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

Primera Prueba – 26 septiembre 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: 26 septiembre de 2018

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.- Responda en forma clara:

a) Explique por qué se usa la FRM para estimar la FRP y no estimar ésta directamente. **(1 punto)**

Resp:
Se estima la FRP a través de la FRM porque en la mayoría de los estudios no se cuenta con datos de la población, usados para estimar la FRP **(0.5 puntos)**. Por lo tanto, esta se debe estimar a partir de alguna muestra obtenida de la población y con ella determinar la FRM, que se usa para estimar la FRP. **(0.5 puntos)**

b) Detalle qué representa y las razones por qué se usa el término de perturbación estocástica o error estocástico u_i **(2 puntos)**

Resp:

Representa los efectos de variables no consideradas en el modelo de manera explícita. **(0.4 puntos)**

Se usa por las siguientes razones:

- La teoría que apoya el fenómeno es incompleta **(0.2 puntos)**
- No se conocen algunas variables que deben ser incluidas **(0.2 puntos)**
- Se usan variables regresoras inadecuadas **(0.2 puntos)**
- No es posible disponer de datos para algunas variables **(0.2 puntos)**
- Los datos disponibles tienen errores originados en la medición **(0.2 puntos)**
- La forma funcional entre las variables del modelo se ha definido de manera incorrecta **(0.2 puntos)**
- El problema estocástico modelado siempre tendrá efectos aleatorios **(0.2 puntos)**
- Se desea mantener la simplicidad del modelo **(0.2 puntos)**

c) Señale los supuestos del MCRL y explique lo que cada uno significa **(1.8 puntos)**

Resp:

1) El modelo de regresión es lineal en los parámetros. **(0.1 punto)**
    - Esto significa que la relación entre la variable regresada y los parámetros de intercepto y de pendiente se relacionan de manera lineal o intrínsecamente lineal **(0.1 punto)**

2) Valores fijos de X o Valores de X independientes del término de error **(0.1 punto)**
    - Esto significa que los valores de las variables regresoras pueden ser establecidos de manera fija, **(0.1 punto)** o

- Que pueden no ser fijos, y ser muestreados con la variable regresada, pero en este caso, se debe tener las variables regresoras deben ser independientes del termino de error, es decir, que la covarianza entre ellas sea nula. **(0.1 punto)**

3) Valor medio de la perturbación igual a 0 **(0.1 punto)**
   - Esto significa que los factores o variables no incluidos en el modelo de manera explícita (contenidos en u_i) no influyen de manera sistemática el valor de la media de Y (variable regresada) **(0.1 punto)**
   - Esto significa que no hay error de especificación **(0.1 punto)**
   - También significa que no hay correlación entre X e u_i **(0.1 punto)**

4) Homoscedasticidad de u_i **(0.1 punto)**
   - Esto significa que la varianza de u_i para cada X_i es constante por lo que las distintas poblaciones de la regresada Y correspondientes a distintos valores de X tienen la misma varianza. Es decir, los cambios en los valores de las regresoras no hacen cambiar la variabilidad de los términos de error. **(0.1 punto)**

5) No hay autocorrelación entre las perturbaciones ui **(0.1 punto)**
   - Esto significa que el modelo no tiene problemas de autocorrelación y en este sentido no está mal especificado **(0.1 punto)**

6) El número de observaciones n debe ser mayor que el número de parámetros por estimar **(0.1 punto)**
   - Se debe tener suficientes datos para al menos resolver los sistemas de ecuaciones normales mediante las que se calculan los estimadores de MCO **(0.1 punto)**

7) No todos los valores de X en una muestra determinada deben ser iguales **(0.1 punto)**
   - Esto significa que la varianza de la variable X debe ser mayor que cero. **(0.1 punto)**
   - Además, no debe haber valores atípicos (muy distintos en relación a la mayoría) **(0.1 punto)**

d) Explique la importancia del teorema de Gauss Markov **(1.2 puntos)**

Resp:

Este teorema señala: "Dados los supuestos del modelo clásico de regresión lineal, los estimadores de mínimos cuadrados, dentro de la clase de estimadores lineales insesgados, tienen varianza mínima, es decir, son MELI". **(0.2 puntos)**
Su importancia radica en que si se cumplen los supuestos del MCRL, los estimadores de los parámetros de intercepto y de pendiente de la FRM,

obtenidos por el método de MCO, serán MELI **(0.2 puntos)**, es decir, serán los mejores estimadores:

- Lineales en los parámetros **(0.2 puntos)**
- Insesgados, es decir, el valor esperado del estimador corresponde al valor del parámetro buscado. (para cada uno de los estimadores determinados por MCO) **(0.2 puntos)**
- Eficientes, es decir, tienen varianza mínima dentro de la clase de los estimadores lineales **(0.2 puntos)**

Esto es importante porque implica que se obtienen mejores inferencias respecto de los parámetros poblacionales buscados en la FRP. **(0.2 puntos)**

2.- Responda en forma clara:

a) Explique para qué sirve el coeficiente de determinación muestral, r^2. **(1 punto)**

Resp.

El **coeficiente de determinación** (muestral), r^2, es la medida más conocida de la "bondad de ajuste" de una línea de regresión. **(0.3 puntos).** Esto porque mide la proporción o porcentaje de la variación total en la regresada, Y, que puede ser explicada por el modelo de regresión definido. **(0.3 puntos).** Es decir, el grado en que la variación de la variable regresora determina la variación promedio de la variable regresada. **(0.4 puntos)**

b) Detalle porqué se usa el supuesto de normalidad de los u_i **(3.5 puntos)**

Resp:
El supuesto de normalidad de los ui se usa por:

1) **Tiene efectos prácticos sobre la modelación**, dado que muchos fenómenos estudiados tienen involucradas variables normalmente distribuidas. **(0.5 puntos)**
2) **Se aprovecha la sencillez de la distribución normal**, dado que sólo requiere dos parámetros (la media y la varianza). **(0.5 puntos)**
3) **Para el uso del TLC**. Dado que los $u_i$representan la influencia combinada de variables regresoras no incluidas explícitamente en el modelo, se puede demostrar que si existe un gran número de v.a. independientes con idéntica distribución, entonces, con pocas excepciones, la distribución de su suma tiende a ser normal cuando crece el número de variables. **(0.5 puntos)**
4) **También aplica una variante del TCL**, que establece que, aunque el número de variables no sea muy grande, o si estas variables no son estrictamente independientes, su suma puede estar aún normalmente distribuida. **(0.5 puntos)**

5) **Se puede aprovechar la propiedad** que señala que cualquier función lineal de variables normalmente distribuidas estará también normalmente distribuida. **(0.5 puntos)**
6) **Se facilitan las pruebas de hipótesis** sobre los coeficientes del modelo, pues dado que los u_i están normalmente distribuidos, los estimadores de los coeficientes de intercepto y de pendiente, $\hat{\beta}_1$ y $\hat{\beta}_2$, también estarán normalmente distribuidos. **(0.5 puntos)**
7) Para el caso que **se trabaje con muestras pequeñas**, se pueden derivar las distribuciones de probabilidad exactas de los estimadores de MCO y también usar las pruebas estadísticas *t*, *F* y $\mathbf{X}^2$**(0.5 puntos)**

c) Responda de manera concisa:
   i. ¿Cuál es la hipótesis nula de la prueba t de los coeficientes de regresión estimados y qué significa dicha hipótesis en el modelo de regresión? **(0.5 puntos)**

Resp:
Que el coeficiente de regresión (de intercepto o de pendiente) es cero, es decir, en el caso de que el intercepto es cero, que la regresión pasa por el origen; en el caso de los coeficientes de pendiente, que la variación de la variable que acompaña a dicho coeficiente, no influye en la variación de la regresada. **(0.5 puntos)**

   ii. ¿Cuál es la hipótesis nula de la prueba F del análisis de varianza en un modelo con una variable regresora? **(0.5 puntos)**

Resp:
En el modelo estudiado de una variable regresora, la hipótesis nula corresponde a que el coeficiente de pendiente es cero, es decir que la variación de dicha variable regresora no influye en la variación de la regresada. **(0.5 puntos)**

   ii. ¿Cuál es la hipótesis nula de la prueba Jarque-Bera? **(0.5 puntos)**

Resp:
La hipótesis nula de la prueba de JB es Ho: Los residuos están normalmente distribuidos. **(0.5 puntos)**

3. Interprete los resultados al estimar la siguiente regresión, que fue estimada a partir de datos registrados entre los años 1996 y 2014:

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

Donde:
Y_i: gasto total en salud pública en año i (Miles de Millones de dólares)
X_i: número de inmigrantes que tiene el país al inicio del año i (millones de personas)

Resultados de la regresión

	Coeficiente	Error estándar
$\hat{\beta}_1$	0.2033	0.0976
$\hat{\beta}_2$	0.6560	0.1961
r^2=0.397	SCR=0.054	SCE=0.0358
Valor p de estadístico de Anderson – Darling (A^2) = 0.044		

a) Determine la estimación de la varianza de la regresión **(1 punto)**

Resp:

Como $SCR = \sum \hat{u}_i^2 = \hat{\sigma}^2(n-2)$

Usando el valor dado de SCR y considerando que

n=19, pues: n=(2014-1996+1)=19 **(0.5 puntos)**

se tiene que

$$\hat{\sigma}^2 = 3.176 * 10^{-3}$$ **(0.5 puntos)**

b) De acuerdo a lo visto en clases, ¿Qué puede decir de la precisión de los estimadores? **(1 punto)**

Resp:

Calculando el nivel de precisión como el ee sobre el valor estimado, se tiene que la estimación de ambos coeficientes es imprecisa **(0.4 puntos)**, pues para el caso del coeficiente de intercepto, el ee equivale al 48% del valor estimado **(0.3 puntos)**, mientras que para el coeficiente de pendientes equivale a un 29.89%.**(0.3 puntos)**

c) Determine el estadístico t de cada coeficiente y señale la significancia estadística de cada una de las estimaciones realizadas (use la tabla de la distribución t de student) **(1 punto)**

Resp:

Considerando el estadístico t para el coeficiente i:

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

y como la Ho para este test es β_i=0 **(0.1 punto)**
se tiene que

$$t = \frac{\hat{\beta}_i}{ee(\hat{\beta}_i)} \quad \textbf{(0.1 punto)}$$

Luego, para el coeficiente de intercepto: t= 2.083 **(0.2 puntos)**
Y para el coeficiente de pendiente: t= 3.346 **(0.2 puntos)**

De acuerdo a la tabla t, para un nivel de significancia de 5% (dado que no se dice nada, se asume ese valor por defecto) y 17 grados de libertad, para una prueba de dos colas, el valor crítico es: 2.110. **(0.2 puntos)**
Dado el valor crítico, el coeficiente de intercepto no es significativo y el coeficiente de pendiente sí es significativo. **(0.2 puntos)**

OJO: No se puede usar la regla práctica de 2t porque el número de grados de libertad es menor a 20.

d) ¿Qué puede decir en general de la regresión estimada? Fundamente detalladamente. **(2 puntos)**

Resp.

- El coeficiente de determinación es bajo lo que es un indicio que no es una buena regresión dado que el porcentaje de la variación de la variable regresada explicada por la variable regresora es muy bajo. (r^2=0.397). Puede ser que le falten variables regresoras (o explicativas) al modelo. **(0.5 puntos)**
- Por otro lado, los coeficientes estimados son poco precisos lo que hace menos confiable la estimación realizada. **(0.5 puntos)**
- El coeficiente de intercepto no es estadísticamente diferente de cero, por lo que se puede pensar en una regresión que pase por el origen si esto es lo que sugiere la teoría que sustenta el modelo. **(0.5 puntos)**
- El test de Anderson-Darling, tiene un valor p que resulta estadísticamente significativo, por lo que se rechaza la hipótesis que los residuos estén distribuidos normalmente (Hipótesis nula de dicho test). **(0.5 puntos)**

e) Interprete el coeficiente de pendiente. **(1 punto)**

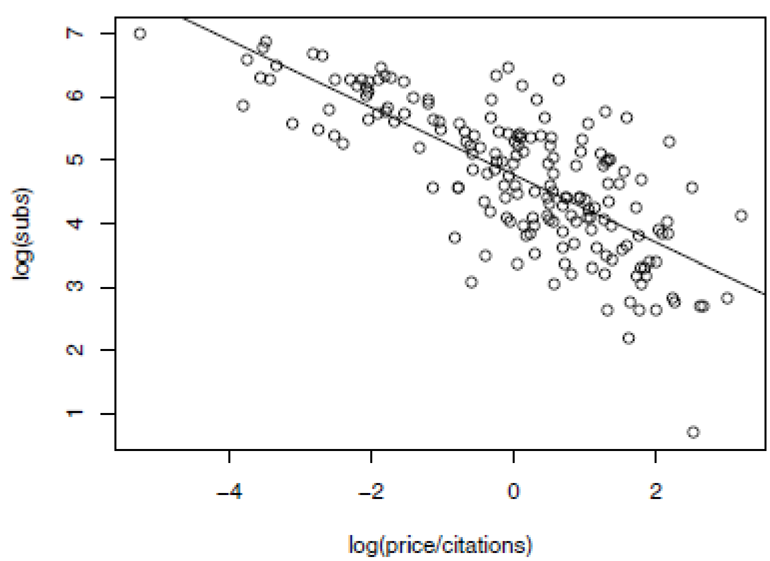
Resp.
Dado el modelo lin-log **(0.3 puntos)**, se tiene que una variación de un 1% en el número de inmigrantes que tiene el país al inicio del año **(0.3 puntos)**, se tendrá un aumento de 0.006560 Miles de Millones de dólares en el gasto total en salud pública

en dicho año. **(0.4 puntos)** (Es decir, el gasto en salud pública aumentará en 6.56 millones de dólares ese año.)

4. Los siguientes resultados corresponden a la estimación de un modelo econométrico simple que trata de explicar la relación entre la cantidad de suscripciones a revistas económicas en USA el año 2000 (subs) a partir de la razón entre la variable precio y la variable número de citas de dichas revistas (Price/citations).[1]. interprete y analice detalladamente dichos resultados, incluido el tipo de datos usados. **(6 puntos)**

```
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)            4.7662     0.0559    85.2   <2e-16
log(price/citations)  -0.5331     0.0356   -15.0   <2e-16

Residual standard error: 0.75 on 178 degrees of freedom
Multiple R-squared: 0.557,       Adjusted R-squared: 0.555
F-statistic:  224 on 1 and 178 DF,  p-value: <2e-16
```



Resp

- Los datos corresponden a datos transversales, dado que se ha medido el número de suscripciones (variable subs), el precio (variable Price) y el

[1] Ejemplo tomado del libro "Applied Econometrics with R" de Christian Kleiber y Archim Zeileis, de la editorial Springer. Páginas 1-2.

número de citas (variable citations) para distintas revistas de Estados Unidos en el mismo tiempo (año 2000) **(0.6 puntos)**

- Este es un modelo lineal simple, con forma funcional log-log, doble log o log-lineal **(0.6 puntos)**
- Dada la forma funcional de las variables, el coeficiente de pendiente proporciona la elasticidad (precio/citación) de la cantidad de suscripciones, de manera directa. Así, esta elasticidad es negativa y equivalente a -0.5331. **(0.6 puntos)**
- La relación negativa entre el logaritmo de la variable dependiente y el logaritmo de la regresora se aprecia además en el gráfico reportado. **(0.6 puntos)**
- El signo de esta elasticidad es respaldado por la teoría económica, que señala que, para la mayoría de los bienes, la cantidad demandada decrece con el precio. Así, el aumento en un uno porciento de la relación precio a citas implicará una disminución de 0.53% en el número (o cantidad) de suscripciones. **(0.6 puntos)**
- Los resultados del modelo son bastante buenos:
  - El coeficiente de determinación es superior a 0.5. es decir, la variación de la relación precio a citas explica al menos la mitad de la variación en la cantidad de suscripciones. **(0.6 puntos)**
  - Los signos de los coeficientes son los esperados **(0.6 puntos)**
  - Las estimaciones son bastante precisas. Para el intercepto, el error estándar corresponde a 1.1% del valor estimado, mientras que, para el coeficiente de pendiente, corresponde al 6.7%. **(0.6 puntos)**
  - Los coeficientes son estadísticamente muy significativos de acuerdo al valor p de los estadísticos t de cada estimador. (prácticamente 0) **(0.6 puntos)**
  - El valor del estadístico F y su correspondiente valor p confirman la significancia estadística del coeficiente de pendiente, por lo que se rechaza la hipótesis nula que el valor de ese parámetro sea nulo. **(0.6 puntos)**

Formulario:

$$FRP: Y_i = \beta_1 + \beta_2 X_i + u_i \;\; ; \;\; FRM: Y_i = \hat{Y}_i + \hat{u}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$$

$$\hat{\beta}_2 = \frac{\sum(X_i - \bar{X}) * (Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2}; \qquad \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

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

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

$\sum(Y_i - \bar{Y})^2 = \sum(\hat{Y}_i - \bar{Y})^2 + \sum \hat{u}_i^2$; SCT=SCE+SCR

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

$$t = \frac{\hat{\beta}_i - \beta_i}{ee(\hat{\beta}_i)} \;\; ; \operatorname{Pr}\left(-t_{\frac{\alpha}{2}} \le \frac{\hat{\beta}_2 - \beta_2^*}{ee(\hat{\beta}_2)} \le t_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$\operatorname{Pr}\left(\beta_2^* - t_{\frac{\alpha}{2}} * ee(\hat{\beta}_2) \le \hat{\beta}_2 \le \beta_2^* + t_{\frac{\alpha}{2}} * ee(\hat{\beta}_2)\right) = 1 - \alpha$$

$$F = \frac{SCP\ de\ SCE}{SCP\ de\ SCR} = \frac{(\hat{\beta}_2^2 \sum x_i^2)}{\hat{\sigma}^2} \;\; ; JB = n\left[\frac{S^2}{6} + \frac{(\mathrm{K}-3)^2}{24}\right]$$