PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
CIENCIAS SOCIALES
CICLO 2022-2

Fundamentos de Econometría
Práctica Dirigida 5 - Solucionario

Profesor:	Juan Palomino	juan.palominoh@pucp.pe
Jefes de Práctica:	Tania Paredes	tania.paredes@pucp.edu.pe

Fecha: 24 – 09 – 2022

1. Problema de multicolinealidad

Y	X_2	X_3
-10	1	1
-8	2	3
-6	3	5
-4	4	7
-2	5	9
0	6	11
2	7	13
4	8	15
6	9	17
8	10	19
10	11	21

a. ¿Es posible estimar el vector de parámetros del modelo $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$?

Si, podemos reexpresar el modelo en forma compacta

$$Y = \beta\,\mathrm{X} + \varepsilon$$

Por lo que, podemos obtener el resultado

$$\hat{\beta} = (X'X)^{-1}XY$$

b. ¿Existe multicolinealidad? ¿Si existiera, especifique e indique cuál sería la combinación lineal detrás del modelo?

Recordemos: El problema de multicolinealidad es cuando 2 o más variables independientes se correlacionan fuertemente. Por tanto, es difícil poder determinar cuál explica a cada variable. Un caso extremo es cuando una variable se construye en combinación de otra variable. En ese caso tenemos una colinealidad perfecta.

El principal problema de la colinealidad es que incrementa la incertidumbre en la estimación de los parámetros del modelo de regresión planteado (errores). En el caso de colinealidad perfecta, no se puede invertir la matriz (X'X) dado que el determinante de la matriz será 0 y la hace una matriz singular.

El supuesto del MRLC es que no haya presencia de multicolinealidad.

De la información mostrada, vemos que X_3 es una combinación de X_2.

$$X_3 = 2X_2 \quad - 1$$

Por lo que podemos reescribir el modelo de la siguiente manera:

$$Y_i = \beta_1 + \ \beta_2 X_{2i} + \beta_3(2X_{2i} - 1) + \varepsilon_i$$

¿Cómo podemos detectar la colinealidad?

- Revisar las matrices de correlaciones de las variables independientes (cuando tenemos un r>0.90)
- Analizamos la significancia individual y conjunta
- VIF>10 (Factor de inflación de la varianza)

c. ¿Cómo podría solucionar este problema de multicolinealidad?

- Replantear el modelo (eliminar una de las variables problema que genera la colinealidad)
- Incrementar el número de observaciones de la muestra

1.2 Se quiere estudiar los determinantes del ahorro. Para ello, se propone el siguiente modelo con una muestra para el periodo 1964-1968:

$$snfam_t = \beta_1 += \beta_2 rndfam_t + \beta_3 tdfam + \varepsilon_t$$

Donde:
snfam: Ahorro neto familiar.
rndfam: Renta disponible familiar.
tdfam: Impuestos directos pagados por las familias.

Importe el archivo Excel "Datos Ahorro" y responda las siguientes preguntas:

a. ¿Puede decir que hay presencia de multicolinealidad en el modelo? Sustente su respuesta.

Solucionario en Do file

b. Analizar la presencia de multicolinealidad mediante el Factor de Inflación de Varianza

Solucionario en Do file

2. Variables cualitativas

2.1. Supongamos un modelo de regresión lineal $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ con k variables en donde los individuos son personas. Se desea controlar por la variable Sexo y para ello se construyen dos variables binarias o "dummy" de esta manera:

$$S_{1i} = \begin{cases} 1 & si\ i\ es\ hombre \\ 0 & si\ i\ es\ mujer \end{cases} \qquad S_2 = \begin{cases} 1 & si\ i\ es\ mujer \\ 0 & si\ i\ es\ hombre \end{cases}$$

a. ¿Por qué no se pueden agregar las dos variables dummy y la constante?

No es posible realizar la estimación en el modelo, pues se cae en "la trampa de las variables dummy", la cual consiste en que no se pueden incluir todas las categorías simultáneamente, pues se genera un problema de multicolinealidad perfecta, porque la regresión tiene como intercepto $1 = S_{1i} + S_{2i} \ \forall i$
Esto generará en la matriz de datos una columna de unos que es colineal con la columna de unos de la constante del intercepto.
OJO: Para solucionar el problema de multicolinealidad perfecta se deberá elegir una variable como grupo base o de referencia.

b. Supongamos que solo se agrega la variable S_{1i} , quedando el modelo como

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_{kX_{ki}} + \gamma S_{1i} + \varepsilon_i$$

¿Cuál es la interpretación del coeficiente γ?

✓ γ será el parámetro asociado a la variable S_{1i}.

✓ Una interpretación sobre el parámetro γ es que si la variable dependiente fuera salario por hora, γ equivale a un incremento o reducción (si $\gamma>0$ o $\gamma<0$) del salario de los hombres frente a las mujeres independiente de las otras variables.

✓ Por lo que, en este caso γ podría evidenciar si hay discriminación en contra de los hombres para un $\gamma<0$.

✓ En forma de expectativas:

$$\gamma = E[Y_i \mid \underbrace{S_{1i}=1}_{\text{Hombre}}, X_{2i} \dots X_{ki}] - E[Y_i \mid \underbrace{S_{1i}=0}_{\text{Mujer}}, X_{2i} \dots X_{ki}]$$

c. Suponga que agrega las dos dummy al modelo y retira la constante, quedando:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_{kX_{ki}} + \alpha_1 S_{1i} + \alpha_2 S_{2i} + \varepsilon_i$$

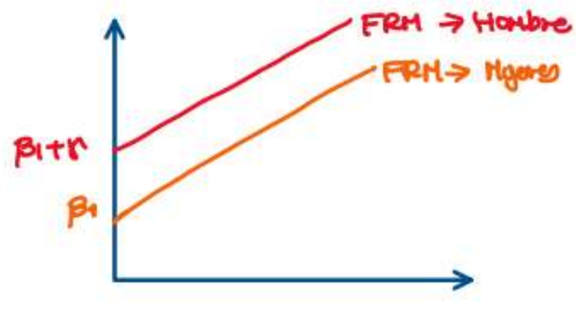
¿Qué relación hay entre $\beta_1, \gamma, \alpha_1$ y α_2?

Primer caso ; Si tenemos el modelo $Y_i = \beta_1 + \beta_2 X_{2i} + \dots \beta_k X_{ki} + \gamma S_{1i} + \varepsilon_i$

Tomamos Esperanzas condicionales:

$E[Y_i \mid S_{1i}=1] = (\beta_1+\gamma) + \underbrace{\beta_2 X_{2i} + \dots + \beta_k X_{ki}}_{\beta X}$ → Función de Regresión Muestral de un hombre

$E[Y_i \mid S_{1i}=0] = \beta_1 + \underbrace{\beta_2 X_{2i} + \dots + \beta_k X_{ki}}_{\beta X}$ → Función de Regresión Muestral de un Mujer



$\beta_1+\gamma$: intercepto de FRM Hombre

β_1 : intercepto de FRM Mujer.

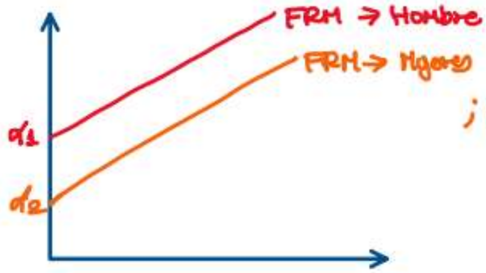
$\boxed{\text{Segundo caso}}$ $Y_i = \beta_2 X_{2i} + \dots \beta_k X_{ki} + d_1 S_{1i} + d_2 S_{2i} + \varepsilon_i$

Tomamos esperanzas condicionales:

$$E[Y_i \mid S_{1i}=1, S_{2i}=0] = d_1 + \overbrace{\beta_2 X_{2i} + \dots + \beta_k X_{ki}}^{\beta X}$$

→ FRM Hombre

$$E[Y_i \mid S_{1i}=0; S_{2i}=1] = d_2 + \underbrace{\beta_2 X_{2i} + \dots + \beta_k X_{ki}}_{\beta X}$$



¡Asumimos $d_2 < d_1$

- d_1 : intercepto de FRM Hombre
- d_2 : intercepto de FRM Mujer.

Por lo que:
✓ $\beta_1 = d_2$ → Intercepto de mujeres FRM
✓ $\beta_1 + \gamma = d_1$ → Intercepto de hombre FRM.
$\gamma = d_1 + \beta_1$

2.2. Se tiene el siguiente modelo:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 S_{1i} + \beta_4 S_{2i} + \varepsilon_i$$

Donde *Y* es el salario en soles, *X* representa la educación medida en años de estudio, S_1 es una variable binaria que toma el valor de 1 si *i* es hombre y 0 si es mujer, y S_2 es una variable binaria que toma el valor de 1 si *i* es mujer y 0 si es hombre.

a. ¿Existe algún problema para estimar dicho modelo? ¿Por qué?

Y_i = salario / ingreso en soles (s)
X_i = años de educación
S_{1i} = (variable binaria) que toma valor de 1 (Hombre) o 0 (Mujer)
S_{2i} = (variable binaria) que toma el valor de 1 (Mujer) o 0 (Hombre)

⇒ Si hay un problema dado que estamos incluyendo 2 variables binarias que representan lo mismo (género) : S_{1i}, S_{2i}

⇒ Gujarati & Porter ⇒ Cuando tenemos una variable cualitativa con "m" categorías ⇒ Solo debemos incluir en el modelo "m−1" categorías.

⇒ Si no se excluye a una categoría, incurrimos en la trampa de las dummys, dado que el det de X'X es igual a cero y no es invertible.

$$\hat{\beta} = (X'X)^{-1}X'Y \rightarrow \text{No invertible} \rightarrow \text{problema para hallar } \hat{\beta}.$$

b. Se decide eliminar del modelo la variable S_2. ¿Cuál sería la interpretación del coeficiente asociado a la variable S_1?

Reescribimos el modelo:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 S_{1i} + \varepsilon_i$$

$$S_{1i} \begin{cases} 1, & \text{si persona "i" es hombre} \\ 0, & \text{si persona "i" es mujer} \end{cases}$$

* $\frac{\partial Y_i}{\partial S_{1i}} = \beta_3 \rightarrow$ La variable dummy equivale a un incremento ($\beta_3>0$) o reducción ($\beta_3<0$) de los salarios de los hombres respecto a las mujeres.

c. A partir de este modelo con una sola variable binaria, se decide agregar una interacción (multiplicación) entre S_1 y X. ¿Cuál sería la interpretación del coeficiente asociado a esta *dummy* interactiva?

Reescribimos el modelo:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 S_{1i} + \beta_4 \underbrace{X_i S_{1i}}_{\text{Relación entre educación y género}} + \varepsilon_i$$

* $\frac{\partial Y_i}{\partial X_i S_{1i}} = \beta_4 \rightarrow$ ¿?

$\beta_4 > 0 \rightarrow$ significaría que los salarios de los hombres incrementan más con los años de educación a diferencia de los salarios de las mujeres.

3. Variables dummy iterativas y categóricas

Utilice la base de datos “trabajo.dta”. Se desea estimar el siguiente modelo:

$$lnSalario_i = \beta_1 + \beta_2 Educación + \beta_3 Edad + \beta_4 Edad^2 + \beta_5 Sexo + \varepsilon_i$$

a. Estime el modelo por MCO y comente sus resultados (signos y significancias)
   Solucionario en Do file

b. Verifique si hay multicolinealidad. Comente.
   Solucionario en Do file

c. Añada la dummy interactiva Sexo x Educación. ¿Qué interpretación tienen los signos calculados de la dummy sexo y de la dummy interactiva?
   Solucionario en Do file

d. Ahora queremos ver si los retornos en educación cambian según el nivel educativo. Estime un nuevo modelo que incorpore el análisis “Educ x Educación Secundaria” y “Educación x Educación Superior”. Interprete sus resultados.
   Solucionario en Do file

e. Finalmente, reestime el modelo analizado en (a) agregando la variable categórica de nivel socioeconómico. Interprete sus resultados.
   Solucionario en Do file