

**K300 (4392) Statistical Techniques (Fall 2007)**  
**Assignment 2: Descriptive Statistics (Due September 17)**

Instructor: Hun Myoung Park  
kucc625@indiana.edu, (317) 274-0573

This assignment contains exercises for summarizing univariate variables using graphical and numerical methods (total 140 points). Please read the instructions carefully to do the homework successfully.

- You **MAY NOT** use a wordprocessor (e.g., Microsoft Word and WordPerfect); **Use separate sheets and write down answers by hand.** Exception is SPSS outputs for questions 12-14.
- Follow all instructions of a question. Do not skip any one.
- Try to show how you obtain the answers (statistics). In some questions, a single number may not be accepted as the answer.
- Useful examples and sections of the textbook are italicized for you.
- Put your answer sheets into an envelope and hand in on Monday, September 17.
- You **MAY NOT** discuss with other classmates when answering the questions.

If you have any problem with any of the questions, please post messages on Oncourse CL or come and see the instructor during office hour MW 2:00-3:00 P.M. Or you may make an appointment with the instructor.

**1. (10 points)** Classify the level of measurement (nominal, ordinal, interval, and ratio) of the following variables. Note that some variables may be used as more than one level of measurement depending on research questions and data at hand. *See page 6-9 and question 7 on page 26.*

- 1) What was your GPA last semester?
- 2) What kind of beer do you prefer?
- 3) Where are you from?
- 4) What do you think about Indiana property tax increase? (Strongly agree—Agree—Indifference—Disagree--Strongly disagree)
- 5) Age ( $\leq 20$ ;  $\leq 25$ ;  $\leq 30$ ;  $\leq 40$ ;  $\leq 50$ ;  $\leq 60$ )
- 6) How many credit hours did you take last semester?
- 7) Grade (from 1<sup>st</sup> through 12<sup>th</sup>)
- 8) Fahrenheit temperature at 9:00 A.M. in IUPUI
- 9) Your monthly gas consumption (gallons)
- 10) which one (1 through 9), if any, can be treated as more than one level of measurement? And why? You may make your own assumptions to support your arguments, if needed.

**2. (10 points)** Explain what the stratified sampling is and then apply it to a study on student satisfaction at IUPUI/SPEA. Given different numbers of students enrolled in each major/department this semester, present your strategy to obtain a representative sample.

Ignore other variables such as gender and race. Sample size is limited to 500. *See page 12 and question 12 on page 26.*

**3. (10 points)** Solve question 9 on page 44. You need to construct a frequency table with five classes and compute the mean on the basis of the frequency table you construct. And then compute the mean of raw data. Which one is larger? *See pages 37-41 for examples.*

**4. (10 points)** Solve question 14 on page 44. And draw a stem-and-leaf plot with 8 stems (leading digits) by hand. *See pages 73-76.*

**5. (10 points)** Solve question 16 on page 59. You may skip the frequency polygon and ogive for the data. *See Figure 2-8 on page 56 for description of the histogram.*

**6. (10 points)** Solve question 18 on page 79. Keep in mind that whole numbers are nonnegative integers (i.e., 0 and positive integers). You may have 10 stems (leading digits). *See example 2.14 on page 76.*

**7. (5 points)** Solve question 24 on pages 97-98. Draw a Pareto chart by hand. *See pages 64-65.*

**8. (5 points)** Solve question 12 on page 117. Simply compute the mean only. You SHOULD show each step you follow as showing in *the example 3-3 on pages 105-106.*

**9. (5 points)** Solve question 15 on page 117. You do not need to compute the mean in order to answer. *See "Properties and Uses of Central Tendency" on page 114.*

**10. (10 points)** Solve question 1 on page 163. Add "25" to the series of data so that you have 8 data points. Obtain key descriptive statistics (minimum, 1Q, 2Q, 3Q, maximum, and mean) and draw the box plot based on the statistics. A box plot may be arranged either horizontally or vertically (either one will do). And then mark all data points using "O" on the box plot. Put "X" on the mean position of the box plot. Is the mean larger than the median? *See pages 147-149 and 159-161 for obtaining the key statistics.*

**11. (15 points)** Solve question 7 on page 135 but compute the mean, variance, and standard deviation ONLY. You SHOULD show steps for computation as we did on an Excel sheet during the lab. DO NOT use Excel though; DO it by hand. *See examples 3-21 (step 6) and 3-22 (step 3) on pages 123-125.* Keep in mind the data points were drawn from the population; you need to compute sample variance and standard deviation (as opposed to the population counterparts).

(Continued...)

\* Now let us turn to our data set. Download the SPSS data set of class survey from OnCourse or course webpage at <http://www.masil.org/method/statistics.html>. If you need questionnaire and data dictionary, download [http://www.masil.org/teach/k300/K300\\_Survey.pdf](http://www.masil.org/teach/k300/K300_Survey.pdf).

Launch SPSS 15 for Windows and read the data set stored in your computer.

**12 (15 points)** Draw a frequency table of `major`. Click Analyze→Descriptive Statistics→Frequencies.... Choose `major` and click arrow to move the variable into the right. Click “Statistics” button on the bottom and check mean, median, variance, and “std deviation.” Click Continue and then OK to get a frequency table. 1) Which major has the smallest and largest frequency? 2) What is the proportion (relative frequency) of Health Administration of the total? How can you calculate by hand? 3) What are the mean and variance of `major`, 4) How do you interpret the mean and variance substantively? This question may remind you of the level of measurement. 5) Report the median. Can you interpret the median substantively? 6) print out the SPSS output and append it to your assignment 2.

**13 (10 points)** Produce descriptive statistics of `credits`. Click Analyze→Descriptive Statistic→Descriptives... and then choose `credits`. Click Options... on the right bottom, check sum, variance, and range, and then click Continue. 1) Print out the SPSS output and attach it to your assignment. 2) Report N, sum, and mean. 3) Show how the mean was calculated using these statistics. 4) Show how the standard deviation was calculated from the variance.

**14 (15 points)** Draw a histogram of `distance`. Click Analyze→Descriptive Statistic→Explore... Choose `distance` and click arrow so that the variable appears on a box under “Dependent List:” Click the “Statistics...” option at the bottom, check Percentiles, and click Continue. Now click “Plots” next to “Statistics...”, check Histogram, and then click Continue. Finally click OK to get descriptive statistics and the histogram. 1) Print out SPSS output and attach to your assignment. 2) Report the mean and variance. 3) Report the key statistics used in a box plot. 4) Look at the box plot. Write down five key statistics and the mean at their right positions on the box plot. Do you have any outliers (extremely large or small)? 5) Look at the histogram and stem-and-leaf plot. Is the histogram similar to the stem-and-leaf plot? If not, what do you think makes a difference? Note that graphs and charts oftentimes are sensitive to scale, number of classes, and outliers. Tell me your idea about difference and similarity of the histogram and stem-and-leaf plot of `distance`. 6) Finally, which statistics (mean or median) are you going to report as a representative value (central tendency) of `distance`? Which plot (histogram, box plot, stem-and-leaf plot) do you think will be best for summarizing this variable? And why?