-3$	$\frac{(1-R^2)(\sum y_i^2)}{(n-3)}$
Total	$\sum y_i^2$	$n-1$	

Contribución incremental de X_3 después de considerar la contribución de X_2:

$$F = \frac{Q_2/_{gl}}{Q_4/_{gl}} = \frac{\frac{(SCE_{nueva} - SCE_{vieja})}{número\ de\ regresoras\ nuevas}}{\frac{SCR_{nueva}}{(n - \#de\ parámetros\ en\ el\ nuevo\ modelo)}}$$

Tabla ANOVA para evaluar la contribución incremental:

Origen de la variación	SC	gl	SCM
SCE debido a X_2	$Q_1 = \hat{\beta}_{12}^2 \sum x_2^2$	1	$\frac{Q_1}{1}$
SCE debido a la adición de X_3	$Q_2 = Q_3 - Q_1$	1	$\frac{Q_2}{1}$
SCE debido tanto a X_2 como a X_3	$Q_3 = \hat{\beta}_2 \sum y_I x_{2i} + \hat{\beta}_3 \sum y_I x_{3i}$	2	$\frac{Q_3}{2}$
SCR	$Q_4 = Q_5 - Q_3$	$n - 3$	Q_4
Total	$Q_5 = \sum y_i^2$	$n - 1$	$n - 3$

10. NOTAS DE APOYO

Modelos con múltiples variables explicativas.

En la regresión multivariada lo que se busca es analizar la relación entre una variable dependiente Y y un conjunto de variables explicativas de carácter independiente $X_1, X_2, \dots, X_k$. En economía es sumamente relevante dicho análisis ya que rara vez las relaciones económicas son de carácter simple, por lo general, existen múltiples causas que explican un mismo fenómeno o proceso económico.

A nivel macroeconómico, por ejemplo, el crecimiento económico puede estar siendo explicado por el nivel de ingresos familiares, la cantidad de empresas, el desarrollo tecnológico alcanzado, y el tipo de cambio, por mencionar algunas variables. El desempeño de una reforma a la política fiscal puede ser determinado por la política de gasto, la magnitud de la recaudación, los ingresos corrientes provenientes de los impuestos sobre la renta, sobre el valor agregado, la evasión fiscal en el pago de impuestos u otras tantas variables.

A nivel microeconómico, el desempeño en las ventas de las empresas, por ejemplo, puede estar siendo determinado por el ingreso disponible del consumidor, los precios del producto ofertado, los precios de los bienes sustitutos y el ingreso disponible de los consumidores.

El modelo general supuesto para esto es el siguiente:

$$Y_i = a + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, \qquad i = 1,2, \dots, n$$

Donde los errores u_i se deben los errores de medición en Y a la especificación de la relación entre Y y las X, haciendo las mismas suposiciones que se hicieron en el análisis de regresión simple, donde es posible demostrar que el método de Mínimos Cuadrados proporciona estimadores $a, \beta_1, \beta_2, \dots, \beta_k$, insesgados y con varianza mínima entre la clase de estimadores lineales insesgados.

Interpretación de los coeficientes de regresión.

En la ecuación de regresión múltiple con dos variables explicativas, por ejemplo, X_1 y X_2 se busca hablar del efecto conjunto de X_1 y X_2 y el efecto parcial de X_1 o X_2 sobre Y, de tal manera que la ecuación de regresión se puede escribir de la siguiente manera:

$$Y - \bar{Y} = \hat{\beta}_1 (X_1 - \bar{X}_1) + \hat{\beta}_2 (X_2 - \bar{X}_2) + \hat{u}_i$$

Donde el efecto parcial de X_1 se mide por $\hat{\beta}_1$ y el efecto parcial de X_2 por $\hat{\beta}_2$, siendo el efecto parcial el que se obtiene al mantener constante la otra variable.De esta forma, los coeficeintes $\hat{\beta}_1$ y $\hat{\beta}_2$ deben interpretarse como una medición del efecto de X_1 o X_2 sobre Y después de eliminar el efecto de la otra variable.

Expresado de otra forma, los coeficientes de regresión $\beta_1, \beta_2, \dots, \beta_k$, se conocen también como los coeficientes de regresión parcial o coeficientes parciales de pendiente, donde β mide el cambio en el valor de la media de $Y, E(Y)$por unidad de cambio en$X$, permaneciéndolas demás X constantes. Es, por tanto, el “efecto neto” o “efecto directo” que una unidad de cambio de X tiene sobre el valor medio de Y.

Correlación parcial y múltiple.

Si se cuenta con una regresión múltiple con tres variables explicativas, por ejemplo, X_1 , X_2 y X_3 y una variable explicada Y, entonces:

- r_{y1}^2, r_{y2}^2; r_{y3}^2 son los cuadrados de las *correlaciones simples* entre Y y X_1 , X_2 y X_3; y miden la proporción de la varianza en Y que explican X_1 , X_2 y X_3 por si solas.
- R_{y123}^2 mide la proporción de la varianza de Y que explican X_1 , X_2 y X_3 en forma *conjunta*.
- $r_{y2.1}^2$ y $r_{y3.12}^2$ son los coeficientes parciales de determinación que explican cuanto afecta en la regresión, X_2 después de incluir a X_1 y cuanto afecta X_3 después de incluir a X_1 y X_2, siendo la primera una correlación parcial de primer orden $(r_{y2.1}^2)$ y la segunda de segundo orden $(r_{y3.12}^2)$

Como se puede observar, la forma usual de denotar las correlaciones simples y parciales es r, mientras que las correlaciones múltiples se denotan con R, donde la r^2 parcial puede ser mayor o menor que la r^2simple, pero r^2 simple y r^2 parcial nunca pueden ser mayores que R^2.

Predicción.

Las fórmulas de predicción en regresión múltiple son sumamente similares a las del caso de la regresión simple, excepto por el hecho de que para calcular el error estándar del valor predicho se necesitan las varianzas y covarianzas de todos los coeficientes de regresión.

Grados de libertad y R ajustada.

Si se tiene n observaciones y se estiman tres parámetros de regresión múltiple:

$$Y_i = a + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i, \qquad i = 1,2, \dots, n$$

Es posible observar que los residuos estimados $\hat{u}_i$, satisfacen tres restricciones lineales:

$$\sum \hat{u}_i = 0, \sum_{X_i} \hat{u}_i = 0, \quad \sum_{X_{2i}} \hat{u}_i = 0,$$

Es decir, que sólo existen $(n-3)$ residuos que pueden variar, ya que, dado cualquiera $n-3$ residuos, es posible obtener los tres residuos restantes.

Dado que, el estimado de la varianza residual σ^2 esta dada por $\hat{\sigma}^2 = \frac{RSS}{Grados\ de\ Libertad}$, a medida que el número de variables explicativas aumenta, la RSS se reduce, pero también se presenta una reducción en los grados de libertado, siendo por tanto que llegara el punto en que σ^2 crecerá a medida que se incorporen más variables explicativas.

De esta forma, en regresión múltiple se acostumbra reportar el $R^2 ajustada$, denotada como $\bar{R}^2$, que es simplemente, la $R^2 ajustada$ de acuerdo a los grados de libertad $1 - \bar{R}^2 = \frac{n-1}{n-k-1}(1 - R^2)$ donde k es el número de regresores

Pruebas de estabilidad.

Para probar la capacidad de predicción futura de una ecuación de regresión múltiple, se supone que los parámetros son constantes a través del tiempo (en todo momento), durante la estimación y predicción, siendo necesario probar dicha hipótesis de comportamiento (estabilidad) mediante las pruebas: a) prueba de análisis de varianza y b) pruebas predictivas

a) Prueba de análisis de varianza (Chow):

Suponiendo dos conjuntos de datos independientes de tamaño de muestra n_1 y n_2 con ecuaciones de regresión

$$Y = a_1 + \beta_{11}X_1 + \beta_{12}X_2 + \cdots + \beta_{1k}X_k + u, \qquad \text{(Conjunto 1)}$$

$$Y = a_2 + \beta_{21}X_1 + \beta_{22}X_2 + \cdots + \beta_{2k}X_k + u, \qquad \text{(Conjunto 2)}$$

Siendo que el primer subíndice de β denota el conjunto de datos y el segundo la variable, la prueba de estabilidad entre las poblaciones de ambos conjuntos es una prueba de hipótesis:

$$Ho: \beta_{11} = \beta_{21},\ \beta_{12} = \beta_{22},\ \ldots,\ \beta_{1k} = \beta_{2k},\ a_1 = a_2$$

De tal manera que si Ho es posible estimar una sola ecuación para el conjunto de ambos grupos de datos.

b) Pruebas predictivas para estabilidad

Suponiendo dos conjuntos de datos independientes de tamaño de muestra n_1 y n_2. Para el caso en que $n_2 < (k-1)$, se utilizan las primeras n_1 observaciones para calcular la regresión y para realizar predicciones para las siguientes n_2 observaciones. Se prueba que la hipótesis de que los errores de la predicción tienen media cero.

Pruebas de LR y W.

En el modelo de regresión múltiple, para probar la hipótesis $\beta_i = 0$ se utilizan las estadísticas de prueba LR, W, LM para muestras grandes, con la estadística $r^2\ parcial$ correspondiente en el lugar de la $r^2\ simple$, donde la prueba tiene una distribución x^2 con k grados de libertad.

- $LR = n\ log_e(\frac{RRSS}{URSS})$
- $W = \frac{RRSS - URSS}{URSS/n}$
- $LM = \frac{RRSS - URSS}{URSS/n}$

Donde:
RRSS = Suma de cuadrados residuales restringida
URSS= suma de cuadrados residuales no restringida.

La prueba de verosimilitud LR tiene una distribución x^2 con k grados de libertad y k número de restricciones.

Las pruebas LR y W tienen una distribución x^2 con k grados de libertad, cumpliendo la desigualdad $W \geq LR \geq LM$

11. AUTOEVALUACIÓN

Lee con atención y responde si los siguientes enunciados son verdaderos o falsos.

1. La modelación econométrica consiste en un proceso de abstracción de los fenómenos económicos complejos para facilitar su comprensión, el estudio de su comportamiento y pronosticarlos en tiempo y circunstancias determinadas.

   Verdadero Falso

2. La ecuación de regresión multivariada o con más de una variable explicativa, donde la variable dependiente es la única variable endógena, se expresa con la siguiente manera:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + \beta_n X_{nt} + u_t \quad , \qquad t = 1,2,\dots,n$$

$$Y_t = \sum_{i=1}^{k} \beta_i X_{it} + u_t \quad , \qquad t = 1,2,\dots,n$$

   Verdadero Falso

3. Son dos razones causantes de una especificación incorrecta de un modelo de regresión multivariado:
   a) Omitir una variable explicativa relevante
   b) Incluir una variable explicativa irrelevante

   Verdadero Falso

4. En la fórmula de correlación parcial, $r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$, se observa que cuando no hay correlación entre *Y* y *X,* es decir $r_{12} = 0$, la correlación parcial entre *Y* y X_2, manteniendo X_3 constante, solo puede ser cero cuando r_{13} o r_{23} o ambos sean iguales a cero.

   Verdadero Falso

5. En el modelo de regresión múltiple cuanto mayor es la distancia entre x_o de $\bar{x}$ la varianza de la predicción se incrementa.

   Verdadero Falso

6. En la conclusión de la prueba de Chow, si el valor de *F* calculado excede el valor crítico obtenido de la tabla *F* en el nivel de significancia elegido, no se rechaza la hipótesis nula de estabilidad paramétrica.

   Verdadero Falso

7. Para probar la estabilidad de las relaciones sostenidas en una regresión múltiple de manera eficiente se debe elegir entre la prueba de análisis de varianza (AV) y la prueba predictiva (prueba de Chow)

   Verdadero Falso