PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
FACULTAD DE CIENCIAS SOCIALES
ESPECIALIDAD DE ECONOMÍA
CICLO 2019-2

Curso: Econometría 1
Profesor: Luis García (lgarcia@pucp.edu.pe)
Jefa de práctica: Yasmeen Destre (y.destre@pucp.pe)

PRÁCTICA CALIFICADA 3

I. TEORÍA [4.5 PUNTOS]

1. ¿Cuáles son las dos características con las que debe contar toda categoría empleada para incorporar información de variables cualitativas en una regresión? **(1.5 puntos)**

Estas deben ser:

i. Exhaustivas: todas las observaciones deben estar clasificadas en al menos una categoría.
ii. Mutuamente excluyentes: ninguna observación puede ser clasificada en más de una categoría

2. Considere el siguiente conjunto hipotético de datos

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

¿Es posible estimar el vector de parámetros del modelo $Y_i = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u_i$? ¿Existe multicolinalidad? Si existiera, especifique e indique cual sería la combinación lineal detrás del modelo. **(1.5 puntos)**

No es posible estimar el vector de parámetros de este modelo por MCO, pues X_3 es una combinación lineal de X_2; en otras palabras, existe multicolinealidad. En particular se cumple que $X_3 = 2X_2 - 1$. En consecuencia, existe multicolinealidad perfecta entre estos regresores y la matriz $X = [i \quad X_2 \quad X_3]$ no es de rango lleno.

3. Se cuenta con la siguiente estimación:

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.6442721	.0538061	11.97	0.000	.5385695	.7499747
exper	.0700954	.0109776	6.39	0.000	.0485297	.0916611
_cons	-3.390539	.7665661	-4.42	0.000	-4.896466	-1.884613

¿Cuál será el salario (en dólares por hora) de un individuo con 15 años de educación y 15 años de experiencia laboral? **(1.5 puntos)**

$$\widehat{wage}_i = -3.390539 + 0.6442721 educ_i + 0.0700354 exper_i$$

Para una persona con 15 años de educación y 15 años de experiencia:

$$\widehat{wage}_f = -3.390539 + 0.6442721(15) + 0.0700954(15) = 7.3249735$$
$$\rightarrow 7.32\ dólares\ por\ hora\ aprox.$$

II. APLICACIONES [11.5 PUNTOS]

4. Después de haberse estimado el modelo de regresión lineal $y = XB + u$ por MCO y haber obtenido $\hat{Y} = X\hat{\beta}$, ¿qué valor tomará la variable endógena Y si las variables exógenas toman algún valor particular?

   a. Plantee el modelo en predicción. **(1 punto)**

$$y = X\beta + u$$
$$\hat{y} = X\hat{\beta}$$

Tenemos la predicción puntual:

$$x_f = [1 \quad X_{2f} \quad X_{3f} \dots \quad X_{kf}]$$

Entonces,

$$\hat{Y}_f = x_f\hat{\beta}$$

$$\hat{Y}_f = \hat{\beta}_1 + \hat{\beta}_2 X_{2f} + \hat{\beta}_3 X_{3f} + \dots + \hat{\beta}_k X_{kf}$$

Donde: $\hat{\beta}_1 + \hat{\beta}_2 X_{2f} + \hat{\beta}_3 X_{3f} + \dots + \hat{\beta}_k X_{kf}$ $es\ la\ predicción\ puntual$

   b. Encuentre el error en la predicción y especifique los argumentos incluidos. **(1.5 puntos)**

$$e_f = Y_f - \hat{Y}_f$$

$$e_f = x_f\beta + u_f - x_f\hat{\beta}$$
$$e_f = x_f(\beta - \hat{\beta}) + u_f$$

Donde: $x_f(\beta - \hat{\beta})$ es la imprecisión en la estimación de β y u_f perturbación imprevisible

5. Para el modelo $Y_i = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u_i$ se tienen los siguientes datos:

$$\boldsymbol{(X'X)^{-1}} = \begin{pmatrix} \mathbf{0.6477} & \mathbf{-0.041} & \mathbf{-0.0639} \\ \mathbf{-0.041} & \mathbf{0.0071} & \mathbf{-0.0011} \\ \mathbf{-0.0639} & \mathbf{-0.0011} & \mathbf{0.0152} \end{pmatrix}$$

$$\boldsymbol{X'Y} = \begin{pmatrix} \mathbf{91} \\ \mathbf{699} \\ \mathbf{448} \end{pmatrix}; \boldsymbol{n} = \mathbf{12}\ ; \boldsymbol{SCT} = \mathbf{104.9167}\ ; \boldsymbol{SCE} = \mathbf{78.2654}$$

a. Encuentre los coeficientes del modelo. **(1 punto)**

La estimación del modelo por MCO se obtiene a partir de:

$$\hat{\beta} = (X'X)^{-1}X'Y = \begin{pmatrix} 0.6477 & -0.041 & -0.0639 \\ -0.041 & 0.0071 & -0.0011 \\ -0.0639 & -0.0011 & 0.0152 \end{pmatrix}\begin{pmatrix} 91 \\ 699 \\ 448 \end{pmatrix} = \begin{pmatrix} 1.6545 \\ 0.7391 \\ 0.2258 \end{pmatrix}$$

Quedando el modelo de la siguiente forma: $\hat{Y}_i = 1.6545 + 0.7391X_2 + 0.2258X_3$

b. Contraste la hipótesis para $\beta_2 + \beta_3 = 1$, donde el valor crítico al 5% de significancia de la tabla F de Fisher es $F_{(1,9)} = 2.962$ **(2 puntos)**

$$R^2 = \frac{SCE}{SCT} = \frac{78.26547}{104.9167}$$

Para el contraste ($H_0 = \beta_2 + \beta_3 = 1$), tendremos en cuenta que rechaza la hipótesis nula si:

$$F = \frac{(R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)/q}{e'e/n-k} > F_{q,n-k}(1-\alpha)$$

De la restricción $\beta_2 + \beta_3 = 1$ se obtiene que $R = (0\ 1\ 1)$, $r = 1$ y $q = 1$, por lo que tenemos:

$$R\hat{\beta} - r = (0\ 1\ 1)\begin{pmatrix} 1.6545 \\ 0.7391 \\ 0.2258 \end{pmatrix} - 1 = -0.0351$$

$$R(X'X)^{-1}R' = (0\ 1\ 1)\begin{pmatrix} 0.6477 & -0.041 & -0.0639 \\ -0.041 & 0.0071 & -0.0011 \\ -0.0639 & -0.0011 & 0.0152 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = 0.0201$$

$$F = \frac{(-0.0351)(49.7512438)(-0.0351)/1}{(104.9167 - 78.26547)/12 - 3} = 0.02069872 \approx 0.0207$$

Si comparamos con $F_{(1,9)} = 2.962$, podemos concluir que no se rechaza hipótesis nula.

c. Encuentre el intervalo de predicción de $E[Y]$ sabiendo que $x_{2f} = 2.5$ y $x_{3f} = -0.3$. **(2 puntos)**

$$\widehat{Y_f} \pm t_{\left(1-\frac{\alpha}{2}\right)(n-k)}.S\sqrt{1 + x_f'(X'X)^{-1}x_f}$$

$$\widehat{Y_f} = 1.6545 + 0.7391(2.5) + 0.2258(-0.3) = 3.43451$$

$$x_0'(X'X)^{-1}x_0 = (1 \quad 2.5 \quad -0.3)\begin{pmatrix} 0.6477 & -0.041 & -0.0639 \\ -0.041 & 0.0071 & -0.0011 \\ -0.0639 & -0.0011 & 0.0152 \end{pmatrix}\begin{pmatrix} 1 \\ 2.5 \\ -0.3 \end{pmatrix}$$
$$= 0.528433$$

El intervalo de confianza al 95% es:

$$3.43451 \pm (2.2622)(1.721)(1.236) = (-1.3775; 8.24656)$$

6. El objetivo de una investigación es analizar las diferencias salariales de maestros por regiones en EEUU. Para ello, se cuenta con una encuesta de 50 maestros quienes trabajan en una de las siguientes regiones: Centro, Norte y Sur. Se desea conocer si el salario de los maestros difiere según la región y para ello realiza la siguiente estimación:

$$Y_i = \alpha + \beta_2 D_{2i} + \beta_3 D_{3i} + u_i$$

Donde Y_i es el salario promedio de los maestros en el estado i y las variables D_{2i} y D_{3i} son variables dummy definidas de la siguiente forma:

$$D_{2i} = \begin{cases} 1 & si\ i\ es\ un\ estado\ del\ Norte \\ 0 & en\ otros\ casos \end{cases}$$

$$D_{3i} = \begin{cases} 1 & si\ i\ es\ un\ estado\ del\ Norte \\ 0 & en\ otros\ casos \end{cases}$$

Luego de realizar una estimación por MCO con la información de estos 50 estados, se obtiene el siguiente resultado (los errores estándar se indican entre paréntesis):

$$\widehat{Y_i} = 48015.6_{(1857.2)} + 1524.1_{(2363.1)}D_{2i} - 1721_{(2467.2)}D_{3i}$$

a. ¿Cuál es el salario promedio de los maestros para cada una de las regiones? **(1 punto)**

Salario promedio de maestros del Norte:
$$E[Y_i|D_{2i} = 1, D_{3i} = 0] = 48014.6 + 1524.1(1) - 1721(0)$$

$$= 48014.6 + 1524.1 = 49538.7$$

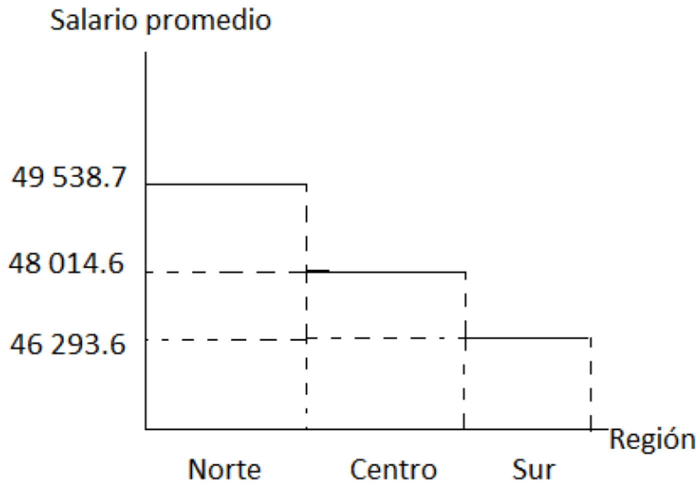
Salario promedio de maestros del Sur:

$$E[Y_i|D_{2i} = 0, D_{3i} = 1] = 48014.6 + 1524.1(0) - 1721(1)$$
$$= 48014.6 - 1721 = 46293.6$$

Salario promedio de maestros del Centro:

$$E[Y_i|D_{2i} = 0, D_{3i} = 0] = 48014.6 + 1524.1(0) - 1721(0)$$
$$= 48014.6$$

b. Ilustre gráficamente el modelo teniendo en cuenta los coeficientes estimados encontrados **(1.5 puntos)**



c. ¿Qué representan los parámetros estimados $\hat{\beta}_2$ y $\hat{\beta}_3$? **(1.5 puntos)**

Salario promedio de maestros del Norte:

$$E[Y_i|D_{2i} = 1, D_{3i} = 0] = \alpha + \beta_2(1) + \beta_3(0) = \alpha + \beta_2$$

Salario promedio de maestros del Sur:

$$E[Y_i|D_{2i} = 0, D_{3i} = 1] = \alpha + \beta_2(0) + \beta_3(1) = \alpha + \beta_3$$

Salario promedio de maestros del Centro:

$$E[Y_i|D_{2i} = 0, D_{3i} = 0] = \alpha + \beta_2(0) + \beta_3(0) = \alpha$$

Donde $\hat{\beta}_2$=1524.1 representa el salario adicional que ganan los maestros del Norte a los maestros del Centro y $\hat{\beta}_3 = 1721$ representa el salario adicional que ganan los del Sur en comparación a los maestros del Centro.

III. LABORATORIO CON STATA [4 puntos]

7. Se quiere estudiar los determinantes del ahorro. Para ello, se propone el siguiente modelo con una muestra para el período 1964 - 1998:

$$snfam_t = \beta_1 + \beta_2 rndfam_t + \beta_3 tdfam_t + u_t$$

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

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

```
. reg SNFAM RNDFAM TDFAM

      Source |       SS       df       MS              Number of obs =      35
-------------+------------------------------           F(  2,    32) =  284.94
       Model |  6.5778e+13     2  3.2889e+13           Prob > F      =  0.0000
    Residual |  3.6936e+12    32  1.1542e+11           R-squared     =  0.9468
-------------+------------------------------           Adj R-squared =  0.9435
       Total |  6.9472e+13    34  2.0433e+12           Root MSE      =  3.4e+05

------------------------------------------------------------------------------
       SNFAM |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      RNDFAM |   .0918641   .0267998     3.43   0.002     .0372747    .1464535
       TDFAM |  -.0935626   .1869882    -0.50   0.620     -.474445    .2873198
       _cons |   29327.91   124724.6     0.24   0.816    -224727.7    283383.5
------------------------------------------------------------------------------
```

a. ¿Puede decir que hay presencia de multicolinealidad en el modelo? Sustente su respuesta. **(2 puntos)**

Prueba F de significancia conjunta alta y significativa y R^2 alto; sin embargo, si analizamos la significancia de cada una de las variables explicativas, solo la variable rndfam es significativa y tdfam no ni la constante. Indicio de multicolinealidad: F-estadístico alto y t-estadísticos bajos.

b. Analizar la presencia de multicolinealidad mediante el Factor de Inflación de Varianza. **(1 punto)**

```
. vif

    Variable |       VIF       1/VIF
-------------+----------------------
      RNDFAM |     66.30    0.015083
       TDFAM |     66.30    0.015083
-------------+----------------------
    Mean VIF |     66.30
```

Los VIF individuales son mayores que 10 y el promedio es mucho mayor que 1. Hay presencia de Multicolinealidad.

c. En caso exista multicolinealidad, ¿qué métodos puede emplear para solucionar dicho escenario? **(1 punto)**

Regresión Ridge: si las variables en la matriz X están fuertemente correlacionadas, entonces la matriz $X'X$ tendrá valores fuera de la diagonal principal muy grandes o pesados. Luego, para contrarrestar este efecto, se le da un mayor peso a la diagonal multiplicándola por un escalar λ. El estimador será sesgado, pero tendrá menores varianzas que el de MCO.

Métodos de reducción de dimensiones: algunos métodos como el de los componentes principales, el análisis factorial y el análisis de correspondencia múltiple construyen variables indicadoras a partir de grupos de variables altamente colineales. Estos "índices" capturan un porcentaje de la variabilidad mostrada por las variables colineales y son utilizados en las regresiones en reemplazo de las variables que representan.

Si bien estas opciones son viables, aun no hay una solución que sea plenamente satisfactoria.