



Pontificia Universidad Católica de Chile
Escuela de Ingeniería
Departamento de Ingeniería Industrial y de Sistemas

ICS 2562: Econometría Sección 02

Profesora: Sonia Vera O. Ayudante: Benjamín Hartmann

Ayudante Coordinador: Matías Vergara

Tercera Prueba – 7 Junio de 2018

Tiempo: 2.5 horas.

COMPROMISO DEL CÓDIGO DE HONOR

Me comprometo a no entregar ni recibir ayuda indebida en esta evaluación. Esto incluye discutir la evaluación con compañeros que aún no la han rendido. También declaro que si me percato de que existe fraude de cualquier tipo en esta evaluación, tengo el deber de comunicárselo a la Profesora Sonia Vera, quien seguirá los procedimientos establecidos en la reglamentación de la Escuela de Ingeniería y de la Pontificia Universidad Católica de Chile para perseguir y sancionar cualquier acto de deshonestidad académica.

Nombre completo: _____

Firma: _____

Fecha: 7 Junio de 2018

**NO OLVIDE COLOCAR NOMBRE A TODAS LAS HOJAS DE RESPUESTA Y
ENTREGAR SEPARADAMENTE CADA PREGUNTA DE ACUERDO A LO
INSTRUIDO POR LOS AYUDANTES**

Está prohibido tener el celular u otro artículo electrónico con usted, salvo su calculadora. Todas sus pertenencias deben estar guardadas y dejadas en la parte delantera de la sala. Del mismo modo, está prohibido salir de la sala durante el desarrollo de la prueba para ir al baño o hablar por teléfono. Si usted debe salir de la sala, deberá entregar su prueba, la que se dará por terminada.

1.- La Sociedad Científica para el Desarrollo de América Latina ha hecho un estudio sobre las consecuencias económicas para los países de la región debido al cambio climático. Para ello trabajó con datos de corte transversal considerando 5200 comunas pertenecientes a los distintos países de la región. Las variables incluidas en el modelo, para el periodo relevante de análisis, son:

Y : Pérdidas monetarias en la comuna, en millones de dólares, acumuladas como consecuencias de eventos climatológicos considerados anómalos (temperaturas extremas, sequías, lluvias sobre abundantes, heladas, granizadas, etc.)

PC : Población de la comuna (Miles de habitantes)

PA : Población activa de la comuna¹

IP : Ingreso per cápita de los habitantes de la comuna (Miles de dólares)

ND : Número de días (acumulados) de duración total de los eventos climatológicos ocurridos²

NI : Nivel de infraestructura de la comuna, clasificada en Precario o Firme

AG : Área geográfica a la que pertenece la comuna, clasificada en Área Norte o Sur.

a) Defina el modelo econométrico que permita explicar las pérdidas monetarias en función de las variables señaladas, si desea con dicho modelo estimar, ceteris paribus, el efecto promedio sobre Y de:

- i. La población
- ii. La población activa
- iii. El Ingreso per cápita
- iv. El número de días acumulados de eventos climatológicos
- v. El efecto diferencial debido al nivel de infraestructura
- vi. El efecto diferencial debido al área geográfica
- vii. El efecto diferencial de ser una comuna con infraestructura precaria del área norte
- viii. El efecto diferencial del número de días de los eventos al tratarse de una comuna con infraestructura precaria

Señale además qué representa cada coeficiente del modelo definido. **(3 puntos)**

Resp:

Dado los requerimientos se debe definir un modelo ANCOVA con variables dicótomas que interactúen.

¹ Persona activa es aquella que trabaja recibiendo remuneración)

² Si hay dos eventos y uno dura 3 días y otro dura 5 días, en total será una duración de 8 días

Como se desean los efectos diferenciales, el modelo debe incluir Intercepto, por lo que se debe incluir una sola variable dicótoma para no caer en la trampa de la variable dicótoma (NI y AG)

Entonces, sea:

NI=1 si el nivel de infraestructura de la comuna es Precario y 0 en otro caso **(0.2 puntos)**

AG=1 si la comuna pertenece al área Norte y 0 en otro caso **(0.2 puntos)**

Luego el modelo es **(0.8 puntos)**:

$$Y_i = \beta_1 + \beta_2 PC_i + \beta_3 PA_i + \beta_4 IP_i + \beta_5 ND_i + \beta_6 NI_i + \beta_7 AG_i + \beta_8 (NI_i AG_i) + \beta_9 (NI_i ND_i) + u_i$$

Donde los coeficientes representan, ceteris paribus:

β_1 : Coeficiente de intercepto base asociado al caso en que la comuna pertenece al área Sur y tiene un nivel de infraestructura firme. **(0.2 puntos)**

β_2 : Coeficiente de pendiente asociado al efecto de la población de la comuna **(0.2 puntos)**

β_3 : Coeficiente de pendiente asociado al efecto de la población activa de la comuna **(0.2 puntos)**

β_4 : Coeficiente de pendiente asociado al efecto del Ingreso per cápita **(0.2 puntos)**

β_5 : Coeficiente de pendiente asociado al efecto del número de días totales de duración de los eventos **(0.2 puntos)**

β_6 : Coeficiente diferencial con respecto al caso de intercepto base, asociado al caso en que la comuna tiene un nivel de infraestructura precario. **(0.2 puntos)**

β_7 : Coeficiente diferencial con respecto al caso de intercepto base, asociado al caso en que la comuna pertenece al área Norte. **(0.2 puntos)**

β_8 : Coeficiente diferencial con respecto al caso de intercepto base, asociado al caso en que la comuna es del área Norte y tiene un nivel de infraestructura precario. **(0.2 puntos)**

β_9 : Coeficiente diferencial de pendiente que muestra como cambia el efecto del número de días cuando la comuna tiene un nivel de infraestructura precaria. **(0.2 puntos)**

Nota. Dado que se piden efectos diferenciales el modelo debe ser con intercepto, al menos que se trabaje sin intercepto y se determinen los efectos diferenciales restando los coeficientes de interés.

- b) De acuerdo a lo visto en clases y al libro guía del curso, explique cuáles supuestos del MCRL cree usted que es probable que no se cumpla(n) y explique qué efectos tendrían dichos incumplimientos. **(3 puntos)**

Resp:

- i.- Se espera que haya problema de colinealidad, dado que el número de habitantes activos es un subconjunto del número de habitantes; y porque es muy probable que las comunas con mayor ingreso per cápita tendrán infraestructura firme. **(0.5 puntos)**

Este problema de colinealidad provoca que:

- Los estimadores de MCO, aunque seguirán siendo MELI, presentarán varianzas y covarianzas grandes, lo que hará imprecisa la estimación. **(0.3 puntos)**
- Dadas las mayores varianzas y por lo tanto la menor precisión, los intervalos de confianza de H_0 : El coeficiente de pendiente es 0, son más amplios, por lo que se puede aceptar erróneamente la hipótesis nula. **(0.3 puntos)**
- Dadas las mayores varianzas, los estadísticos t de una o más variables pueden aparecer como estadísticamente no significativos **(0.3 puntos)**
- Los estimadores MCO y sus respectivos errores estándar son sensibles a cambios en los datos, aunque estos sean pequeños. **(0.3 puntos)**

- ii.- Se espera que haya problemas de heteroscedasticidad dado que los datos agrupados por nivel de infraestructura serán tales que a mayor nivel de ingreso por cápita, ceteris paribus, tendrán menores pérdidas (y viceversa) por lo que no se puede esperar que la varianza sea constante. **(0.5 puntos)**

Este problema de heteroscedasticidad provoca:

- Que los estimadores de los coeficientes determinados por el método de MCO, aunque siguen siendo insesgados y consistentes, ya no son eficientes (no tienen varianza mínima) **(0.4 puntos)**

- Dado que los estimadores MCO ya no son MELI, entonces, al usar las varianzas asociadas a dichos estimadores, las pruebas t y F pueden llevar a conclusiones muy equivocadas. **(0.4 puntos)**

2.- En un estudio se encuestó a 5000 familias agrupadas por nivel de ingresos (familiares), con el fin de determinar las posibilidades en favor de que el hijo(a) mayor termine estudios superiores (universitarios o técnico profesional), a través de la siguiente regresión del modelo Logit a un nivel de significancia de 0.05:

$$\hat{L}_i^* = -2.005\sqrt{w_i} + 0.043I_i^* - 0.017H_i^*$$

$$ee = 0.42 \quad 0.0058 \quad 0.0021$$

Donde:

$\sqrt{w_i} = \sqrt{\frac{1}{\sigma_i^2}} = \frac{1}{\sigma_i}$, ponderador de los Mínimos Cuadrados Ponderados para la corrección de la heteroscedasticidad.

\hat{L}_i^* : es el logit ponderado

\hat{I}_i^* : es el ingreso ponderado en miles de dólares al año

\hat{H}_i^* : es el número de hijos ponderado

a) Interprete detalladamente los coeficientes de pendiente de ambas variables regresoras I_i^* y H_i^* **(2 puntos)**

Resp:

Si se toma el antilogaritmo (función “e”) del coeficiente de pendiente de I_i^* , se tendrá el valor aproximado 1.044, que significa que un aumento unitario en el ingreso ponderado, las posibilidades ponderadas en favor de terminar los estudios superiores aumentarán en 1.044, o en término de probabilidades, aumentarán en 4.4%. **(1 punto)**

Del mismo modo, para el caso del coeficiente de pendiente del número de hijos, H_i^* , un aumento unitario en el número de hijos, las posibilidades ponderadas se reducirán en 0.98, o en términos de probabilidades, se reducirán en un 1.7% **(1 punto)**

- b) Qué puede decir de la significancia de los resultados obtenidos. Justifique. **(2 puntos)**

Resp:

Usando

$$t = \frac{\hat{\beta}_i}{ee(\hat{\beta}_i)}$$

Se puede obtener el estadístico t, para cada coeficiente:

$$\hat{L}_i^* = -2.005\sqrt{w_i} + 0.043I_i^* - 0.017H_i^*$$

$$ee = \quad 0.42 \quad \quad 0.0058 \quad \quad 0.0021$$

$$t = \quad -4.774 \quad \quad 7.41379 \quad \quad -8.0952 \quad \quad \textbf{(1 punto)}$$

Los que para un nivel de significancia de 0.05 y 5000 observaciones, son todos significativos. **(1 punto)**

- c) ¿Cuál es la probabilidad que una familia de tres hijos y un nivel de ingresos de 18000 dólares al año tenga un hijo o hija mayor que termine sus estudios superiores? **(2 puntos)**

Resp:

Dada la estimación de la regresión, se puede obtener \hat{L}_i al reemplazar $18\sqrt{w_i}$ en el valor de I_i^* y $3\sqrt{w_i}$ en el valor de H_i^* y dividir todo por $\sqrt{w_i}$ obteniéndose:

$$\hat{L}_i = -2.005 + 0.043 * 18 - 0.017 * 3 = -1.282 \quad \textbf{(1 punto)}$$

Como:

$$\hat{L}_i = \ln\left(\frac{\hat{P}_i}{1 - \hat{P}_i}\right) = -1.282 \rightarrow \frac{\hat{P}_i}{1 - \hat{P}_i} = 0.2775$$

Y, por lo tanto, la probabilidad es:

$$\hat{P}_i = 0.2172 \quad \textbf{(1 punto)}$$

3.- Responda

- a) Explique los cuatro problemas y sus consecuencias, del modelo lineal de probabilidades **(4 puntos)**

Resp:

- i.- Las perturbaciones u_i no están normalmente distribuidas. Esto ocurre porque, al igual que Y_i , los u_i sólo toman dos valores: el 0 y el 1. por lo que también distribuyen Binomial (Bernoulli). **(0.5 puntos)**

Esto trae como consecuencia que no se pueden hacer inferencias sobre los verdaderos parámetros poblacionales. **(0.5 puntos)**

- ii.- Las perturbaciones no son homoscedásticas, porque dado que el problema de modela a través de una distribución Binomial (Bernoulli), con media p y varianza $p(1 - p)$, con p la probabilidad de éxito, las varianzas ya no son constantes, sino que dependen de la dicha probabilidad de éxito, es decir de la media. **(0.5 puntos)**

Esto trae como consecuencia que los estimadores de MCO ya no son eficientes (no tienen varianza mínima) y por lo tanto, las pruebas t y F pueden llevar a conclusiones erradas. **(0.5 puntos)**

- iii.- No hay garantía que los \hat{Y}_i , estimadores de $E(Y_i|X_i)$ obtenidos a partir del método de MCO y que representan probabilidades, estén entre 0 y 1. **(0.5 puntos)**

Esto tiene como consecuencia que no se cumple la restricción fundamental de toda probabilidad y debilita el modelamiento del problema. **(0.5 puntos)**

- iv.- La probabilidad $P_i = E(Y_i|X_i)$ aumenta linealmente con X_i , es decir, la variable regresora tiene un efecto incremental sobre la regresada que es constante siempre. **(0.5 puntos)**

Esto tiene como consecuencia que el modelo puede, en la mayoría de los casos, no ser realista en la representación del fenómeno analizado. **(0.5 puntos)**

- b) Analice los resultados de la regresión que modela los gastos promedios mensuales en salud (en miles de pesos) pagados directamente por las personas por concepto de visitas a médicos y exámenes, de los adultos entre 35 y 50 años de edad, afiliados a Isapres en Chile (no incluye el pago que se le hace a la Isapre por concepto de cotización de salud). Realice todos los cálculos necesarios para apreciar la bondad del modelo y la significancia de las estimaciones de los parámetros. Concluya a partir de su análisis.

	Coeficiente	Error estándar
Intercepto	12,25	0,987
X1	24,779	0,554
X2	35,259	19,99
X3	-13,069	-7,697
X4	-10,889	-1,006

Origen de la variación	Suma de cuadrados promedio (SPC)
Debido a la regresión	3284,988
Debido a los residuos	28,759

Donde:

Y	Gasto promedio mensual de adultos entre 35 y 50 años en Chile
X1	Número de días que dura el tratamiento indicado por médico
X2	Edad
X3	Número de enfermedades al año
X4	Número de días de vacaciones al año

Resp:

A partir de los datos proporcionados, se puede determinar:

	Coeficiente	Error estándar	t	%(ee/Coef)
Intercepto	12,25	0,987	12,4113475	8%
X1	24,779	0,554	44,7274368	2%
X2	35,259	19,99	1,76383192	57%
X3	-13,069	-7,697	1,69793426	59%
X4	-10,889	-1,006	10,8240557	9%

Origen de la variación	Suma de cuadrados promedio (SPC)	Estadístico F
Debido a la regresión	3284,988	114,2246949
Debido a los residuos	28,759	$R^2=0.99$

Por lo que se puede apreciar:

- i. Hay dos coeficientes individuales (de las variables X2 y X3) que no son estadísticamente significativos, por lo que no se rechaza la H_0 respecto a que dichos coeficientes de manera individual son 0. Junto con esto el R^2 es muy alto. **(0.3 puntos)**
- ii. El estadístico F es estadísticamente muy significativo, (valor grande) por lo que se rechaza la H_0 respecto a que de manera conjunta todos los coeficientes son nulos **(0.3 puntos)**
- iii. Se esperaría por la teoría, que el signo de X3 fuera positivo, pero es negativo **(0.3 puntos)**
- iv. Los errores estándar de las variables X2 y X3 son altos. Con respecto al valor del coeficiente respectivo, son ambas cercanos al 60% **(0.3 puntos)**
- v. La teoría podría indicar que hay una correlación positiva entre edad (X2) y número de enfermedades (X3) **(0.3 puntos)**

Conclusión: se sospecha que hay colinealidad entre las variables X2 y X3 **(0.5 puntos)**

(Nota: si se concluye sin el análisis anterior, no hay puntaje)

Formulario:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} ; \quad \hat{\beta}_2 = \frac{\sum (X_i - \bar{X}) * (Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} ; \quad \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} ;$$

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2} ; \quad ee(\hat{\beta}_2) = \frac{\sigma}{\sqrt{\sum (X_i - \bar{X})^2}}$$

$$\text{var}(\hat{\beta}_1) = \frac{\sum X_i^2}{n \sum (X_i - \bar{X})^2} \sigma^2 ; \quad ee(\hat{\beta}_1) = \sqrt{\frac{\sum X_i^2}{n \sum (X_i - \bar{X})^2}} \sigma ;$$

$$r^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum (Y_i - \bar{Y})^2} ; \quad SCR = \sum \hat{u}_i^2 = \hat{\sigma}^2 (n-2) ; \quad r^2 = 1 - \frac{SCR}{SCT}$$

$$z = \frac{x - \mu}{\sigma} ; \quad \hat{\sigma}^2 = s^2 = \frac{\sum (x_i - \bar{x})^2}{(n-1)} ; \quad T = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} ; \quad \bar{X} \pm t_{\frac{\alpha}{2}, n-1} * \frac{s}{\sqrt{n}}$$

$$\Pr \left(\hat{\beta}_i - t_{\frac{\alpha}{2}} * ee(\hat{\beta}_i) \leq \beta_i \leq \hat{\beta}_i + t_{\frac{\alpha}{2}} * ee(\hat{\beta}_i) \right) = 1 - \alpha$$

$$\Pr \left((n-2) \frac{\hat{\sigma}^2}{X_{\frac{\alpha}{2}}^2} \leq \sigma^2 \leq (n-2) \frac{\hat{\sigma}^2}{X_{1-\frac{\alpha}{2}}^2} \right) = 1 - \alpha$$

$$t = \frac{\hat{\beta}_i - \beta_i}{ee(\hat{\beta}_i)} ; \quad F = \frac{SCP \text{ de } SCE}{SCP \text{ de } SCR}$$

$$E(Y_i | X_i) = \beta_1 + \beta_2 X_i = P_i ; \quad P_i = \frac{1}{1 + e^{-(Z_i)}} = \frac{e^Z}{1 + e^Z}$$

$$\hat{P}_i = \frac{n_i}{N_i} ; \quad \hat{L}_i = \ln \left(\frac{\hat{P}_i}{1 - \hat{P}_i} \right) = \hat{\beta}_1 + \hat{\beta}_2 X_i ; \quad \sigma_i^2 = \frac{1}{N_i P_i (1 - P_i)}$$