UCLA Extension Data Science Intensive

Instructor: William Yu

Project 2

10/1/2018

- Submit your results (including R script and any output you got) through Canvas.

A. Regression: An Example of Turning A Small Dataset Into Knowledge

- Read my article: "Will A Lower Corporate Tax Rate Boost Economic Growth?"
- In the Course Website, download p02_Corporate tax.csv and save it into your computer.
- Complete the R script to answer the following questions:
- On Page 55, there are three equations. Replicate these regression results by using R.
- Plot a chart similar to Figure 4 by using R. (The red line is a regression, fitted line).
- The dataset provides more variables (description as follows). Play around by adding these variables. And present the best model and briefly explain why.

	OECD 35 Countries
Variable	Description
ypcg	GDP per capita growth rate, average from 2000 to 2015
ctax	Corporate tax rate (%), average from 2000 to 2008
ypc2000	GDP per capita (US$) in 2000
dty	Debt to GDP ratio (%), average from 2000 to 2008
trade	Trade (imports and exports) as percentage of GDP (%)
ihc	Index of country's human capital
y2000	GDP in 2000 (economy size, US billion $)

B. Calculate the City Human Capital Index (CHCI) from the American Community Survey (ACS) Data

- Browse the following webpage about CHCI developed by me:
- http://www.anderson.ucla.edu/centers/ucla-anderson-forecast/projects-and-partnerships/city-human-capital-index
- We want to calculate the CHCI by a weighting average of educational attainment of local adult residents. See the following table. The data is from the 5-year American Community Survey summary file (DP02) from 2009 to 2016. Note: in addition, there are DP03, DP04, and DP05 files. And the data is obtained through American Fact Finder (https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t). We might revisit these dataset later on.
- Download all the ACS data into your computer. The data is shown by all the counties in the U.S. _metadata is the documentation for variables. _with_ann is the data file.

- You need to use the following variables: #2: Id2, #3: Geography, #232: Adult population, #238: % in less than 9th grade, #242: % in 9th to 12th grade, #246, #250, #254, #258, and #262.
- Hint: the formula is CHCI = 50*% in less than 9th grade+100*% in 9th to 12th grade+…
- Add two variables into the data frame: CHCI growth from 2009 to 2016, and CHCI growth rate from 2009 to 2016.
- Your expected output should look like chci.csv attached in the folder.
- Reorder the data based on the following variables (e.g. you can use "arrange" function under library dplyr.
- Select the top 20 counties for (1) CHCI in 2016, (2) the growth of CHCI from 2009 to 2016, (3) adult population in 2016, (4) population growth from 2009 to 2016, (5) population growth rate from 2009 to 2016, and bottom 20 counties for (6) CHCI in 2016, (7) the growth of CHCI from 2009 to 2016, (8) population growth from 2009 to 2016, and (9) population growth rate from 2009 to 2016. Your expected output should look like chci.top20.csv.

The CHCI is calculated based on the following education attainment data (**multiplied by 10**) for adult residents in each area:

Education Attainment	Assigned Schooling Year
Less than 9th grade	5
9th to 12th grade, no diploma	10
High school graduate (includes equivalency)	12
Some college, no degree	13
Associate's degree	14
Bachelor's degree	19
Graduate or professional degree	23