



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

ESCUELA DE VERANO EN MÉTODOS MIXTOS - 2018

Prof. Francisco Urdinez

Ayudante de laboratorio: Andrés Cruz

Laboratorio práctico – tiempo estimado 2 horas

1. Analizaremos una base de datos sobre desigualdad para una muestra de 118 países a mediados de la década de 1990.

Comience cargando la base de datos en R, usando el siguiente comando:

```
inequality <- read.csv("c:/<Path Where You've Saved the File>/inequality.csv")
```

2. Mire los nombres de las variables en la base de datos:

```
names(inequality)
```

¿La democracia de un país causa un aumento relativo en la igualdad económica?

Lleve a cabo una regresión bi-variada, utilizando el coeficiente de Gini (como medida de la desigualdad social) como la variable dependiente y la variable Polity (como medida de democracia) como la variable independiente. Informe todos los resultados relevantes de este análisis. ¿Qué significan los diversos tests-t? ¿Qué significa el test-F? ¿De qué manera el test-F es similar o diferente del test-t para el coeficiente Polity? ¿Qué expresan los dos valores diferentes de R^2 ?

3. Ahora debe expandir el modelo incorporando el PBI como una segunda variable independiente (use la variable logarítmica del PBI), para así controlar el nivel de desarrollo de un país como un posible factor de confusión. Los comandos R serían:

```
ineqlm2 <- lm(Gini ~ Polity + log(GDP), data=inequality)
summary(ineqlm2)
```

4. Ahora, basado en la Tabla de selección de casos de Seawright (a continuación) determine cuáles serían los 4 países más recomendables para utilizar como caso de estudios siguiendo los criterios de (a) caso típico, (b) caso desviado, (c) caso influyente; (d) extremo en X y (e) extremo en Y. Esta tarea puede demandarle mucho tiempo así que la haremos en grupo si hay dificultades.

Selection Rule	Implementation
Random	Select case(s) randomly with equal probability from the entire data set.
Typical	Select case(s) to minimize $ Y_i - \hat{Y}_i $.
Deviant	Select case(s) to maximize $ Y_i - \hat{Y}_i $.
Influential	Select case(s) to maximize the DFBETA connected with X_i .
Extreme X	Select case(s) to maximize $ X_i - \bar{X}_i $.
Extreme Y	Select case(s) to maximize $ Y_i - \bar{Y}_i $.
Most Similar	Select paired case(s) that such that $X_{i,1}$ is “high” and $X_{i,2}$ is “low” while minimizing the overall difference between $Z_{i,1}$ and $Z_{i,2}$.
Most Different	Select paired case(s) that minimize $ X_{i,1} - X_{i,2} $ while maximizing the overall difference between $Z_{i,1}$ and $Z_{i,2}$.
Contrast	Select paired case(s) that maximize $ X_{i,1} - X_{i,2} $ while minimizing the overall difference between $Z_{i,1}$ and $Z_{i,2}$.
Pathway	Select case(s) for which $ Y_i - \hat{Y}_i $ is made as much smaller as possible by including X in the regression.

Table 4.1: Techniques for Choosing Cases.

5. En 1855, el epidemiólogo inglés, John Snow, publicó un libro que identificaba correctamente el modo de transmisión del cólera. Este libro se ha convertido posteriormente en un caso paradigmático de inferencia causal exitosa para los metodólogos de las ciencias sociales. El libro está disponible en línea aquí:

<http://www.ph.ucla.edu/epi/snow/snowbook.html> Examine el estudio, identificando los componentes cuantitativos y cualitativos de la inferencia causal. (Tenga en cuenta que las estadísticas no estaban completamente desarrolladas en ese momento, por lo que los componentes cuantitativos pueden carecer de algunas características que esperamos en una investigación contemporánea, como valores p).

Concéntrese principalmente en la Parte 3 del libro que explica qué contribuye cada componente a la inferencia causal. ¿La investigación de métodos mixtos de Snow hace que la inferencia causal sea más poderosa de lo que sería un estudio puramente cuantitativo? Si es así, explica por qué.

6. Retomando la regresión del punto 3, y usando las herramientas de John Snow, diseñe un plan de investigación para el análisis de estudios de casos que pueda probar suposiciones de inferencias causales para esta regresión. Los dos casos en los que debe centrarse son la República del Yemen y Zimbabwe. Describe la investigación que has diseñado para evaluar cada supuesto de interés, y si su estrategia conlleva alguna modificación al análisis de regresión.
7. Observando los datos, determine qué regla de selección de casos se utilizó en la pregunta anterior para seleccionar a la República del Yemen y Zimbabwe. Realice un estudio de caso similar al de la pregunta anterior, pero esta vez usando cada una de las siguientes reglas: caso típico, caso extremo sobre Y, caso extremo sobre X. ¿Cómo se comparan las inferencias generadas por estos estudios de casos con las que produjo en la pregunta 5? ¿se le ocurren variables que pueden mejorar su modelo?
8. En este ejercicio, utilizaremos matching para estudiar con métodos mixtos si la democracia causa (o reduce) la desigualdad económica. Cree una versión dicotómica del indicador de régimen Polity:

```
inequality$polbin <- 1*(inequality$Polity > 0)
```

Cargue con R el siguiente paquete de matching

```
install.packages("Matching")
```

```
library(Matching)
```

Ahora, construya un puntaje de propensión inicial para la democracia:

```
dempensitylogit <- glm(polbin ~ log(GDP) +  
Industry + FuelExports + CommunistLegacy,  
data=inequality, family=binomial(link=logit))  
inequality$pscores <- NA  
inequality$pscores[as.numeric(  
names(dempensitylogit$fitted))]  
<- dempensitylogit$fitted
```

Ahora descartamos los datos faltantes y realizamos el matching:

```
ineqtrim <- inequality[!is.na(inequality$Gini) &
!is.na(inequality$pscores),]
ineqmatch <- Match(Y=ineqtrim$Gini, Tr=ineqtrim$polbin,
X=ineqtrim$pscore, estimand="ATE")
summary(ineqmatch)
```

Los propensity scores son útiles para elegir casos por medio de la técnica de par más similar y par más diferente (ver tabla 4.1) ¿Cuáles serían sus pares más similares y más diferentes en este caso? ¿Qué concluyes de estos resultados? ¿Cuáles son las suposiciones clave que se realizan para llegar a la inferencia causal con este análisis? ¿Qué concluye sobre el análisis original y cómo lo rediseñaría?