Solucionario Examen Parcial

Fundamentos de Econometría
1ECO1 - Horario: 0723

Profesor: Juan Palomino
24 de octubre, 2022

TEORÍA

Pregunta 1

(Conceptos): Responder brevemente los siguientes conceptos:

a. Multicolinealidad imperfecta y brinde un ejemplo. (1 punto)

- **Multicolinealidad imperfecta**: correlación entre las variables explicativas es alta, pero no perfecta. Ejemplo:

$$Consumo_i = \beta_1 + \beta_2 Ingreso_i + \beta_3 Riqueza_i + \epsilon_i$$

b. Error Tipo 1 y Error Tipo 2. Ejemplifique cada una. (1 punto)

- **Error Tipo 1 (Falso Positivo)** es el error que se comete cuando el investigador rechaza la hipótesis nula siento esta verdadera en la población. Ejemplo: Se considera que el paciente está enfermo, a pesar de que en realidad está sano; (H_0: el paciente está sano).
- **Error Tipo 2 (Falso Negativo)** se comete cuando el investigador no rechaza la hipótesis nula siendo esta falsa en la población. Ejemplo: Se considera que la bomba está desactivada pero en realidad está activada; (H_0: la bomba está desactivada).

c. Defina Perturbaciones No Esféricas. (1 punto)

- **Perturbaciones No Esféricas**: cuando la varianza de las perturbaciones no es constante y la covarianza entre diferentes perturbaciones es diferente de cero.

Pregunta 2

(Pruebas de Hipótesis): Asumir que tienes una muestra aleatoria del modelo:

$$\begin{aligned} y_i &= \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i \\ E(\epsilon_i|x_i) &= 0 \end{aligned}$$

donde y_i es el salario medido en soles por hora, y x_i es la edad. Describir como podrías testear la hipótesis que el salario esperado para un trabajador de 60 años es 20 soles una hora (2 puntos).

Aplicando esperanza a y_i se tiene:

$$\begin{aligned} E(y_i|x_i = 60) = 60\beta_1 + 60^2\beta_2 + E(\epsilon_i|x_i) &= 20 \\ = 60\beta_1 + 3600\beta_2 &= 20 \\ 3\beta_1 + 180\beta_2 &= 1 \end{aligned}$$

La hipótesis nula y alternativa serían:

$$\begin{aligned} H_{0:}\ 3\beta_1 + 180\beta_2 &= 1 \\ H_{1:}\ 3\beta_1 + 180\beta_2 &\neq 1 \end{aligned}$$

Sea $\hat{\beta}_1$ y $\hat{\beta}_2$ los coeficientes estimados por MCO, y sea $\widehat{V}$ la matriz de varianza-covarianza estimada. El estadístico F para esta hipótesis es:

$$F = (R\hat{\beta} - r)'[R\widehat{V}R']^{-1}(R\hat{\beta} - r)$$

donde

$$\begin{aligned} R &= \begin{bmatrix} 3 & 180 \end{bmatrix}_{1x2} \qquad r = 1 \\ R\hat{\beta} - r &= \begin{bmatrix} 3 & 180 \end{bmatrix}_{1x2} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}_{2x1} - 1 = 0 \end{aligned}$$

Entonces,

$$F = \frac{(3\hat{\beta}_1 + 180\hat{\beta}_2 - 1)^2}{R\widehat{V}R'}$$

Este tiene una distribución F bajo H_0. Una prueba de tamaño del 5 % es rechazar H_0 si F excede el valor crítico del 5 %. De lo contrario, H_0 no se rechaza.

EJERCICIOS

Pregunta 3

(Regresión Lineal): Considere la siguiente Tabla con observaciones anuales del logaritmo del Consumo (cp), de logaritmo del ingreso disponible (yd) y de la tasa de interés de referencia (i).

Año	Intercepto	yd	cp	i
2010	1	6.62	6.51	0.02
2011	1	6.66	6.52	0.03
2012	1	6.70	6.60	0.04
2013	1	6.76	6.64	0.052
2014	1	6.74	6.64	0.043
2015	1	6.77	6.66	0.030
2016	1	6.81	6.70	0.03
2017	1	6.85	6.76	0.05
2018	1	6.90	6.81	0.055
2019	1	6.83	6.83	0.06

a. Halle los estimadores $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, $\hat{\sigma}^2$ de la regresión $cp_t = \beta_1 + \beta_2 yd_t + \beta_3 i_t + \epsilon_t$. (1 punto)

Para empezar cargamos la siguiente librería

```
library(readxl)
```

Luego, importamos la base de datos:

```r
Datos <- read_excel("Datos.xlsx")
names(Datos)
```

```
## [1] "year" "cp"   "yd"   "i"
```

Usando la librería `lm` estimamos el modelo por MCO:

```r
ols <- lm(cp ~ yd + i, data=Datos)
summary(ols)
```

```
##
## Call:
## lm(formula = cp ~ yd + i, data = Datos)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.041323 -0.016908 -0.002920  0.005073  0.063832
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.3066     1.1856  -0.259 0.803403
## yd            1.0208     0.1804   5.658 0.000768 ***
## i             1.6733     1.1938   1.402 0.203763
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03266 on 7 degrees of freedom
## Multiple R-squared:  0.9315, Adjusted R-squared:  0.9119
## F-statistic:  47.6 on 2 and 7 DF,  p-value: 8.411e-05
```

Si se quiere extraer los coeficientes de esta estimación:

```r
b_hat <- summary(ols)$coefficients[ , 1]  # Extrayendo coeficientes
b_hat
```

```
## (Intercept)          yd           i
##  -0.3065575   1.0208387   1.6732843
```

Para extraer el valor del $\hat{\sigma}^2$ usamos lo siguiente:

```r
sigma(ols)^2
```

```
## [1] 0.001066706
```

b. Pruebe la hipótesis que $\beta_2 = 0$. Use como valor crítico $t_{\alpha/2} = 2.306$. ¿Rechaza o no rechaza la hipótesis nula? (1 punto)

Bajo la hipótesis nula $H_0 : \beta_2 = 0$, el estadístico t se define como:

$$t_2 = \frac{\hat{\beta}_2 - \beta_2}{SE(\hat{\beta}_2)}$$

y se distribuye como t_{n-k} es decir, distribución t con $(n - K)$ grados de libertad.

Entonces, para extraer los errores estándar del modelo estimado se programa lo siguiente:

```
se    <- summary(ols)$coefficients[ , 2]  # Extrayendo errores estándar
se
```

```
## (Intercept)          yd           i
##   1.1855723   0.1804301   1.1937806
```

De esa manera, se programa la siguiente fórmula $t_2 = \frac{\hat{\beta}_2}{SE(\hat{\beta}_2)}$

```
t <- b_hat[2]/se[2]
t
```

```
##       yd
## 5.657808
```

El valor del t estadístico es mayor al valor crítico $t_{\alpha/2} = 2.306$. Por lo tanto, se rechaza la hipótesis nula de que $\beta_2 = 0$, es decir, el coeficiente estimado es significativo.

c. Halle un intervalo de confianza al 95 % para $\hat{\beta}_3$. Use como valor crítico $t_{\alpha/2} = 2.306$.(1 punto)

Para hallar los intervalos de confianza con los límites inferiores y superiores al 95 % se usa la siguiente fórmula:

$$[\hat{\beta}_3 - t_{\alpha/2}SE(\hat{\beta}_3); \hat{\beta}_3 + t_{\alpha/2}SE(\hat{\beta}_3)]$$

De esa manera, se programa lo siguiente y se obtiene:

```
ic_low <- b_hat[3]-(2.306*se[3])
ic_high <- b_hat[3]+(2.306*se[3])
cat("El intervalo inferior es ", ic_low,", mientras que el intervalo superior es", ic_high)
```

```
## El intervalo inferior es  -1.079574 , mientras que el intervalo superior es 4.426142
```

Se observa que el intervalo de confianza traslapa el cero, por lo que el coeficiente $\hat{\beta}3$ no es significativo.

d. Halle el coeficiente R^2 y $\bar{R}^2$. (1 punto)

El valor del R^2 es:

```
summary(ols)$r.squared
```

```
## [1] 0.9315022
```

El valor del $\bar{R}^2$ es:

```
summary(ols)$adj.r.squared
```

```
## [1] 0.9119314
```

e. Halle el estadístico F para verificar la prueba de hipótesis $H_0 : \beta_2 = \beta_3 = 0$. Usa el valor crítico 4.46. ¿Rechaza o no rechaza la hipótesis nula? (1 punto)

Para evaluar la hipótesis lineal se instala el paquete `car`:

```
install.packages("car")
```

Y se carga la librería:

```r
library(car)
```

Una vez instalado, se evalúa la prueba de hipótesis $H_0 : \beta_2 = \beta_3 = 0$:

```r
linearHypothesis(ols, c("yd=0","i=0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## yd = 0
## i = 0
##
## Model 1: restricted model
## Model 2: cp ~ yd + i
##
##   Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1      9 0.109010
## 2      7 0.007467  2   0.10154 47.596 8.411e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

De aquí se obtiene que el valor del F estadístico es:

```r
linearHypothesis(ols, c("yd=0","i=0"))$F[2]
```

```
## [1] 47.59653
```

Se observa que el F estadístico es mayor al valor crítico 4.46. Por lo tanto, se rechaza la hipótesis nula de que $\beta_2 = \beta_3 = 0$, es decir, ambos coeficientes en conjunto son distintos de cero.

Pregunta 4

(Propiedades Asintóticas): Considerar el siguiente modelo $y_i = x_i\beta_0 + \epsilon_i$, con el supuesto $(\epsilon_i|x_i) = 0$ tal que $x_i, \beta, \epsilon_i \in \mathbb{R}$. Ahora considerar el siguiente estimador:

$$\tilde{\beta} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}$$

Asumir que $\{y_i, x_i\}$ es una muestra aleatoria i.i.d

a. ¿Es el estimador insesgado? (1 punto)

Introducimos y_i en el estimador $\tilde{\beta}$:

$$\begin{aligned}
\tilde{\beta} &= \frac{\sum_{i=1}^{n} x_i\beta_0 + \epsilon_i}{\sum_{i=1}^{n} x_i} \\
\tilde{\beta} &= \frac{\beta_0 \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i} + \frac{\sum_{i=1}^{n} \epsilon_i}{\sum_{i=1}^{n} x_i} \\
\tilde{\beta} &= \beta_0 + \frac{\sum_{i=1}^{n} \epsilon_i}{\sum_{i=1}^{n} x_i}
\end{aligned}$$

Luego:

$$\begin{aligned}
\mathbb{E}(\tilde{\beta}|x_i) &= \beta_0 + \frac{1}{\sum_{i=1}^n x_i}\mathbb{E}(\sum_{i=1}^n \epsilon_i|x_i) \\
&= \beta_0 + \frac{1}{\sum_{i=1}^n x_i}\sum_{i=1}^n \mathbb{E}(\epsilon_i|x_i) \quad \text{Propiedad de la esperanza} \\
&= \beta_0 + \frac{1}{\sum_{i=1}^n x_i}0 \quad \text{Por LIE y exogeneidad} \\
&= \beta_0 \quad \text{El estimador es insesgado}
\end{aligned}$$

b. Encontrar $Var(\tilde{\beta}|x_i)$. (1 punto)

$$\begin{aligned}
Var(\tilde{\beta}|x_i) &= Var(\beta_0 + \frac{\sum_{i=1}^n \epsilon_i}{\sum_{i=1}^n x_i}|x_i) \\
&= Var(\frac{\sum_{i=1}^n \epsilon_i}{\sum_{i=1}^n x_i}|x_i) \quad \text{ya que } \beta_0 \text{ es constante} \\
&= \frac{1}{(\sum_{i=1}^n x_i)^2}\sum_{i=1}^n Var(\epsilon_i|x_i) \quad \text{por propiedad de la varianza} \\
&= \frac{\sum_{i=1}^n \sigma_i^2}{(\sum_{i=1}^n x_i)^2}
\end{aligned}$$

c. ¿Es el estimador consistente? Realizar una lista de todos los supuestos que has realizado. (1.5 puntos)

Tenemos el sampling error:

$$\begin{aligned}
\tilde{\beta} &= \beta_0 + \frac{\sum_{i=1}^n \epsilon_i}{\sum_{i=1}^n x_i} \\
\tilde{\beta}_n &= \beta_0 + \frac{\frac{1}{n}\sum_{i=1}^n \epsilon_i}{\frac{1}{n}\sum_{i=1}^n x_i} \quad \text{dejando en promedio}
\end{aligned}$$

Entonces:

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \epsilon_i \xrightarrow{p} E(\epsilon_i) \quad \text{by Weak Law of Large Numbers} \\
\text{if} \quad &\rightarrow \quad (1)\ \{y_i, x_i\} \text{ es i.i.d random sample, entonces } \epsilon_i \text{ es i.i.d.} \\
&\rightarrow \quad (2)\ E(\epsilon_i) < \infty
\end{aligned}$$

También:

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n x_i \xrightarrow{p} E(x_i) \quad \text{by Weak Law of Large Numbers} \\
\text{if} \quad &\rightarrow \quad (1)\ \{y_i, x_i\} \text{ es i.i.d random sample} \\
&\rightarrow \quad (2)\ E(x_i) < \infty
\end{aligned}$$

Tener en cuenta que:

$$[\frac{1}{n}\sum_{i=1}^n x_i]^{-1} \xrightarrow{p} [E(x_i)]^{-1} \quad \text{by Continuous Mapping Theorem}$$

Entonces por el supuesto: $E(\epsilon_i|x_i) = 0$, tengo:

$$E(\epsilon_i) = E(E(\epsilon_i|x_i)) = 0 \quad \text{by LIE}$$

Luego, por Teorema de los grandes números:

$$\Rightarrow \quad \tilde{\beta}_n \xrightarrow{p} \beta_0 + \frac{E(\epsilon_i)}{E(x_i)} \text{ if } E(x_i) \neq 0$$
$$\tilde{\beta}_n \xrightarrow{p} \beta_0 + \frac{E(\overbrace{E(\epsilon_i)}^{\approx 0})}{E(x_i)} \text{ LIE}$$
$$\tilde{\beta}_n \xrightarrow{p} \beta_0 \text{ Es consistente}$$

Los supuestos son:

- $n \to \infty$.
- $\{y_i, x_i\}$ es i.i.d random sample, entonces ϵ_i es i.i.d.
- $E(\epsilon_i) < \infty$.
- $E(x_i) < \infty$ y $E(x_i) \neq 0$
- $E(\epsilon_i|x_i) = 0$

d. Hallar $\sqrt{n}(\tilde{\beta} - \beta_0)$ a medida que $n \to \infty$. Realizar una lista de todos los supuestos que has realizado. (1.5 puntos)

$$\tilde{\beta} - \beta_0 = \frac{\sum_{i=1}^n \epsilon_i}{\sum_{i=1}^n x_i}$$

Multiplicando por $\sqrt{n}$: %

$$\sqrt{n}(\tilde{\beta} - \beta_0) = \sqrt{n}\frac{\sum_{i=1}^n \epsilon_i}{\sum_{i=1}^n x_i}$$
$$\sqrt{n}(\tilde{\beta} - \beta_0) = \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^n \epsilon_i}{\frac{1}{n}\sum_{i=1}^n x_i} \quad \text{dejando en promedios}$$

Como:

$$\frac{1}{n}\sum_{i=1}^n x_i \xrightarrow{p} E(x_i) \quad \text{by Weak Law of Large Numbers}$$
$$\sqrt{n}(\frac{1}{n}\sum_{i=1}^n \epsilon_i) \xrightarrow{d} N(E(\epsilon_i), Var(\epsilon_i)) \quad \text{by CLT}$$

Sabemos que por lo mostrado anteriormente, entonces: %

$$\sqrt{n}(\frac{1}{n}\sum_{i=1}^n \epsilon_i) \xrightarrow{d} N(0, Var(\epsilon_i))$$

También sabemos que $var(x) =$, entonces: %

$$
\begin{aligned}
Var(\epsilon_i) &= E(\epsilon_i^2) - [\underbrace{E(\epsilon_i)}_{=0}]^2 \\
Var(\epsilon_i) &= E(\epsilon_i^2) \text{ if } E(\epsilon_i^2) < \infty
\end{aligned}
$$

Entonces tenemos: %

$$\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n}\epsilon_i) \xrightarrow{d} N(0, E(\epsilon_i^2))$$

Por lo tanto,

$$
\begin{aligned}
\Rightarrow \quad & \sqrt{n}(\tilde{\beta} - \beta_0) \xrightarrow{d} \frac{N(0, E(\epsilon_i^2))}{\xrightarrow{p} E(x_i)} \text{ as } n \to \infty \\
& \sqrt{n}(\tilde{\beta} - \beta_0) \xrightarrow{d} N(0, \underbrace{[E(x_i)]^{-1}E(\epsilon_i^2)}_{V}) \\
& \sqrt{n}(\tilde{\beta} - \beta_0) \xrightarrow{d} N(0, V)
\end{aligned}
$$

Supuesto adicional:

- $E(\epsilon_i^2) < \infty\}$
- $[E(x_i)]^{-1} \neq 0$

Pregunta 5

(Interpretación): Se tiene la siguiente ecuación salarial:

$$log(wage) = \beta_1 + \beta_2 sch_i + x_i^{'}\gamma + \epsilon_i$$

donde $log(wage)$ es el salario por hora, sch son los años de escolaridad del individuo. Asimismo, usando la Encuesta Nacional de Hogares (ENAHO) se estima por MCO y se obtiene los siguientes resultados:

Ecuación Salarial usando ENAHO

	Modelo 1	Modelo 2	Modelo 3	Modelo 4
Constante	6.493^{***}	6.341^{***}	5.742^{***}	5.804^{***}
	(0.008)	(0.009)	(0.014)	(0.014)
$Escolaridad$	0.066^{***}	0.070^{***}	0.098^{***}	0.094^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
$Mujer$		-0.184^{*}	-0.178^{**}	-0.167^{**}
		(0.006)	(0.006)	(0.006)
$Experiencia$			0.013^{***}	0.010^{***}
			(0.000)	(0.000)
$Permanencia$				0.007^{***}
				(0.000)
N	84251	84251	84251	84251
R^2	0.089	0.099	0.132	0.136
$R^2 ajustado$	0.084	0.093	0.122	0.130

Notes: *, **, *** denota significancia estadística al 10 %, 5 % and 1 %, respectivamente.

Asimismo, considerar que $E(\hat{\beta}_2|sch) = \beta_2 + \beta_3\hat{\gamma}$ donde β_3 es la relación entre salario y mujer, y $\hat{\gamma}$ es el coeficiente de la regresión de mujer sobre escolaridad. Asumir que para esta muestra las mujeres ganan en promedio menos que los hombres $\beta_3 < 0$. Y de una regresión entre escolaridad y mujer se obtiene que $\hat{\gamma}$ es positivo.

Responder lo siguiente:

a. ¿Cómo se interpreta el coeficiente de escolaridad para el modelo 1? ¿Es significativo? (1 punto)

Un año de escolaridad más tiene un efecto de 6.6 % en los ingresos de los individuos, significativo al 1 %.

b. En el modelo 2 el coeficiente de escolaridad es sesgado hacia arriba o hacia abajo con respecto al modelo 1? (2 puntos)

Dado que $\beta_3 < 0$ y $corr(mujer, sch)$>0, el sesgo es negativo, es decir, el coeficiente de escolaridad en el modelo 1 es sesgado hacia abajo. Inversamente, el coeficiente de escolaridad en el modelo 2 es sesgado hacia arriba con respecto al modelo 1.

c. Se podría incorporar una variable de $Hombre$ al modelo? ¿Qué problemas estaría ocasionando? ¿Por qué? (1 punto)

No se puede incorporar una variable de $Hombre$ al modelo ya que está presente el coeficiente de la categoría $Mujer$. En caso se incorpore estaría generando problemas de multicolinealidad.

d. ¿Qué es el $R^2 ajustado$ y por qué tiene un valor más pequeño que el R^2? (1 punto)

$$R^2 ajustado \quad = \quad 1 - \frac{(n-1)}{(n-K)}\frac{SCR}{SCT}$$

El R^2 ajustado castiga la inclusión de muchas variables, en el sentido que si K (# variables explicativas) aumenta, la SCR disminuye y paralelamente $\frac{n-1}{n-K}$ aumenta. Para que R^2 ajustado aumente, $SCR > \frac{n-1}{n-K}$, entonces la variable incluída si es relevante.