Valeria Salinas-Lopez

Fall 2018

Sep 25, 2018

SOC 3305

Lab 3

(Due in class on September 27, 2018)

Please save your final work as lastname_lab3.doc (put your name above the title page) and email it to instructor. Please explain your results with your own words. (You must use Stata program to do this lab). Copy your Stata command for each question.

1. **Use Canada1.dta from data we used in class. Generate a type as a new variable and assign 1 for nation, 2 for province and 3 for Yukon and Northwest territories. Provide histogram of unemployment by type. (20 pts) (copy paste the histogram to word document)**

   I first used generate type=2 in order to create the type column. I then used replace type=1 if place=="Canada" in order to change the type for the nation of Canada. Then I used replace type=3 if place=="Yukon" | place =="Northwest Territories" in order to change the type for Yukon and the Northwest Territories. The provinces remained as type 2.

2. **Please explain with your own words what are "append", "merge" and "collapse" commands do and how they are different from each other. (30 pts)**

    Append: This command is used in order to add data onto the same spreadsheet of older and saved file. The data is added to the bottom of the other data. This makes the dataset longer.

    Merge: This is used in order to add data from a previous dataset to the dataset you are working on that has matching variables. This makes the dataset wider.

    Collapse: This command breaks down the data into means, medians, sums, range, and more. This doesn't use 2 datasets, just the one being worked on.

3. **Use Granite2011_6 data from class dataset. Generate genint3 variable similar to what we did in class. Make sure genint3 is equal to five subtracted by genint variable if genint is less than 10. Label new variable you generated as "Interest in 2012 election – brand new" and define genint3 as "1"Not very" 2 "Somewhat" 3 "Very" 4 "Extremely" .a "DK" .b "NA" ". Is there a difference from what we did in class? If yes/no, why? (20)**

    I used generate genint3= 5-genint if genint<10 and then manually changed the label by editing the spreadsheet. Then to define genint3 I used label define genint3 1"Not very" 2 "Somewhat" 3 "Very" 4 "Extremely" .a "DK" .b "NA" and the only difference between that command and the one we did in class is that the name of the variable was genint2 not genint3. The command was the same so there is no difference.

4. **Use Arctic9 data and convert it to wide format. (10)**

   **Command:** reshape wide tempN, i(year) j(month)

5. **Explain histogram, box plots, scatter plots and matrix plots. Give an example for each plot using Nations2 data and copy each graph to your word document. Also, explain your graphs. (Avoid using same examples that we did in class, use different variable). (20 pts)**

    Histograms: This form of graph is done with one variable. The bars in the graph are right next to each other and connected with no space in between.
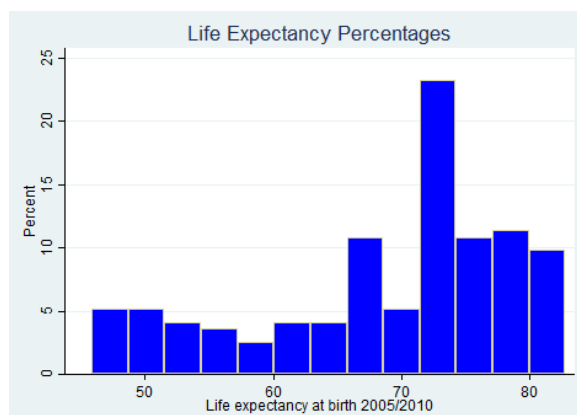
Explanation: About 23 or 24 percent of the population of the study, have a life expectancy of about 71 to 73 years of life from birth.

**Box Plots:** Show the center (median), spread (range), and outliers of the data.
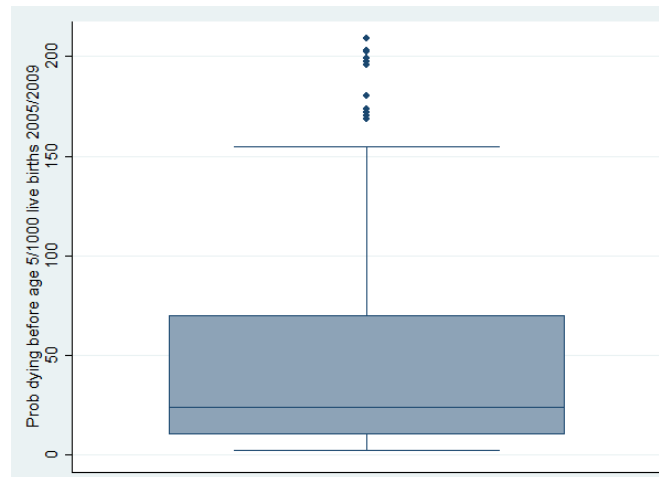
Explanation: The median shows that the probability of dying before age 5 is about 25 deaths out of 1000 live births. The range is from 0 because that is the lowest logical number (you cannot have negative deaths) and goes up to about 155 deaths per 1000 live births. The data shows that many countries have outliers, or numbers that are higher than the median and the range of the data.

Scatter Plot: This graph compares two variables and shows the relationship between them.
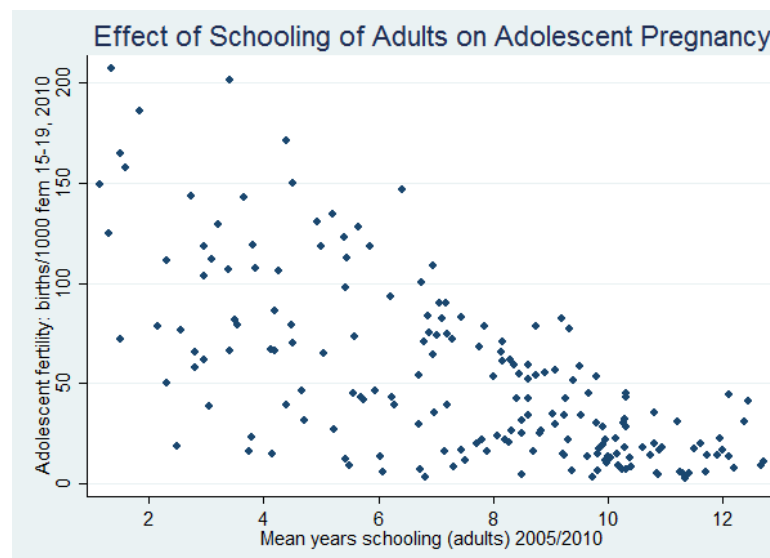
Explanation: There is a negative indirect correlation. As the schooling of the adults in a country increases, adolescence fertility of the country decreases.

Matrix Plot: This graph allows you to compare as many variables are you would like from the data you have.
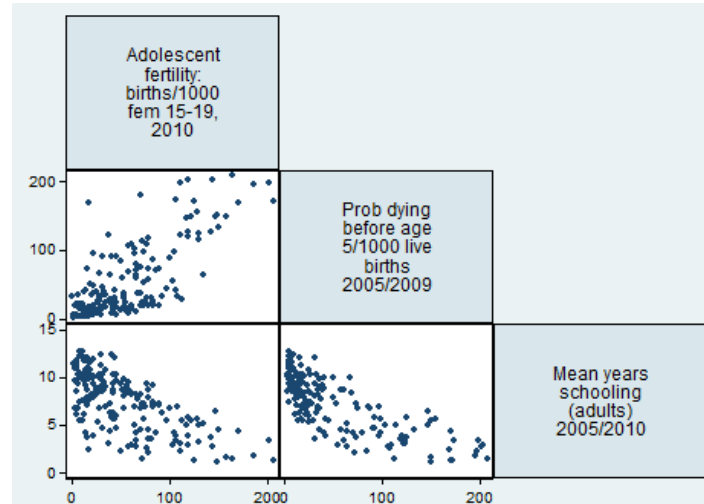
Explanation: Here the matrix plot shows that adolescent fertility and the probability of dying at age 5 are directly correlated. As adolescent fertility goes up so does the probability of dying by age 5. Adolescent fertility is inversely related to the mean of years of schooling of adults, as adolescent fertility goes up, the mean in years in schooling of the adults is going down. The probability of dying by age 5 is inversely related to the mean of years of schooling of adults. As the mean of years in schooling increases, the probability of death by age 5 decreases.