DEPARTAMENTO DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS
FACULTAD DE INGENIERÍA
PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

ECONOMETRÍA APLICADA

Profesora: Javiera Vásquez

Pauta Interrogación Nº2

Miércoles 18 de octubre de 2017

TIEMPO: 90 minutos

I. COMENTES (30 puntos)

Para cada una de las siguientes afirmaciones indique si es Falsa, Verdadera o Incierta. Siempre debe justificar su respuesta.

1. El test de normalidad de Jarque-Bera tiene como hipótesis nula que el coeficiente de asimetría es igual a cero. Comente. (5 puntos)

R: Falso, el test de Jarque-Bera tiene como hipótesis nula que la variable tiene una distribución normal, por lo tanto, se testea conjuntamente que el coeficiente de asimetría es cero y el coeficiente de curtósis es igual a 3.

2. En presencia de multicolinealidad, los intervalos de confianza deben ser obtenidos mediante bootstrap. Comente. (5 puntos)

R: Falso, cuando hay multicolinealidad se generan problema de eficiencia en el estimador MCO, la varianza se ve inflada con respecto a una situación donde las variables explicativas no son colineales entre ellas. Esto no tiene que ver con el supuesto de normalidad del error, y por lo tanto, mientras este supuesto no se rompa se pueden seguir obteniendo los intervalos de confianza bajo a partir de la distribución t-student.

3. La inclusión de variables explicativas en potencias siempre me permite mejorar la especificación del modelo, y mientras más potencias se incluyan mejor será el ajuste. Comente. (5 puntos).

R: Falso, no siempre las variables explicativas en potencias permiten mejorar la especificación del modelo, dependerá si existen o no, no linealidades omitidas. Si existen no linealidades omitidas, es posible que la inclusión de potencias de algunas variables explicativas permitan eliminar este sesgo, ahora no siempre es bueno incluir muchas potencias, generalmente con el cuadrado y el cubo es suficiente, más allá de estas las variables se comienzan a parecer mucho entre ellas generando un problema de multicolinealidad.

4. Ante presencia de heterocedasticidad, se deben calcular las varianzas robustas, esto me permite solucionar el problema de ineficiencia del estimador MCO en presencia de este problema. Comente. (5 puntos).

R: Falso, en presencia de heterocedasticidad el estimador MCO deja de ser el más eficiente, en este caso, el estimador eficiente es el estimador de Mínimos Cuadrados Generalizados, el que consiste en estimar por MCO un modelo transformado donde se elimina la heterocedasticidad mediante esta transformación. La varianza robusta corresponde al cálculo de las varianzas ineficientes en presencia de heterocedasticidad, que permite hacer inferencia basada en varianzas efectivas (ineficientes).

5. Si se rechaza la hipótesis nula del test Reset (test Ramsey), podemos estar tranquilos de que no existen variables no lineales omitidas. Comente. (5 puntos).

R: Falso, la hipótesis nula en el test Reset-Ramsey es que no existen no linealidades omitidas, por lo tanto, si se rechaza esta hipótesis nula, significa que existen no linealidades omitidas y por lo tanto debemos tratar de incorporar esto al modelo para corregir el sesgo por la omisión de estas variables no lineales.

6. Mientras mayor sea el nivel de significancia, mayor será el rango del intervalo de confianza. Comente. (5 puntos).

R: El intervalo de confianza de un parámetro poblacional corresponde al valor estimado del parámetro más y menos el valor de la distribución t con n-k grados de libertad y para el nivel de significancia establecido (α) por el error estándar. Mientras mayor sea el nivel de significancia, menor es el valor de la distribución t con n-k grados de libertad, y por lo tanto menor es el rango del intervalo de confianza.

II. Ejercicios aplicados

1. Una vez estimador el modelo de regresión lineal:

$$lprice = \beta_0 + \beta_1 rooms + \beta_2 stratio + \beta_3 lnox + u$$

Se ejecuta el siguiente comando:

```
. predict u, resid
```

a) ¿Qué es lo se está haciendo con este comando? (2 puntos)

R: con este comando se está generando una variable llamada u, que corresponde a los residuos (errores) del modelo, es decir:

$$\hat{u} = lprice - \hat{\beta}_0 - \hat{\beta}_1 rooms - \hat{\beta}_2 stratio - \hat{\beta}_3 lnox$$

La siguiente tabla muestra el test de normalidad sobre la variable `u`, creada a través del comando anterior:

```
. sktest u

                    Skewness/Kurtosis tests for Normality
                                                          ------- joint ------
    Variable |        Obs  Pr(Skewness)  Pr(Kurtosis) adj chi2(2)   Prob>chi2
-------------+---------------------------------------------------------------
           u |        506     0.0148        0.0000        55.25        0.0000
```

b) ¿Qué puede concluir con respecto a la hipótesis nula de que el coeficiente de asimetría es cero? (1 punto)

R: Al 5% de significancia esta hipótesis nula se rechaza, ya que el valor-p (`Pr(Skewness)=0.0148`) es menor a 0.05.

c) ¿Qué puede concluir con respecto a la hipótesis nula de que el coeficiente de curtósis es igual a 3? (1 punto)

R: Al 5% de significancia esta hipótesis nula se rechaza, ya que el valor-p (`Pr(Kurtosis)=0.0000`) es menor a 0.05.

d) ¿Qué puede concluir con respecto a la hipótesis nula de que la variable `u` tiene una distribución normal? (1 punto)

R: con respecto a la hipótesis nula de normalidad, es decir, conjuntamente que asimetría es cero y curtósis igual a 3, se tiene que el valor del estadístico es 55.25 el que se debe comparar con una chi cuadrado con 2 grados de libertad, pero también se tiene que el valor-p asociado a este estadístico es 0.0000, con lo cual se rechaza la hipótesis nula de normalidad del error.

2. Con respecto a la siguiente tabla:

```
. estat ovtest

Ramsey RESET test using powers of the fitted values of lrd
       Ho:  model has no omitted variables
                 F(3, 26) =      1.65
                 Prob > F =      0.2029
```

a) ¿Qué hace este comando? (1 punto)

R: este comando hace el teste Reset-Ramsey de no linealidades omitidas.

b) ¿Cuál es la hipótesis nula que se está testeando? (1 punto)

R: se está testeando la hipótesis nula de que no existen no linealidades omitidas, más específicamente se hace una regresión auxiliar del modelo original agregando tres variables explicativas: $\hat{y}^2, \hat{y}^3$ y $\hat{y}^4$ y se testea conjuntamente que los coeficientes de estas variables son cero.

c) ¿Por qué el estadístico F tiene 3 grados de libertad en el numerador y 26 grados de libertad en el denominador? (1 punto)

R: Como se explicaba en la parte b) el test corresponde a un test de 3 restricciones conjuntas, que los coeficientes de estas tres variables de potencias incluidas con cero, por lo que corresponde a un test F, con tres restricciones, por eso tiene 3 grados de libertad en el numerados, y los grados de libertad del denominador corresponde a n-k.

d) ¿Cuál es la conclusión con respecto a la hipótesis nula testeada? Justifique su respuesta. (2 puntos)

R: se obtiene un valor-p de 0.2029 para el valor del estadístico F de 1.65, con este valor p se tiene que no se puede rechazar la hipótesis nula de que los coeficientes asociados a las varianzas en potencias son cero, por lo tanto, se puede concluir que no existen no linealidades omitidas.

3. Suponga que usted está interesado la estima la elasticidad precio de la demanda por lentes de sol, para esto dispone de datos mensuales para los años 2014, 2015 y 2016 (36 observaciones) del logaritmo del precio y del logaritmo de la cantidad.

a) Plantee un modelo que le permita estimar el efecto marginal (porcentual) de un aumento de precio sobre la cantidad demandada de lentes de sol. (1 punto)

R: para estimar esta elasticidad (cambios porcentuales) se debe estimar el siguiente modelo:

$$lnq_t = \beta_0 + \beta_1 \cdot lnp_t + u_t$$

b) Defina una variable dummy temporada que tome valor 1 para los meses de noviembre, diciembre, enero, febrero y marzo, y cero para los restantes meses. Sea preciso en cómo se define esta variable. (1 punto)

R: la variable temporada toma valor 1 para t=2014-11;2014-12;2015-01;2015-02;2015-03;2015-11;2015-12;2016-01;2016-02;2016-03, y para las restantes fechas toma valor 0.

c) Al modelo estimado en a) agregue una o más variables que le permita testear la hipótesis que la demanda por lentes de sol en los meses de temporada es mayor que en los meses que no son de temporada. Sea preciso en el modelo a estimar (2 punto)

R: Para estimar el "efecto intercepto" es decir que en promedio la demanda por lentes de sol es mayor en meses de temporada que en el resto de los meses, basta con agregar la variable dummy temporada al modelo:

$$lnq_t = \beta_0 + \beta_1 \cdot lnp_t + \beta_2 \cdot temporada_t + u_t$$

De esta forma, el coeficiente que acompaña a esta variable dummy mide la diferencia porcentual promedio entre la demanda por lentes de sol en los meses de temporada y los meses fuera de temporada:

$$\beta_2 = E[lnq_t|temporada = 1] - E[lnq_t|temporada = 0]$$

d) A partir del modelo anterior, ¿cómo podría testear la hipótesis planteada?, sea especifico en qué tipo de test y como lo realizaría. (1 punto)

R: para testear que la demanda por lentes de sol es mayor en meses de temporada que en el resto de los meses, se debe hacer el siguiente test:

$$H_0: \beta_2 = 0$$
$$H_1: \beta_2 > 0$$

Lo que se hace mediante un test-t.

e) Al modelo estimado en c) agregue una o más variables que le permita testear la hipótesis de que la elasticidad precio de la demanda difiere entre los meses de temporada y los meses no de temporada. Sea preciso en el modelo a estimar (2 punto)

R: Para poder estimar una elasticidad precio de la demanda diferenciada entre meses de temporada y el resto de los meses, se debe interactuar la variable logaritmo del precio con la dummy temporada, así se debe estimar el siguiente modelo:

$$lnq_t = \beta_0 + \beta_1 \cdot lnp_t + \beta_2 \cdot temporada_t + \beta_3 \cdot lnp_t \cdot temporada_t + u_t$$

De esta forma, la elasticidad en los meses de temporada es:

$$\frac{\Delta lnq}{\Delta lnp} = \frac{\Delta\%q}{\Delta\%p} = \beta_1 + \beta_3$$

Y la elasticidad en los meses fuera de temporada:

$$\frac{\Delta lnq}{\Delta lnp} = \frac{\Delta\%q}{\Delta\%p} = \beta_1$$

f) A partir del modelo anterior, ¿cómo podría testear la hipótesis planteada?, sea especifico en qué tipo de test y como lo realizaría. (1 punto)

R: Para testear la hipótesis planteada se debe testear que el coeficiente β_3 es estadísticamente significativo:

$$H_0: \beta_3 = 0$$
$$H_1: \beta_3 \neq 0$$

Lo que se hace mediante un test-t.

4. Después de estimar el siguiente modelo para el logaritmo del precio de la vivienda:

$$lprice = \beta_0 + \beta_1 rooms + \beta_2 stratio + \beta_3 lnox + u$$

Se ejecuta el siguiente comando:

```
estat hettest, rhs
```

Obteniendo el siguiente resultado:

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: rooms stratio lnox

         chi2(3)      =   158.28
         Prob > chi2  =   0.0000
```

a) ¿Qué es lo que hace este comando? (1 punto)

R: test de heterocedasticidad

b) ¿Cuál es la hipótesis nula que se está testando? (1 punto)

R: se está testeando la hipótesis nula de que los errores del modelo tienen varianza constante, es decir, la hipótesis nula es de homocedasticidad.

c) ¿Por qué el estadístico chi-cuadrado tiene 3 grados de libertad? (1 punto)

R: porque en la regresión auxiliar del error al cuadrado, se inlcuyen las tres variables explicativas del modelo romos, stratio y lnox, y se hace un test de que estos tres coeficientes son cero.

d) ¿Qué puede concluir sobre la hipótesis nula testeada? Justifique su respuesta (2 punto)

R: El valor del estadístico chi-cuadrado (158.28) tiene un valor-p asociado de 0.0000 menor a 0.05, con lo cual se rechaza la hipótesis nula de que las varianzas son constantes y se puede concluir que el error del modelo es heterocedástico.

5. Con los datos de la encuesta Casen 2015, se define el salario por hora (`yph`) y luego se aplica logaritmo para obtener la variable `lyph` que será utilizada como variable dependiente.

Se tiene como variable explicativa el nivel educacional:

1= Sin educación
2= Básica Completa
3= Media Completa
4= Superior Completa

Además de la variable edad, edad al cuadrado y una variable dummy que toma valor 1 si la persona es hombre (`dhombre`).

A partir de la variable nivel educacional se generan 4 dummies, pero solo se incluyen tres dummies la modelo: `DE_2` que toma valor 1 si la persona tiene básica completa, `DE_3` que toma valor 1 si la persona tiene media completa y `DE_4` que toma valor 1 si la persona tiene educación superior completa.

De esta forma, el modelo estimado es el siguiente:

```
. reg lyp edad edad2 dhombre DE_2 DE_3 DE_4

      Source |       SS       df       MS              Number of obs =  106412
-------------+------------------------------           F(  6,106405) = 6831.28
       Model |   17641.219     6  2940.20316           Prob > F      =  0.0000
    Residual |  45797.0382106405  .430403066           R-squared     =  0.2781
-------------+------------------------------           Adj R-squared =  0.2780
       Total |  63438.2572106411  .596162588           Root MSE      =  .65605

------------------------------------------------------------------------------
        lyph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        edad |   .0291159   .0008634                      .0274237     .030808
       edad2 |  -.0002751   9.72e-06                     -.0002942   -.0002561
     dhombre |   .1863797   .0041228                       .178299    .1944604
        DE_2 |   .1945299   .0071621                      .1804923    .2085675
        DE_3 |   .5401112   .0066882                      .5270024      .55322
        DE_4 |   1.471183   .0084059                      1.454707    1.487658
       _cons |   6.252713   .0192148                      6.215052    6.290374
------------------------------------------------------------------------------
```

a) Interprete el coeficiente de la variable `dhombre`. (1 punto)

R: el coeficiente de la variable dhombre es igual a 0.1864, lo que nos indica que en promedio los hombres ganan 18.64% más que las mujeres.

b) ¿La variable `edad2` es estadísticamente significativa? Justifique. (1 punto)

R: Si, es significativa al 5% de significancia, ya que el cero no pertenece al intervalo de confianza al 95%.

c) ¿Cuál es la diferencia porcentual promedio entre el salario por hora entre las personas con educación superior completa y las personas con educación media completa? (1 punto).

R: De la regresión se tiene que el coeficiente que acompaña a DE_4 mide cuanto más ganan las personas con educación superior completa versus las personas sin educación, y el coeficiente que acompaña a DE_3 cuanto más ganan las personas con educación media completa con respecto a la personas sin educación, de esta forma la diferencia entre el coeficiente de DE_4 (1.4712) y el coeficiente de DE_3 (0.5401) indica cuanto más ganan en promedio las personas con educación superior completa con respecto a las personas con educación media completa. Esta diferencia es 93%.

d) ¿Cómo se interpreta el coeficiente asociado a la variable DE_4? (1 punto).

R: como se explicaba en la parte c) el coeficiente que acompaña a DE_4 mide la diferencia porcentual en salario por hora entre las personas con educación superior completa y las personas sin educación. El valor del coeficiente de 1.47 nos indica que en promedio las personas con educación superior ganan 147% más por hora que las personas sin educación.

La siguiente tabla muestra el factor de inflación de la varianza:

```
. estat vif
```

Variable	VIF	1/VIF
edad2	35.20	0.028411
edad	34.82	0.028721
DE_3	2.76	0.362695
DE_2	2.20	0.454660
DE_4	1.81	0.551244
dhombre	1.02	0.983356
Mean VIF	12.97	

e) ¿Existen problemas de multicolinealidad en este modelo? ¿Qué haría al respecto? (2 puntos)

R: Si, existe problemas de multicolinealidad ya que las variables edad y edad al cuadrado tienen un factor de inflación de la varianza (vif) mayor a 10. En este caso no haría nada ya que el aumento en la varianza no está afectando la significancia de las variables explicativas.