ECONOMETRÍA I

Apuntes Examen [1]

Primavera 2019

Anáiticos y Comentes

1. En el contexto de un Modelo de Regresión Lineal (MRL), la recta de regresión pasa por los valores promedios de la variable dependiente. Comente.

> **Solution:** Este se cumple siempre que el modelo de regresión incluya un término constante, es la constante justamente quien cumple el rol de asegurar que el promedio muestral del valor estimado de Y se iguale con el promedio muestral de los valores observados de Y . Si el modelo incluye término constante por condición de primer orden se cumple que $\sum u_i = 0$, y que el valor estimado de la constante (β) es $\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2\overline{X}$, de lo que deduce inmediatamente que la recta de regresión pasa por los valores promedio: $\overline{Y} = \hat{\beta}_1 + \hat{\beta}_2\overline{X}$. Por lo tanto, esta afirmación es verdadera sólo si estamos hablando de un modelo que incluye término constante.

2. Si un estimador es insesgado, significa que no existe error de estimación. Comente.

> **Solution:** Falso, al tratar de aproximar un parámetro poblacional a partir de una muestra siempre existe un error de estimación, es inevitable debido a las fluctuaciones muestrales. Lo que sucede si es el estimador es insesgado es que este error de estimación en promedio es igual a cero y por lo tanto, en valor esperado el estimador es igual a su valor poblacional.

3. Explique en qué consiste el sesgo en la estimación de un parámetro.

> **Solution:** El sesgo en un parámetro corresponde a la diferencia entre el valor esperado del estimador del parámetro y el valor verdadero del parámetro. Por lo tanto, esta medida se conoce como $E(\hat{\beta}) - \beta = 0$. Si este es insesgado, entonces $E(\hat{\beta}) = \beta$.

4. Explique la diferencia entre causalidad económica y correlación estadística.

> **Solution:** La principal diferencia que existe entre causalidad económica y correlación estadística es que la primera es una relación de causa-efecto en un sentido económico. En cambio, la correlación estadística es simplemente una observación estadística que indica un grado de relación lineal. La correlación estadística no necesariamente implica que una variable correlacionada con otra se comporte de una forma cuando la variable con la que tiene la correlación cambie, no hay necesariamente una causalidad.

[1] Material preparado por Juan Felipe Ly para el curso Econometría I, segundo semestre 2019. Cualquier error en este documento es responsabilidad del autor. Comentarios a juan.ly@mail.udp.cl

5. Mientras mayor es el tamaño de la muestra que disponemos, más se aproxima un estimador a su valor poblacional. Comente.

Solution: Si bien cuando el tamaño de muestra aumenta, esta cada vez se parece más a la población, un estimador para que en el límite sea igual al valor poblacional tiene que cumplir con la propiedad de consistencia. Recordar que un estimador es simplemente una fórmula o método que nos dice como aproximar un parámetro poblacional a través de una muestra, existen estimadores consistentes y otros que no lo son, a pesar que la muestra sea a infinito un estimador puede ser distinto a su valor verdadero, en este caso es inconsistente.

6. La estimación por el método de variables instrumentales es muy útil para los casos en que el residuo está correlacionado con las variables independientes.

Solution: Verdadero. El método sirve para esos casos, puesto que se viola el supuesto de que las variables independientes y el residuo no estén correlacionadas. Un ejemplo que vimos en clases es cuando hay error de medida en las variables dependientes. Esto bajo el supuesto de que tenemos buenos instrumentos, que estén correlacionados con la variables independientes y no correlacionados con el residuo.

7. El estimador de 2 etapas utilizando Variables Instrumentales (VI) es siempre consistente, aún si no hay problemas de endogeneidad. Es por esto que es conveniente utilizar dicho estimador para asegurarse de no cometer problemas de sesgo e inconsistencia. Comente.

Solution: Efectivamente dicho estimador es consistente aún sin problemas de endogeneidad. Sin embargo, el estimador con VI es ineficiente, por lo que sólo debiese ser ocupado si es que efectivamente hay un problema de endogeneidad que resolver.

Para testear la existencia de endogeneidad se puede realizar un test de Hausman. De rechazar la nula de exogeneidad, se podría utilizar VI.

8. La estimación por variables instrumentales siempre es mejor en términos de consistencia y sesgo que la estimación por mínimos cuadrados ordinarios. Comente.

Solution: Falso. Si existe problema de endogeneidad y se tiene un instrumento exógeno y relevante el estimador de VI es consistente mientras que MCO es sesgado e inconsistente. Pero si se tiene el problema de endogeneidad y se utiliza un instrumento débil (que no sea exógeno o que sea irrelevante) la estimación VI puede ser más sesgada e inconsistente que el estimador MCO.

9. En una prueba de hipótesis cualquiera, la zona de rechazo nunca cambia al cambiar la hipótesis nula. Comente.

Solution: Falso, cuando hacemos un test de hipótesis de la forma $H_0 : R\hat{\beta} = r$ para ver si rechazamos o no la hipótesis nula debemos comparar el valor calculado del estadístico con el valor de tabla de una distribución F con q grados de libertad en el numerador y $n - k$ grados de libertad en el denominador. Si bien el valor de tabla de la distribución F no cambia con r (los valores que testeamos bajo la hipótesis nula) si cambia con el número de hipótesis que estemos testeando(q).

10. Es imposible que en el siguiente modelo $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$, los valores de $\hat{\beta}_1$ y $\hat{\beta}_2$ sean individualmente no significativos, pero que el test F de significancia global del modelo nos diga que este es significativo. Comente.

Solution: Es posible que los test de significancia individual de $\hat{\beta}_1$ y $\hat{\beta}_2$ nos digan que estos no son significativos individualmente. Sin embargo, esto no quiere decir que en conjunto sean estadísticamente no significativos. Lo anterior podría darse por la covarianza existente entre los coeficientes estimados: puede darse el escenario donde la covarianza existente entre ambos coeficientes sea tal que en conjunto logren explicar la variabilidad de la variable dependiente.

11. La significancia global del modelo depende de la significancia individual de las variables independientes incluidas.

Solution: Falso. Es posible que, aun cuando los coeficientes estimados no sean individualmente significativos, sí lo sean en conjunto. Lo anterior podría darse en el caso que la covarianza de los regresores sea capaz de "explicar" la variabilidad de la variable dependiente. Como existe un alto grado de correlación entre las variables independientes del modelo (es decir, un alto grado de multicolinealidad), entonces, al menos una de estas variables tiene una influencia significativa sobre la variable dependiente, pero no se puede establecer cual. Por otro lado, al existir un alto grado de multicolinedlidad, es muy probable que los coeficientes, a nivel individual, no sean estadísticamente significativos. Recuerde que:

$$var(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1}$$

Si existe un alto grado de colinealidad entre las variables independientes del modelo, el determinante de $(X'X)$ tiende a cero, por lo que se obtienen varianzas "gigantes".

12. E R^2 mide la proporción de la varianza de la variable dependiente que es explicada por la varianza de las variables explicativas, de esta forma, siempre es un número que esta entre 0 y 1. Comente.

Solution: Depende, cuando el modelo de regresión incluye un término constante, se cumple que: $ST = SR + SE$, lo que se conoce como la descomposición de varianza, lo que garantiza que el R^2 sea siempre positivo. Sin embargo, si el modelo no incluye constante, no se cumple la ecuación anterior de descomposición de varianza y el R^2 puede tomar valores negativos, pero sigue siendo siempre menor a 1.

13. La omisión de una variable relevante siempre sesga la estimación MCO de β. Comente.

Solution: Falso, la omisión de una variable relevante siempre produce sesgo en los parámetros a menos que la correlación entre la variable omitida y las explicativas incluidas sea cero, el signo del sesgo depende de dos cosas: la correlación entre la variable omitida y las variables explicativas incluidas y el valor del parámetro que tendría asociado la variable omitida(valor poblacional). Para un modelo sencillo con una variable explicativa (x_1) y una variable omitida (x_2) se mostró en clases que el sesgo es: $\frac{cov(x_1,x_2)}{V(x_1)}\beta_2$.

14. Un R^2 alto siempre nos indica que nuestro modelo esta bien especificado. Comente.

Solution: Falso, modelos con presencia de multicolinealidad pueden arrojar elevados R^2. Sin embargo, la estimación de los parámetros del modelo será ineficiente, arrojando conclusiones erróneas sobre la significancia de los parámetros involucrados.

15. Miguel le comenta a su amigo Pedro que incluir una gran cantidad de variables a un modelo de regresión lineal, no tiene mayor incidencia sobre el R^2 del modelo. Pedro le refuta que inevitablemente el R^2 tenderá a aumentar. Explique el razonamiento de Pedro para realizar este postulado. Demuestre que en el caso extremo en que $n = k$ el R^2 del modelo es igual a uno.

Solution: Pedro está en lo correcto puesto que al agregar variables al modelo, independiente de si estas son significativas o no, el R^2 se mantendrá igual o aumentará de valor pero nunca disminuirá. Esto dado que los cuadrados de los residuos nunca aumentan cuando se añaden variables.

En el caso que $n = k$ se tiene que la matriz X se vuelve cuadrada, y al ser de rango completo será invertible. Luego podemos redefinir nuestro estimador MCO de la siguiente forma:

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

Luego, los residuos del modelo quedan de la siguiente manera:

$$\hat{u} = Y - X\hat{\beta}$$
$$\hat{u} = Y - XX^{-1}Y$$
$$\hat{u} = 0$$

Luego:

$$SRC = \sum \hat{u}^2 = 0$$
$$R^2 = 1$$

16. Dado que el estimador de mínimos cuadrados ordinarios (MCO) es insesgado, si $\hat{\beta}_1 = 2,07$, entonces, el verdadero parámetro β_1 es igual a 2,07.

Solution: El insesgamiento indica que la esperanza del estimador es igual al parámetro poblacional. Para una muestra en particular, el estimador no tiene por qué igualar al verdadero valor y, de hecho, podría estar más bien alejado del verdadero parámetro poblacional, dependiendo de qué tan grande sea la varianza.

17. Siempre es posible obtener los estimadores de MCO en el modelo múltiple utilizando la expresión $\hat{\beta} = (X'X)^{-1}X'Y$.

Solution: Si la matriz $(X'X)$ no es invertible, no será posible computar la fórmula señalada. Ahora bien, para que la matriz $(X'X)$ sea invertible, es necesario que el rango de la matriz X sea completo, es decir, que sea igual al número de regresores incluidos en el modelo múltiple ($rango(X) = k$).

18. Considerando el siguiente modelo tentativo para modelar un fenómeno económico $Y = \beta + u$. Si el error está bien comportado, el estimador de MCO de β, es igual a la media muestral de la variable dependiente ($\overline{Y}$).

Solution: El problema que se debe resolver es el siguiente:

$$min \sum \hat{u_i}^2 = min \sum (Y_i - \hat{\beta})^2$$

Derivando con respecto a $\hat{\beta}$ e igualando a cero:

$$\frac{\partial \sum (Y_i - \hat{\beta})^2}{\partial \hat{\beta}} = -2 \sum (Y_i - \hat{\beta})^2 = 0$$

$$\hat{\beta}_{MCO} = \frac{1}{n} \sum Y_i = \overline{Y}$$

19. Los residuos del modelo de regresión lineal se asumen normales pero no independientes e idénticamente dristribuidos.

Solution: Dentro de los supuestos dados a los residuos del modelo a estimar, está el de normalidad, lo que junto a los supuestos de media condicional igual cero y la varianza constante (homocedasticidad), implica que los residuos son idénticamente distribuidos. Por otra parte, se asume que los residuos son independientes entre ellos, es decir, que la covarianza entre e_i y e_j para todo i distinto de j es cero. Así, se tiene que con los supuestos estándares sobre los residuos, éstos son normales y al mismo tiempo independientes e idénticamente distribuidos.

20. La significancia estadística de una variable x_i está por completo determinada por la magnitud de $\hat{\beta}_i$, mientras que la significancia económica de esa variable está relacionada con la magnitud (y signo) de $t_{\hat{\beta}_i}$.

Solution: Es al revés, la significancia estadística de una variable x_i está por completo determinada por la magnitud de $t_{\hat{\beta}_i}$, mientras que la significancia económica de esa variable está relacionada con la magnitud (y signo) de $\hat{\beta}_i$.

21. Un síntoma de la existencia de multicolinealidad es la presencia de un R^2 ajustado alto.

Solution: El síntoma corresponde no simplemente a la observación de un R^2 alto, sino que además acompañado por estadísticos t que nos hacen concluir que los parámetros no son significativos (t bajos). Esto porque en presencia de multicolinealidades, las varianzas de los estimadores son altas, pues $(X'X)$ es cercano a cero. Lo que resulta en estadísticos t pequeños.

22. Usted observa en los datos que existe un elevado grado de asociación (negativo) entre la variable antiguedad y escolaridad del trabajador. ¿Cómo podría usted determinar si tiene en verdad un problema de multicolinealidad en los datos? En caso de haberla, ¿cuáles serían las consecuencias para la estimación de MCO?

Solution: Se debe mencionar y describir el test de Belsey y uno o más síntomas que indiquen la presencia de multicolinealidad. Algunos de ellos podrían ser: R^2 alto con coeficientes estadísticamente no significativos, pequeño cambio en los datos produce grandes variaciones en estimaciones por MCO, y coeficientes con signos y magnitudes distintas a las esperadas. En presencia de multicolinealidad, MCO sigue siendo MELI; sin embargo, las varianzas serán muy grandes.

23. El estimador de White me permite solucionar el problema de ineficiencia en el estimador MCO cuando los errores son Heterocedásticos. Comente.

Solution: Falso, el estimador de White sólo me permite estimar en forma consistente la matriz de varianzas y covarianzas de los parámetros en presencia de heterocedasticidad estimados por MCO. El estimador MCO siempre será ineficiente;el estimador eficiente en este contexto es el de Mínimos Cuadrados Generalizados. Por lo tanto, White me permite estimar consistentemente la matriz de varianzas y covarianzas del estimador MCO (que es ineficiente) para realizar la inferencia en forma correcta, utilizando esta matriz.

24. Es equivalente realizar un Test F de significancia global del modelo y realizar varios test de significancia individual de todos los parámetros.

Solution: Falso, aunque realicemos todos los test de significancia individual para las pendientes del modelo este resultado no va a ser equivalente al test de significancia global, ya que este último considera la correlación entre las variables explicativas. Puede existir un caso extremo que las variables tengan una correlación muy alta y esta varianza conjunta explique el comportamiento de la variable independiente, aunque cada una por separado no es capaz de explicar el comportamiento de Y .

Matemáticos

25. Explique por qué usando el test VIF (Variance Inflation Factor) se puede testear multicolinealidad. Discuta si excluir la variable colineal es la solución.

Solution: El test VIF efectivamente nos sirve para testear multicolinealidad. Este test tiene en su construcción el coeficiente de determinación R_j^2, que es el R^2 no centrado de una regresión entre la variable x_j y el resto de regresores X_j. Si este R_j^2 es alto, entonces tendremos una alta correlación entre variables explicativas y el coeficiente VIF

$$VIF = \frac{1}{1 - R_j^2}$$

también será alto. Un número que nos debiese llamar la atención es un VIF mayor a 10 (que implicitamente nos indica un R_j^2 de 0,9).

26. Considere el siguiente modelo:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

donde u_i se distribuye independiente e idénticamente con media condicional nula y varianza σ^2. Un econometrista ha estimado este modelo con una muestra de datos X e Y, y tiene lo siguiente:

$$X'X = \begin{pmatrix} 5 & 20 \\ 20 & 120 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 40 \\ 230 \end{pmatrix}$$

Por otro lado usted sabe que:

$$\sum_{i=1}^{n} (Y_i - \overline{Y})^2 = 124$$

En base a esta información, conteste lo siguiente:

(a) ¿Cuál es el tamaño de la muestra con que se llevó a cabo esta estimación?

Solution: Para el modelo simple, la matriz $X'X$ tiene la siguiente forma:

$$X'X = \begin{pmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{pmatrix}$$

Por otra parte, la matriz $X'Y$ tiene la siguiente forma:

$$X'Y = \begin{pmatrix} \sum Y_i \\ \sum X_i Y_i \end{pmatrix}$$

Por lo tanto el tamaño de la muestra es $n = 5$.

(b) Encuentre $\overline{X}$ e $\overline{Y}$.

Solution: Los promedios de X e Y vienen dados por:

$$\overline{X} = \frac{\sum X_i}{n} = \frac{20}{5} = 4$$

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{40}{5} = 8$$

(c) Encuentre $\hat{\beta}_1$ y $\hat{\beta}_2$

Solution: El estimador de MCO viene dado por:

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (X'X)^{-1}X'Y$$

Luego:

$$(X'X)^{-1} = \begin{pmatrix} 0,6 & -0,1 \\ -0,1 & 0,025 \end{pmatrix}$$

Por lo tanto:

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 0,6 & -0,1 \\ -0,1 & 0,025 \end{pmatrix} \begin{pmatrix} 40 \\ 230 \end{pmatrix} = \begin{pmatrix} 1 \\ 1,75 \end{pmatrix}$$

(d) Sabiendo que $\sum \hat{u_i}^2 = 1,5$, calcule la matriz de covarianzas para $\hat{\beta}_1$ y $\hat{\beta}_2$. Utilice el estimador insesgado de σ^2.

Solution: La matriz de varianzas y covarianzas viene dada por:

$$var(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

$$var(\hat{\beta}) = \frac{\sum \hat{u_i}^2}{n-k-1}(X'X)^{-1}$$

$$var(\hat{\beta}) = \frac{1,5}{3} \begin{pmatrix} 0,6 & -0,1 \\ -0,1 & 0,025 \end{pmatrix}$$

$$var(\hat{\beta}) = \begin{pmatrix} 0,3 & -0,05 \\ -0,05 & 0,0125 \end{pmatrix}$$

(e) Calcule el R^2 y el R^2 ajustado ($\overline{R}^2$).

Solution: El R^2 viene dado por:

$$R^2 = 1 - \frac{\sum \hat{u_i}^2}{\sum (Y_i - \hat{Y})^2}$$

$$R^2 = 1 - \frac{1,5}{124}$$

$$R^2 = 0,9879$$

27. Considere el modelo:

$$y_i = \beta_0 + \beta_1 x_i + u_i \qquad (1)$$

donde $u_i \sim N(0, \sigma_u^2)$. Se sabe que la variable x_i está medida con error, de modo que sólo se observa:

$$x_i{}^* = x_i + v_i$$

donde $v_i \sim N(0, \sigma_v^2)$ y es independiente de u_i, de x_i y de y_i.

(a) Reescriba el modelo en la ecuación (1) en función sólo de las variables observables y de los términos de error.

Solution:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$
$$y_i = \beta_0 + \beta_1 (x_i{}^* - v_i) + u_i$$
$$y_i = \beta_0 + \beta_1 x_i{}^* - \beta_1 v_i + u_i$$
$$y_i = \beta_0 + \beta_1 x_i{}^* - \beta_1 v_i + u_i$$
$$y_i = \beta_0 + \beta_1 x_i{}^* + (u_i - \beta_1 v_i)$$
$$y_i = \beta_0 + \beta_1 x_i{}^* + \varepsilon_i$$

(b) Muestre que en el modelo en (a) el error está correlacionado con la variable observable $x_i{}^*$. Explique las implicancias de este problema.

Solution:

$$Cov(\varepsilon_i, x_i{}^*) = Cov(u_i - \beta_1 v_i\ ,\ x_i + v_i)$$
$$Cov(\varepsilon_i, x_i{}^*) = Cov(u_i, x_i) + Cov(u_i, v_i) - \beta_1\ Cov(v_i, x_i) - \beta_1\ Cov(v_i, v_i)$$
$$Cov(\varepsilon_i, x_i{}^*) = 0 + 0 - \beta_1 \cdot 0 - \beta_1 \cdot {\sigma_v}^2$$
$$Cov(\varepsilon_i, x_i{}^*) = -\beta_1 \cdot {\sigma_v}^2$$

Se observa que el error ϵ_i está correlacionado con la variable observable $x_i{}^*$. Esto implica que la estimación por MCO generará estimadores sesgados de los parámetros β.

(c) Explique cómo se podría solucionar el problema en (b).

Solution: Una forma de correción es la utilización de variables instrumentales. La idea es utilizar un instrumento en vez de la variable medida con error. Para ello, se requiere que el instrumento no esté correlacionado con el término de error y que el instrumento sí esté correlacionado con la variable explicativa para la cual actúa como instrumento (la variable medida con error).

(d) En el modelo en la ecuación (1) considere ahora que sólo la variable dependiente y_i está medida con error $w_i \sim N(0, \sigma_w^2)$, el que es independiente de x_i y de u_i. ¿Cuáles son las implicancias de este problema de medición? Explique y justifique algebraicamente.

Solution: Si sólo la variable dependiente está medida con error, se tiene que sólo se observa:

$$y_i{}^* = y_i + w_i$$

De este modo, el modelo observable es

$$y_i{}^* = \beta_0 + \beta_1\ x_i + u_i$$

$$y_i + w_i = \beta_0 + \beta_1\ x_i + u_i$$
$$y_i = \beta_0 + \beta_1\ x_i + u_i - w_i$$
$$y_i = \beta_0 + \beta_1\ x_i + (u_i - w_i)$$
$$y_i = \beta_0 + \beta_1\ x_i + \epsilon_i$$

Se tendrá entonces que la covarianza entre el término de error y las variables explicativas es igual a cero:

$$Cov(\epsilon_i, x_i) = Cov(u_i - w_i,\ x_i)$$
$$Cov(\epsilon_i, x_i) = Cov(u_i, x_i) + Cov(w_i, x_i)$$
$$Cov(\epsilon_i, x_i) = 0$$

Por lo tanto, no habrá problema de sesgo en la estimación de los parámetros.

28. Considerando el siguiente modelo de regresión lineal con error de medición en la variable explicativa

$$Y_i = X_i \cdot \beta + u_i$$

donde u es un shock estocástico bien comportado y $X_i{}^* = X_i + e_i$

Demuestre que el estimador por variables instrumentales se puede escribir como:

$$\hat{\beta}_{VI} = (\hat{X}^{*'}\hat{X}^*)^{-1}\hat{X}^{*'}Y$$
$$\hat{\beta}_{VI} = [\hat{X}^{*'}Z(Z'Z)^{-1}Z'\hat{X}^*]^{-1}\hat{X}^{*'}Z(Z'Z)^{-1}Z'Y$$

Solution: Modelo verdadero: $Y_i = X_i\beta + u_i$, se observa: $X_i{}^* = X_i + e_i$. Como $X_i{}^*$ está medida con error, se puede utilizar el instrumento Z. Para obtener el estimador de variables instrumentales hago un estimador en dos etapas:

Primera etapa: regresión entre $X_i{}^*$ y Z_i, para obtener $\hat{X_i}^*$:

$$X_i{}^* = Z_i\rho + v_i$$
$$\hat{\rho} = (Z'Z)^{-1}Z'X^*$$
$$\hat{X}^* = Z(Z'Z)^{-1}Z'X^*$$

Segunda etapa: regresión entre Y y $\hat{X}^*$ del modelo original.

$$Y_i = \hat{X_i}^*\beta + u_i$$
$$\hat{\beta}_{VI} = (\hat{X}^{*'}\hat{X}^*)^{-1}\hat{X}^{*'}Y$$

Dado que $\hat{X}^* = Z(Z'Z)^{-1}Z'X^*$, podemos escribir $\hat{\beta_{VI}}$ en función de Z.

$$\hat{\beta}_{VI} = (\hat{X}^{*'}Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}\hat{X}^*)^{-1}\hat{X}^{*'}Z(Z'Z)^{-1}Z'Y$$
$$\hat{\beta}_{VI} = [\hat{X}^{*'}Z(Z'Z)^{-1}Z'\hat{X}^*]^{-1}\hat{X}^{*'}Z(Z'Z)^{-1}Z'Y$$

29. Suponga un modelo dado por:

$$y_t = \alpha + \beta x_t + u$$

Determine si los siguientes estimadores de la pendiente son sesgados:

(a)

$$\hat{\beta}_1 = \frac{\sum(y_t - \overline{y})}{\sum(x_t - \overline{x})}$$

Solution:

$$\hat{\beta}_1 = \frac{\sum(y_t - \overline{y})}{\sum(x_t - \overline{x})}$$
$$= E(\frac{\sum(y_t - \overline{y})}{\sum(x_t - \overline{x})})$$
$$= \frac{1}{\sum(x_t - \overline{x})} E(\sum(y_t - \overline{y}))$$

Considerando que:

$$(y_t - \overline{y}) = \beta(x_t - \overline{x}) + u_t$$

Reemplazando:

$$= \frac{1}{\sum(x_t - \overline{x})} E(\sum(\beta(x_t - \overline{x}) + u_t))$$
$$= \frac{1}{\sum(x_t - \overline{x})} \sum(\beta(x_t - \overline{x}) + E(u_t))$$
$$= \frac{1}{\sum(x_t - \overline{x})} \sum(\beta(x_t - \overline{x}))$$
$$= \beta \frac{\sum(x_t - \overline{x})}{\sum(x_t - \overline{x})}$$
$$E(\hat{\beta}_1) = \beta$$

Por lo tanto, es insesgado.

(b)

$$\hat{\beta}_2 = \frac{\sum y_t}{\sum x_t}$$

Solution:

$$\hat{\beta}_2 = \frac{\sum y_t}{\sum x_t}$$
$$E(\hat{\beta}_2) = E(\frac{\sum y_t}{\sum x_t})$$
$$= \frac{\sum E(y_t)}{\sum x_t}$$

Reemplazando el modelo:

$$= \frac{\sum E(\alpha + \beta x_1 + u)}{\sum x_t}$$
$$= \frac{\sum(\alpha + \beta x_t)}{\sum x_t}$$
$$= \frac{n\alpha + \beta \sum x_t}{\sum x_t}$$
$$E(\hat{\beta}_2) = \frac{n\alpha}{\sum x_t} + \beta$$

Por lo tanto, es sesgado.

30. Considere el modelo
$$y = \beta_0 + \beta_1\ x_1 + \beta_2\ x_2 + u$$
donde u se distribuye independiente e idénticamente con media condicional nula y varianza σ^2. Un econometrista ha estimado este modelo con una muestra de datos X e Y, y tiene lo siguiente:
$$(X^tX)^{-1} = \begin{pmatrix} 32,1 & 3,5 & -7,5 \\ 3,6 & 2,0 & -1,0 \\ -7,0 & -1,5 & 3 \end{pmatrix}, \quad X^tY = \begin{pmatrix} 4 \\ 6 \\ 19 \end{pmatrix}$$
Además, se tiene que $s^2 = 0,6$. En base a esta información, conteste lo siguiente:

(a) Encuentre los valores de los parámetros estimados usando Mínimos Cuadrados Ordinarios.

Solution:
$$\hat{\beta} = (X^tX)^{-1}\ X^tY$$
$$\hat{\beta} = \begin{pmatrix} 32,1 & 3,5 & -7,5 \\ 3,6 & 2,0 & -1,0 \\ -7,0 & -1,5 & 3 \end{pmatrix} \cdot \begin{pmatrix} 4 \\ 6 \\ 19 \end{pmatrix}$$
$$\hat{\beta} = \begin{pmatrix} 6,9 \\ 7,4 \\ 20 \end{pmatrix}$$

(b) Calcule la matriz de varianzas y covarianzas, $Var(\hat{\beta})$.

Solution:
$$Var(\hat{\beta}) = s^2(X^tX)^{-1}$$
$$Var(\hat{\beta}) = 0,6 \cdot \begin{pmatrix} 32,1 & 3,5 & -7,5 \\ 3,6 & 2,0 & -1,0 \\ -7,0 & -1,5 & 3 \end{pmatrix}$$
$$Var(\hat{\beta}) = \begin{pmatrix} 19,26 & 2,1 & -4,5 \\ 2,16 & 1,2 & -0,6 \\ -4,2 & -0,9 & 1,8 \end{pmatrix}$$

(c) Construya un intervalo de confianza al 95 % para la predicción cuando $x_1 = 15$ y $x_2 = 5$.

Solution: Al 95

$$IC : \hat{y_f} \pm 1,96\sqrt{c'Var(\hat{\beta})c}$$

Donde,

$$\hat{y_f} = c' \cdot \hat{\beta}$$

$$\hat{y_f} = \begin{pmatrix} 1 & 15 & 5 \end{pmatrix} \cdot \begin{pmatrix} 6,9 \\ 7,4 \\ 20 \end{pmatrix}$$

$$\hat{y_f} = 217,9$$

Ahora, se tiene que:

$$c'Var(\hat{\beta})c = \begin{pmatrix} 1 & 15 & 5 \end{pmatrix} \cdot \begin{pmatrix} 19,26 & 2,1 & -4,5 \\ 2,16 & 1,2 & -0,6 \\ -4,2 & -0,9 & 1,8 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 15 \\ 5 \end{pmatrix}$$

$$c'Var(\hat{\beta})c = \begin{pmatrix} 30,66 & 15,6 & -4,5 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 15 \\ 5 \end{pmatrix}$$

$$c'Var(\hat{\beta})c = 242,16$$

Por lo tanto el Intervalo de confianza nos queda,

$$IC : \hat{y_f} \pm 1,96\sqrt{c'Var(\hat{\beta})c}$$

$$IC : 217,9 \pm 1,96 \cdot \sqrt{242,16}$$

$$IC : 217,9 - 30,500 \leq y_f \leq 217,9 + 30,500$$

$$IC : 187,399 \leq y_f \leq 248,400$$

$$IC : (187,399 \ , \ 248,400)$$

31. Suponga que Ud. Quiere estimar el siguiente modelo:

$$ln(y_i) = \beta_0 + \beta_1 \ Mujer_i + \beta_2 \ Educ_i + \beta_3 \ Mujer_i \times Educ_i + u_i$$

donde $ln(y_i)$ es el logaritmo natural del salario por hora trabajada, $Educ$ es años de escolaridad y $Mujer$ es una variable dummy que toma el valor 0 si el trabajador es hombre y 1 en el caso de ser mujer.

(a) ¿Cuál es la interpretación del coeficiente β_0. ¿Cuál es el retorno a la educación para las mujeres?

Solution: β_0 es el logaritmo natural del salario por hora trabajada esperado para los hombres con 0 años de escolaridad en esta muestra. El retorno a la educación para las mujeres es $\beta_2 + \beta_3$ (β_2 es el retorno para los hombres).

(b) Suponga que quiere testear la hipótesis que existe un premio por título universitario que para los hombres es el doble que para las mujeres, mientras que el retorno a los años de educación es el mismo para hombres que para mujeres. Explique cómo testearía la hipótesis anterior, señalando claramente cómo cambiaría el modelo, cuál es la hipótesis nula, y qué test utilizaría.

Solution: Para incorporar el premio al título universitario, modificamos el modelo anterior de la siguiente manera:

$$ln(y_i) = \beta_0 + \beta_1\ Mujer_i + \beta_2\ Educ_i + \beta_3\ Mujer_i \times Educ_i + \beta_4\ Titulo_i + \beta_5\ Titulo_i \times Mujer_i + u_i$$

donde la variable Título es una variable dicotómica que toma el valor 1 si la persona obtuvo un título universitario. Las hipótesis a testear son las siguientes:

$$H_0:\ \beta_3 = 0$$

$$H_0:\ \beta_4 = 2(\beta_4 + \beta_5)$$

Para testear las hipótesis anteriores se realiza un test F.

32. Suponga que usted tiene que estimar el siguiente modelo:

$$y_i = \beta_1\ x_{1,i} + \beta_2\ x_{2,i} + u_i \qquad donde\ \ i = 1102$$

De la estimación, usted obtiene los siguientes resultados:

$$\hat{\beta} = \begin{pmatrix} 0,5 \\ 0,4 \end{pmatrix} \quad (X'X)^{-1} = \begin{pmatrix} 10 & 8 \\ 8 & 10 \end{pmatrix} \quad Var(\hat{\beta}) = \begin{pmatrix} 0,7 & -0,56 \\ -0,56 & 0,7 \end{pmatrix} \quad Var(y) = 25,2$$

(a) Realice los test de significancia para cada uno de los parámetros.

Solution: Test de significancia de β_1:

$$t^c = \frac{0,5}{\sqrt{0,7}} = 0,59 \sim t^t = t_{100,0,95} \approx 1,98$$

No se puede rechazar la hipótesis nula de que β_1 sea 0, el parámetro no es estadísticamente significativo.
Test de significancia de β_2:

$$t^c = \frac{0,4}{\sqrt{0,7}} = 0,47 \sim t^t = t_{100,0,95} \approx 1,98$$

No se puede rechazar la hipótesis nula de que β_2 sea 0, el parámetro no es estadísticamente significativo.

(b) Determine el número de condición de la matriz X, para ver la presencia de colinealidad entre las variables explicativas (coeficiente de Belsley).

Solution: El número de condición de la matriz X se obtiene con el siguiente coeficiente de Belsey:

$$\gamma = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}$$

donde λ_{max} y λ_{min} son los valores propios de la matriz $B = S(X'X)^{-1}S$. Además, S es una matriz de $n \times n$ diagonal, en que su diagonal está formada por $\frac{1}{\sqrt{x_i{}^2}}$. Para este ejercicio S está dado por:

$$S = \begin{pmatrix} \frac{1}{\sqrt{10}} & 0 \\ 0 & \frac{1}{\sqrt{10}} \end{pmatrix}$$

Por lo que la matrix $B = S(X'X)^{-1}S$ queda:

$$B = \begin{pmatrix} 1 & \frac{4}{5} \\ \frac{4}{5} & 1 \end{pmatrix}$$

Ahora hay que calcular los valores propios de B, lo que se hace a continuación:

$$|B - \lambda\ I| = 0$$

$$|\begin{pmatrix} 1 & \frac{4}{5} \\ \frac{4}{5} & 1 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}| = 0$$

$$|\begin{pmatrix} 1-\lambda & \frac{4}{5} \\ \frac{4}{5} & 1-\lambda \end{pmatrix}| = 0$$

Calculando el valor anterior (determinante) queda:

$$(1-\lambda)^2 - (4/5)^2 = 0$$

Los valores propios son $\lambda_1 = 1/5$ y $\lambda_2 = 9/5$. Reemplazando estos valores en el índice de Belsley:

$$\gamma = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} = \sqrt{\frac{9/5}{1/5}} = 3$$

Como el índice es menor que 25, no estamos en presencia de multicolinealidad.