Assignment 3

Released: 09/15/2020 Due: 09/29/2020

The objective of this homework is to identify appropriate statistical methods for given problems, apply and interpret statistical techniques, evaluate applications of statistical techniques and use a statistical application software package.

This assignment will get you working with Stata and check the review of descriptive statistics (Chapters 1 and 2 from OIS)

You are provided with a dataset titled "Florida County Data.csv" that includes the following variables:

1. **RW**: The typical ratio of the wage for a specific job (e.g. framing carpenters) in each county relative to the state average wage for that job in 2013.
2. **PCI**: 2012 per capita income.
3. **POP**: 2012 population.
4. **WDEN**: Weighted density. Population density, typically calculated as population divided by land area, is of limited usefulness as much land area is irrelevant, e.g. land in a national park can't be developed. Instead, imagine asking residents how many people per square mile live near them, and taking the average of their answers. This would measure the average density perceived by residents. Weighted density measures this by calculating density at fine geographic levels and taking a resident weighted average.
5. **SH65UP**: The share of the 2012 population age 65 or older. This proxy the relative importance of in-migrant retirees in the local economy.
6. **SHLH**: The share of 2012 employment in the Leisure and Hospitality sector. This proxy the importance of tourism in the local economy.

The goal of this assignment is to write a brief report to tell someone who knows nothing about FL, how relative wages vary across the state. Use State to do the following analyses prior to writing anything.

1. Calculate summary statistics for each variable.
2. Produce box and whisker plots and histograms for each variable.
3. Calculate summary statistics for each variable, weighted by county population. That is, calculate the statistics from the point of view of individuals, not counties. This involves modifying the summarize command using analytic weights. Use help and documentation to figure out the syntax. (The point of this one is mainly to make you read through the help documentation on your own.)
4. Create new variables that equal the natural log of POP, WDEN, PCI, and RW. Name them lnPOP, lnWDEN, lnPCI, and lnRW.
5. Produce the correlation matrix for lnRW, lnPCI, lnPOP, lnWDEN, SH65UP, and SHLH. (You might type something like "help cor" to find the right syntax.)

6. Produce scatter plots for lnRW (vertical axis) against lnPCI, lnPOP, lnWDEN, SH65UP, and SHLH.

Report Requirements

Do not just turn in a bunch of code or tables of figures from 1-6 followed by some bullet points. Instead, write up your description in the form of a professional report on an exploratory analysis of the data. Include all the summary statistics, the correlation matrix, and the figures you actually refer to in your report (neatly and sized appropriately) in the body of your report, with appropriate labels and captions. If you do not refer to some parts of the output from 1-6, place them in an appendix, again appropriately labeled or captioned, rather than in the body of the report.

More Practice for Extra Credit (20 points)

a) Change the markers to hollow circles and use weighting to scale the circles in each plot to be proportional to POP (except for the plot against lnPOP).
b) Change the labels on the titles on the scatterplots to be self-explanatory.
c) Try making the scatterplots using a log scale with the untransformed variables—you will get axis labels that are easier to understand but still the log scaling.
d) See if you can figure out how to create a variable that is the first 4 letters of the county names and then make a set of scatter plots with those shortened names as marker labels.
e) Use "graph matrix" to produce a matrix of scatter plots of all interesting bivariate relationships.

Grading Rubric

Analysis – 50 points
Write-Up – 25 points
Report organization and formatting – 25 points. Please make sure your report includes the following at a minimum

- cover page
- table of contents
- list of figures and tables
- Executive Summary (similar to abstract)
- Professional headings for each section
- captions for tables and figures
- Conclusion
- Appendix (includes Stata do-file)