ECONOMETRIA: PRIMERA PARTE

Table of Contents

[2 Sumatoria 3](#_Toc145955867)

[3 Esperanza 4](#_Toc145955868)

[4 Covarianza y Correlación 4](#_Toc145955869)

[4.1 Simetría 6](#_Toc145955870)

[4.2 Mismo signo 6](#_Toc145955871)

[4.3 Unidades de medida 6](#_Toc145955872)

[4.4 Domino entre -1 y 1 6](#_Toc145955873)

[4.5 = 1 si X Función Lineal Exacta de Y 7](#_Toc145955874)

[5 Supuestos Clásicos 7](#_Toc145955875)

[5.1 Linealidad 7](#_Toc145955876)

[5.2 X no aleatoria 7](#_Toc145955877)

[5.3 Exogeneidad 7](#_Toc145955878)

[5.4 Homocedasticidad 8](#_Toc145955879)

[5.5 No correlación serial 8](#_Toc145955880)

[5.6 No multicolinealidad perfecta 8](#_Toc145955881)

[6 Propiedades del MCO 8](#_Toc145955882)

[6.1 Suma de los residuos es 0 8](#_Toc145955883)

[6.2 Suma del producto de los residuos con x es 0 8](#_Toc145955884)

[6.3 Medias iguales 9](#_Toc145955885)

[6.4 La recta pasa por las medias 9](#_Toc145955886)

[6.5 Otra formulación 9](#_Toc145955887)

[6.6 Otra formulación 9](#_Toc145955888)

[7 Insesgadez en MCO 10](#_Toc145955889)

[7.1 AMELIA 10](#_Toc145955890)

[7.2 WOOLDRIDGE 10](#_Toc145955891)

[7.3 INTERCEPTO 11](#_Toc145955892)

[8 Varianza en MCO 11](#_Toc145955893)

[9 R-Cuadrado 12](#_Toc145955894)

[10 Modelo sin intercepto 12](#_Toc145955895)

[11 Regresión múltiple 12](#_Toc145955896)

[11.1 Teorema de Frisch-Waugh-Lovell (Partialling-out) 13](#_Toc145955897)

[11.2 Relación entre estimadores simples y múltiples 14](#_Toc145955898)

[12 Variables binarias 14](#_Toc145955899)

[13 Relaciones funcionales 15](#_Toc145955900)

[13.1 Logarítmicas 15](#_Toc145955901)

[13.2 Cuadráticas 16](#_Toc145955902)

[14 Inferencia 17](#_Toc145955903)

[15 Forma matricial 17](#_Toc145955904)

# Sumatoria

**1.**

**2.**

**3.**

**4.**

Demostración:

**5.**

Demostración:

**6.**

Demostración:

**7. (generalización de 6.)**

Demostración:

**8. (implicaciones de 7.)**

Demostración:

# Esperanza

1.

2-

3.

Demostración:

4.

:

5.

6.

# Covarianza y Correlación

Covarianza poblacional:

Demostración:

Por linealidad de la esperanza:

Las esperanzas son números, no son variables aleatorias,

Covarianza muestral:

Insesgadez de la covarianza :

Demostración:

Correlación muestral:

## Simetría

Demostración: conmutatividad de la multiplicación

## Mismo signo

son positivas, entonces el cambio en los signos de la correlación debe ser por la covarianza

## Unidades de medida

Demostración:

Demostración:

## Domino entre -1 y 1

Desigualdad de Cauchy-Schwarz:

## = 1 si X Función Lineal Exacta de Y

Si , entonces la correlación queda:

# Supuestos Clásicos

## Linealidad

Es lineal en los parámetros (no en las variables).

## X no aleatoria

## Exogeneidad

## Homocedasticidad

Demostración: por y por exogeneidad

## No correlación serial

Demostración: , por exogeneidad la esperanza de los errores es 0,

## No multicolinealidad perfecta

Para que exista un estimador de beta tiene que las x no ser todas iguales. No hay relaciones lineales exactas entre las variables independientes: mirar *cada* una de las variables independientes contra cada una de las otras. Si una variable independiente del modelo es combinación lineal de otra, tiene colinealidad perfecta. Las variables pueden estar correlacionadas, pero no pueden tener correlación perfecta. Se tiene que cumplir que

# Propiedades del MCO

## Suma de los residuos es 0

Esto también es la contraparte muestral del supuesto de media de los residuos 0.

Es una condición de primer orden al minimizar SRC.

## Suma del producto de los residuos con x es 0

Esto también es la contraparte muestral del supuesto de media condicional de los residuos 0.

Demostración: es 0 por la condición anterior.

## Medias iguales

Demostración:

## La recta pasa por las medias

## Otra formulación

## Otra formulación

Considerando a como una constante para los índices de arriba,

Cada es una ponderación de las *y*-es.

## Diferencia entre error y residuo

el modelo poblacional es , con el error de la obsercacion i. tambien se puede expresar en terminos de su valor ajustado y residuo en . El error aparece en la ecuacion con los parametros poblacionales y ; los residuos aparecen con los estimadores. los errores son inobervables, los residuos se computan de los datos.

Se puede escribir el residuo en función del error

La esperanza del valor esperado de cada parámetro beta estimado es beta, pero no es lo mismo que : la diferencia tiene esperanza 0

# Insesgadez en MCO

## AMELIA

## WOOLDRIDGE

Usando la propiedad de que

Ahora hay que ver a beta como una variable aleatoria ya que queremos analizar su comportamiento en cualquier muestra. Beta era:

En términos del coeficiente poblacional y errores (poblacionales), sustituyendo, queda:

Dado que

queda que beta es

Nota: la insesgadez es una propiedad de la distribución muestral de las betas, no de los estimadores que se obtengan para alguna muestra en particular. Si la muestra que tenemos es 'típica' entonces el estimado debería ser cercano al valor poblacional.

Entonces:

Asumiendo que las *x* son no-aleatorias

Usando el supuesto de exogeneidad

Supuestos usados:

- *x* e *y* estar relacionados linealmente

- *x* no aleatorias

- Exogeneidad

### Caso múltiple

variación total de J

r-cuadrado de la regresión de en todas las demás variables independientes incluyendo el intercepto.

## INTERCEPTO

# Varianza en MCO

## AMELIA

Por la propiedad de que : para cualquier muestra

Usando

Usando no correlación serial :

Usando homocedasticidad :

## WOOLDRIDGE

varianza constante. si quisiéramos saber la varianza del estimador de beta

El error tiene la misma varianza para cualquier valor de x

y entonces

Con homocedasticidad la varianza de u condicional en x NO depende de x

la esperanza condicional de y dado x es lineal en x, pero la varianza de y dado x es constante

Varianza muestral de MCO

Como no se puede conocer el , se estima

## Estimadores de la Varianza

Ahora bien, sabiendo que por lo que una estimación insesgada seria : , pero no se puede estimar porque los errores son inobservados. pero los u se pueden estimar con los residuos. Entonces si se usa el cual ES un estimador poque es computable para cualquier muestra de x e y.... pero es sesgado, (no tan sesgado si n es grande). De todas formas, se puede ibtener uno insesgado:

El estimador SSR/n es sesgado porque no contempla las condiciones de primer orden de MCO:

Pero si conocemos n-2 residuos se pueden obtener los otros dos usando las restricciones que implican las condiciones de primer orden: hay n-2 grados de libertad de los residuos MCO (a diferencia de n grados de libertad en los errores)

Es insesgado

porque es el estimador insesgado de la varianza de u

usando el estimador de la varianza de mu se puede obtener un estimador insesgado de beta

## INTERCEPTO

## GAUSS-MARKOV

Los supuestos de homocedasticidad, no correlación lineal, exogeneidad, no multicolinealidad son las condiciones que aseguran que OLS es el mejor estimador de beta.

# R-Cuadrado

La forma de medir la bondad de ajuste:

La variación total de *y* (sus desviaciones de la media) es la suma de la variación de los valores predichos por el modelo y la variación de los datos con el modelo.

# Modelo sin intercepto

# Regresión múltiple

es el cambio en *y* con respecto a manteniendo todas las demás constantes.

La relación entre el error y las *x* es:

Es redundante en el modelo cuadrático.

Ejemplo con dos parámetros (k=2)

Las *i* son las observaciones; el subíndice es *ik*; observación - parámetro/variable. Entonces en el caso general, con la recta estimada

Hay en total  estimadores de los parámetros porque hay parámetros. El problema de MCO es

Y quedan las condiciones de primer orden:

La interpretación es igual que en el modelo simple. El modelo permite interpretar los cambios en la variable dependiente controlando las otras variables.

Los residuos también son

Se cumple la propiedad de que entonces . También . Y los promedios están en la línea.

## Teorema de Frisch-Waugh-Lovell (Partialling-out)

Corres la regresión simple de en .

Después haces una regresión de y en los residuos

Y ese al estimarlo es el del modelo de dos variables:

son los residuos de la regresión simple de en . O sea, usar como dependiente de . Después una regresión simple de *y* sobre *r* te da .

## Relación entre estimadores simples y múltiples

Si tenés dos modelos estimados uno simple: y otro múltiple: . Podes hacer la regresión de en o sea: , queda que

O sea, los parámetros con y sin la otra variable son iguales solo si no tiene efecto sobre y también no tiene efecto sobre .

Si se tienen k variables independientes, el estimador de en un modelo simple y otro con k variables va a ser igual solamente si: los coeficientes de las variables es 0 y/o no tiene correlación con cada una de las variables

## Sesgo de variable omitida

Si planteas dos modelos, uno simple y otro con otra variable, el simple va a estar sesgado:

Usando que son fijos porque depende de las variables x que tratábamos como no estocasticas

Solo si es 0 puede no estar sesgado el parámetro del modelo simple, ósea tienen que no estar correlacionadas las x-es. Si están correlacionadas va a tener el mismo signo que la correlacion entre ellas.

A blue and white box with black text

Description automatically generated

# Variables binarias

Aplicando media condicional 0:

Usando los valores que puede tomar *x*

La beta es la diferencia promedio en el valor de los y que cumplen ser de la categoría .

No cambian las propiedades de MCO, los residuos tienen media, no tienen correlación con los *x*, r-cuadrado tiene la misma interpretación

Propiedad:

Modelo binario con varias variables:

Diferencia esperada para hombres y mujeres para el mismo nivel de ingresos. El modelo binario te permite diferenciar los efectos para diferentes tratamientos.

Si se quiere estudiar el modelo

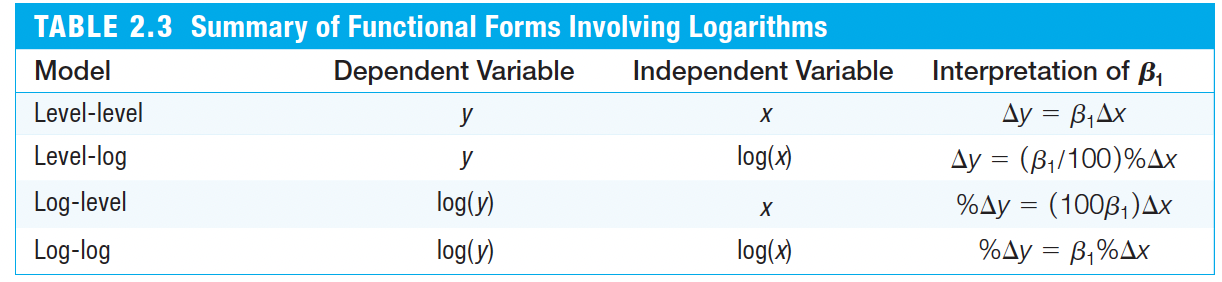
Si se quiere medir para hombres y mujeres con una dummy, este modelo está mal: hay multicolinealidad perfecta. Base de datos de ejemplo:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ingreso | Educación | Mujer | Hombre | Z |
| 12 | 10 | 1 | 0 | 1 |
| 30 | 8 | 1 | 0 | 1 |
| 25 | 5 | 0 | 1 | 1 |

Como Mujer y Hombre son combinaciones lineales, es imposible estimar el modelo. Si eliminas el intercepto no tenés el problema. La opción más usual es poner una categoría y dejar el resto como omitidas.

# Relaciones funcionales

## Logarítmicas



### Log-lin

En el modelo estimado log-lin:

La interpretación de cambia porque

X no es aleatoria

### Binaria log-lin

Ejemplo: si tenés una binaria en un log-lin

Para los hombres:

Para las mujeres:

Entonces:

Buscando el efecto marginal de ser hombre

El efecto marginal va a ser

## Cuadráticas

Se interpreta que para verlo como una regresión múltiple., las propiedades son las mismas para computar el MCO. Lo que cambia la interpretación de los parámetros entre este modelo. En el modelo cuadrático no mide el efecto de sobre *y* porque implicaría fijar y eso no tiene sentido: a medida que cambia uno cambia el otro. Para medir el cambio en *y* con respecto al cambio en *x* hay que hacer:

El cambio depende de los dos parámetros.

# Inferencia

## Intervalo Z

Significatividad:

C = probabilidad de rechazar si es verdadera

## Intervalo T

Para cuando no conoces . Usas el estadístico t

percentil de la

Bajo la hipótesis nula, x no explica y (beta es 0)

Bajo la alternativa, sí lo explica

Quisiera rechazar el estadístico tiene que dar un p-valor bajo (tiene que ser poco probable, dados los datos insertados en el estadístico, que la hipótesis nula sea verdadera)

# Forma matricial

En el modelo de k-variables, los valores poblacionales tienen la forma,

Extendiendo a cada observación se puede presentar como matriz

Modelo estimado:

En forma matricial,

Y también,

N x 1. N x K K x 1. N x 1

El MCO es la suma de los residuos cuadrados

1 x N · N x K · K x 1 = 1 x 1

**=**

**=** 1 x K · K x N · N x 1 = 1 x 1

# Ejercicios

A white paper with black text

Description automatically generated

a.V

a. Suponga el siguiente modelo: In(Y) = a + X + ui

y sean y los estimadores mínimos cuadráticos del intercepto y de la pendiente, respectivamente. Ahora, defina ã y como los estimadores correspondientes a la regresión In(kYi) en Xi en donde k . Entonces, los estimadores de ambos modelos pueden relacionarse de la siguiente manera: y

v

b. F: no cambia

c. V. pero entonces

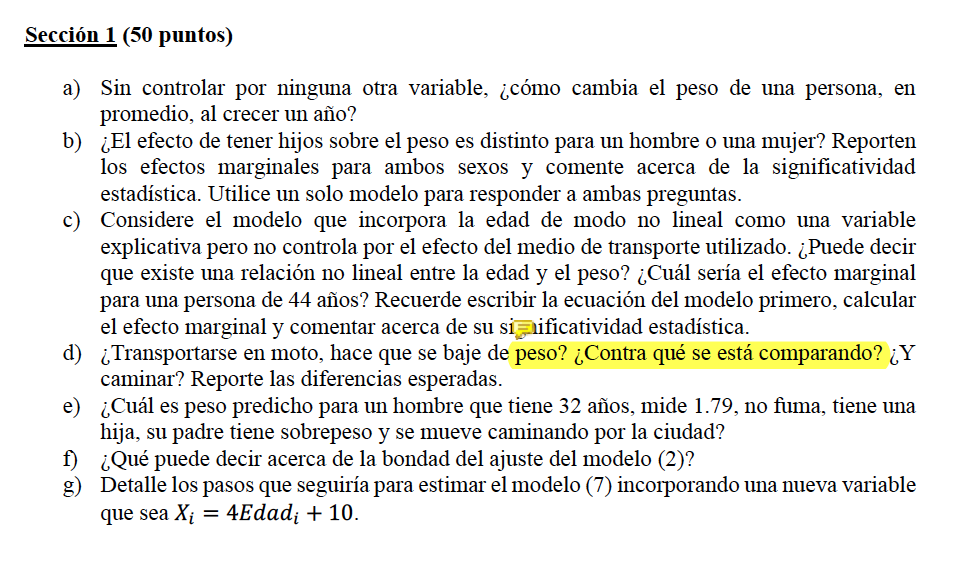
A white text with black text

Description automatically generated

B es insesgado si E( b\* ) = E( (Z’X)^-1 Z’Y ) = b

Exogeneidad

Insesgado



Sección 1 (50 puntos)

﻿﻿﻿Sin controlar por ninguna otra variable, ¿cómo cambia el peso de una persona, en promedio, al crecer un año?

﻿﻿﻿¿El efecto de tener hijos sobre el peso es distinto para un hombre o una mujer? Reporten los efectos marginales para ambos sexos y comente acerca de la significatividad estadística. Utilice un solo modelo para responder a ambas preguntas.

﻿﻿﻿Considere el modelo que incorpora la edad de modo no lineal como una variable explicativa pero no controla por el efecto del medio de transporte utilizado. ¿Puede decir que existe una relación no lineal entre la edad y el peso? ¿Cuál sería el efecto marginal para una persona de 44 años? Recuerde escribir la ecuación del modelo primero, calcular el efecto marginal y comentar acerca de su si nificatividad estadística.

﻿﻿﻿¿Transportarse en moto, hace que se baje de peso? ¿Contra qué se está comparando? ¿Y caminar? Reporte las diferencias esperadas.

﻿﻿﻿¿Cuál es peso predicho para un hombre que tiene 32 años, mide 1.79, no fuma, tiene una hija, su padre tiene sobrepeso y se mueve caminando por la ciudad?

﻿﻿﻿¿Qué puede decir acerca de la bondad del ajuste del modelo (2)?

﻿﻿﻿Detalle los pasos que seguiría para estimar el modelo (7) incorporando una nueva variable que sea Xi = 4Edadi + 10.

a. 0.830 k

b. Mujer\*Hijo : 0.0938\*\*\* sobre ln(peso).

Para los hombres no es siginificativo el efecto de un hijo sobre el peso

= 0

Usar que