

Trabajo 1: Econometría – Análisis de Datos Empíricos

Universidad de Santiago de Chile
Magíster en Economía
Semestre 1, 2019

Este trabajo es una oportunidad para aplicar las técnicas que vemos en clases a datos empíricos que son interesantes para ti. Para hacer esto, es necesario contar con una base de datos. Estos datos pueden estar relacionados con cualquier cosa. Los únicos requisitos son (a) que los datos están libremente disponibles, (b) que tienen (por lo menos) cuatro variables que están de alguna manera correlacionadas entre si (c) que por lo menos una de estas 4 variables es discreta, y por lo menos una es continua, y (d) que el tamaño total (el N) de los datos es mayor a 100.

Hay muchas fuentes de datos—tanto a nivel micro como a nivel macro—que podrían ser consultados. Algunos ejemplos incluyen el [portal de datos abiertos de Chile](#), el [banco de datos del Banco Mundial](#) (también disponible usando el programa [wbopendata](#) de Stata), encuestas, incluyendo [la CASEN](#), [la Encuesta de Protección Social](#) y [la Encuesta Longitudinal Docente](#), o datos recién publicados, como los resultados de la prueba [SIMCE](#). Puedes bajar y utilizar cualquier base que te parece interesante para este trabajo, incluyendo datos que no están en este listado de sugerencias.

La entrega consiste en un documento de trabajo corto (entre aproximadamente 1000 y 2000 palabras), y la base de datos y código que utilizas para hacer el análisis detallada a continuación. El análisis se puede hacer con cualquier paquete de software, pero se sugiere utilizar Stata, ya que es probable que en otros cursos del magíster utilizarán este programa. La entrega se hace por correo electrónico (a damian.clarke@usach.cl) antes de las 23:59 del jueves 25 de abril de 2019. Si ocupan Stata, la entrega debe consistir del documento (en formato .pdf), el archivo .do que contiene el código del análisis, el archivo .ado que escriben para la pregunta C4, y la base de datos en formato .dta. Por favor, cuando entreguen la base de datos, solamente incluye las variables que utilizaron en sus trabajos (no toda la base con variables que no utilizaron). Si utilizan algún otro idioma computacional, los archivos .do, .ado y .dta deben ser reemplazados por los archivos correspondiente en este idioma. **Todos los archivos entregados deben ser enviados en un solo archivo .zip llamado “NombreApellido.zip” (con tu nombre y apellido en vez de “Nombre” y “Apellido”).** Si no tienen un programa de compresión en tu computador para crear archivos .zip, hay sitios que permite hacerlo, por ejemplo: <https://archivo.online-convert.com/es/convertir-a-zip>.

Evaluación:

Este trabajo forma 15% de la nota final del curso. La evaluación tomará en cuenta la parte escrita, y también el código de análisis que demuestra los cálculos de las estadísticas que utilizas. La nota final se calculará de acuerdo a las siguientes características:

Elemento	Porcentaje
Código: ¿Es replicable en otro computador?	10%
Informe: Parte A – Resultados y redacción	23%
Informe: Parte B – Resultados y redacción	27%
Informe: Parte C – Resultados y redacción	38%
Presentación y organización del documento	2%

Análisis:

(Parte A) Estadísticas Descriptivas Abre tu set de datos, y genera las siguientes estadísticas (para las 4 variables de interés) para tener una idea de cómo las variables se “ven”:

1. La esperanza, y la desviación estándar
2. la cantidad total de observaciones
3. El valor mínimo y máximo
4. La función de probabilidad, o la función de densidad de probabilidad

Para los puntos 1-3, crea una tabla que contiene una fila para cada variable, y una columna para cada estadística, para presentar las estadísticas descriptivas de cada variable juntas. Incluye esta tabla y los cuatro gráficos del punto 4 en tu documento de trabajo, y describe brevemente los patrones que se ven en las variables. ¿Son “razonables” los valores promedios, mínimos, y máximos utilizando tu intuición económica? Asegura que has limpiado las variables correctamente (por ejemplo que no hay valores como 999 cuando la variable no tiene una respuesta).

Pista: Existen algunos programas como `estpost` y `estout` para generar tablas así y exportarlas a \LaTeX y MS Word de forma automática. Es necesario instalar estos paquetes utilizando el comando `ssc install estout`.

(Parte B) Análisis Univariada

1. Utilizando una variable continua y la f.d.p. de la parte A4, agrega una curva normal al gráfico con la misma esperanza y varianza de la variable ($\mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$). Es posible que tu variable no es normal, y en este caso, la distribución normal teórica no se verá parecida a la distribución empírica.
2. Imagina que la variable empírica fuese normal, y calcula el punto de la distribución abajo de donde cae 5% de la masa de probabilidad (es decir, $F^{-1}(0.05)$).
3. Por último, cuál es el valor real de la distribución empírica abajo de donde cae 5% de la masa de probabilidad (nota que para calcular un cuantil de una variable en Stata, puedes usar el comando `sum var, d`, donde `var` es el nombre de la variable).

Incluye el gráfico del punto 1 en tu documento de trabajo, y también el cálculo de los dos puntos en la distribución (el teórica y el verdadero). Describe brevemente la distribución empírica, y su relación (o falta de relación) con la distribución normal teórica. ¿Cómo comparan los valores calculados en las partes 2 y 3?

(Parte C) Análisis Bivariada En esta parte, consideraremos dos variables: una continua (C), y una discreta (D). Debes tener alguna hipótesis o razón para creer que las variables tienen alguna correlación entre si. Con estas dos variables:

1. Genera un gráfico de la distribución de C condicional en $D = d$ en dos distintos puntos. Uno de los puntos de D , ($D = d_{\text{bajo}}$) debe ser un valor bajo, y el otro, ($D = d_{\text{alto}}$), debe ser un valor alto.
2. Agrega una línea que representa la esperanza condicional de C (condicional en el valor particular de $D = d$) en cada gráfico.

3. Demuestra que la ley de expectativas iteradas cumple con $E[C]$ (y todos los valores posibles de D), mediante el cálculo de $E_D[E[C|D]]$.
4. Revisando “Clase Computacional: Escribiendo Programas Computacionales” (pp. 56–59) en los apuntes del curso, escribe un programa simple (un ado en Stata) que permite calcular la expectativa de una variable mediante la fórmula de expectativas iteradas como en la pregunta anterior (C3). Este programa puede tener una estructura bastante simple, como por ejemplo un `syntaxis` como `espiterada var1 var2` donde `var1` es la variable que intentamos calcular su promedio, y `var2` es la variable sobre cual estamos iterando.

Pista: Para iterar sobre todos los niveles de una variable, el comando `levelsof` en Stata es muy útil. Se sugiere consultar el archivo `help` de `levelsof` para un ejemplo muy relevante.

Incluye los dos gráficos en tu documento de trabajo. Describe la relación que existe considerando $f(C|D = d_{\text{bajo}})$ y $f(C|D = d_{\text{alto}})$. ¿Es la relación que esperaste? ¿Cual es la correlación y covarianza global entre las dos variables? ¿Piensas que esta correlación refleja una relación causal entre las dos variables? ¿Por qué, o por qué no? Además, incluye el archivo `ado` que escribiste con su entrega de este trabajo. Nota que si utilizas algún programa que no sea Stata para realizar este trabajo, debería entregar el equivalente a un programa escrito en este idioma (eg una función en R y MATLAB, etc.)