

Principles of Biostatistics

Lab 1

High density lipoprotein cholesterol (HDLc) is the good kind of cholesterol which contributes to better health.

The file HDLc.xlsx contains data on the HDLc levels for a sample of sedentary and active men. The goal of this lab is to familiarize yourself with some of the descriptive statistics tools in SPSS by exploring this data set.

Instructions

1. Go to my web site faculty.tarleton.edu/crawford and click on the link for Math 350.
2. Download the file HDLc.xlsx (Right-click and “Save Link As...” to save it somewhere. You might want to save it to your student drive).
3. Start SPSS (called PASW statistics now).
4. Choose “Open an existing data source” and click OK.
5. Browse to find your file and change the file type to Excel. Open your file.
6. Make sure “Read variables names...” is checked and click OK.
7. Go to “Analyze”, “Descriptive Statistics”, “Explore”.
8. Put HDLc in the dependent list and Lifestyle in the factor list.
9. Click the Statistics button, and make sure “Descriptives” and “Percentiles” are both checked.
10. Click the Plots button, and make sure “Histogram” is checked. “Stem-and-leaf” doesn’t need to be checked.
11. Click OK. SPSS will now generate all of the statistics you need to complete this lab.
12. Answer the questions on the next page in an MS Word document. Make sure your document has a heading (your name) and that each problem is numbered. Copy graphs from SPSS into your word document when appropriate. Your goal is to correctly answer the questions *and* to present your results in a professional manner.

Questions

1. What are the average HDLc levels for the sedentary and active groups? The medians? Which group has a higher average HDLc level?
2. Find the range, standard deviation, and variance for both groups. Do they both exhibit about the same amount of spread, or does one group exhibit much more spread than the other?
3. What is the smallest and largest value in each data set?
4. Copy the histogram for each data set into your Word document. Are these histograms consistent with the medians and smallest/largest values you obtained in the previous problems? Where are the mean and median located in each histogram (center, far left, or far right)?
5. Find the percentiles in the SPSS output.
 - (a) What percentage of active men have HDLc levels under 64?
 - (b) What percentage of active men have HDLc levels above 39.25?
 - (c) What percentage of sedentary men have HDLc levels between 19.43 and 44.35?
 - (d) What percentage of sedentary men have HDLc levels between 13.04 and 34.40?
6. Copy the boxplot into your Word document.
 - (a) Are these boxplots consistent with the means, percentiles, and max/min values for each data set?
 - (b) Based on the boxplots, which group appears to have higher HDLc levels?
 - (c) How much overlap exists between the two groups in terms of HDLc levels?
 - (d) Would you be able to easily determine if someone is active or sedentary based on his HDLc level?
7. Based on this data, does an active lifestyle appear to promote more healthy levels of HDLc?