

MIT OpenCourseWare
<http://ocw.mit.edu>

11.220 Quantitative Reasoning & Statistical Methods for Planners I
Spring 2009

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Quantitative Reasoning and Statistical Methods
(11.220)
Problems Set #2

Ezra Haber Glenn, AICP

Due: In section 3/5/09 or 3/6/09

1 Problems

1.1 Some Calculations

1. An instructor in a large class gave an exam to 1,386 students. The lowest grade was a 41. The mean was 78 and the median was 80. The standard deviation was 11.93. The interquartile range (IQR) was 16.5. The max was 100. How many students scored above the 92nd percentile?
2. (a) Find some data giving the population and area of each of the 50 U.S. states. Change the figures for area into hectares. Change the figures for population into 1,000-people. With this data set, please tell us which state represents (or is closest to) each figure in the “Five Number Summary”:¹
 - i. for land area;
 - ii. for total population;
 - iii. for total population per hectare.(b) Present these figures in plain English so that someone unfamiliar with these terms would understand what they tell us about the size and population of these states.
(c) The standard deviation of the size of the states, in square miles, is 88,278. From this, can you compute the standard deviation of the areas of the states in hectares?
3. You have been analyzing a dataset of the salaries of 367 municipal employees for the town of Weymouth, and have computed the following “Five Number Summary”:

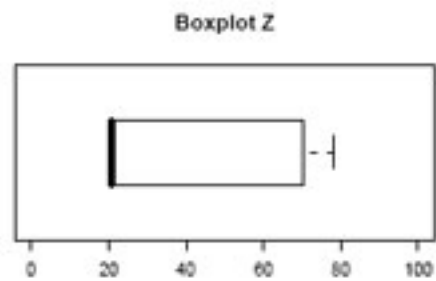
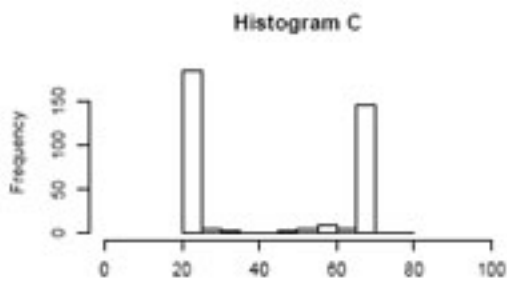
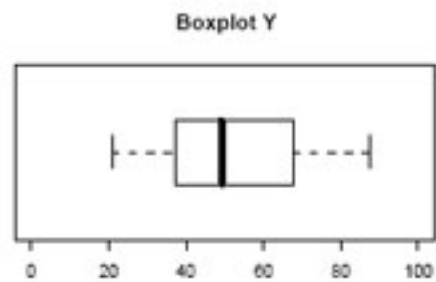
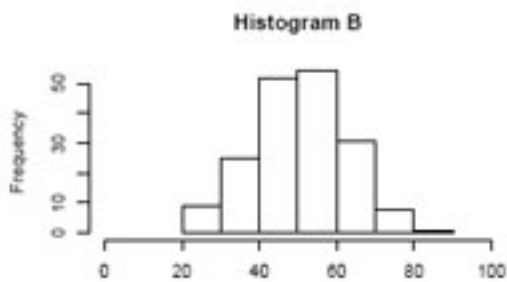
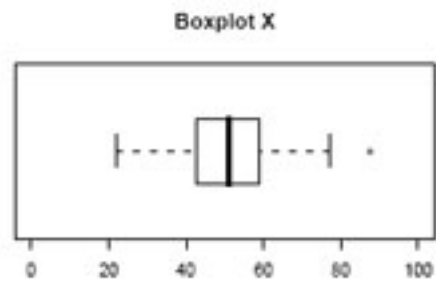
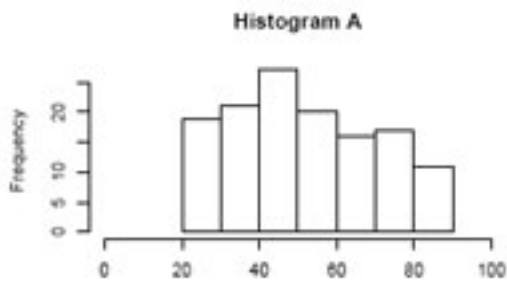
¹Actually, please do this for an “augmented” five number summary, showing min, max, median, quartiles, and the mean.

min	1stQ	median	mean	3rdQ	max
18,367	27,800	30,106	34,787	32,666	97,800

After doing all this work, you learn that the salary data did not include a \$600 bonus that all 367 employees received at the end of the year. Please recompute the table to include this extra income.

1.2 Looking at Plots

For the following plots, match each histogram with its accompanying boxplot.



1.3 Playing with Numerical Summaries

1. Create an imaginary dataset with 10 different values where the median is greater than the mean. Pretend that the values stand for the results of water quality testing for arsenic levels (typically measured in parts per billion, but don't worry too much about reality here.) You can assume this to be at a *ratio* level of measurement.
2. Add one value to make the mean greater than the median.
3. Compute the range, IQR, and the standard deviation. Present these figures in plain English so that someone unfamiliar with these terms would understand what they said about these different water quality samples.
4. Make a “dot plot” of the data on a number line, label the “five number summary” and the mean, and depict the IQR and the standard deviation.

1.4 Numbers in the News

Find an article in the newspaper or a magazine that presents *descriptive statistics*. Briefly describe the figures presented, and critique the presentation. What else might you want to know to put these numbers in a context?

1.5 Assembling Data

1. Create a dataset of your own with three variables and 15-20 observations (cases) of each. One of the variables must be nominal, and at least one should be at the interval or ratio level of measurement. For example, you could gather REGION (e.g., “West”, “Midwest”, “Northeast”, etc.), POPULATION, and LAND AREA for each state in the U.S.² You can download the data from some on-line source, or gather it yourself in “raw” form, or develop it in some combination, but either way you should be thinking about the following questions:
2. For each variable:
 - (a) What is the level of measurement?
 - (b) Is the data discrete or continuous?
 - (c) How was it gathered, and what problems (if any) do you anticipate concerning “noise” and “bias”?
 - (d) Are there any missing values? Are all possible values represented?
3. Enter everything in a table (in a spreadsheet or word processing program, or preferably Stata or R) and save it as a “.csv” (text) file.

²Please *don't* do these exact ones—use your imagination a bit and find something interesting to you.

4. Add a new variable that is based on a calculation of two of the existing variables (for example, in the case of states, you could add a variable showing people per square mile).
5. Compute the mean, median, mode, standard deviation (as a population), and IQR for each variable.
6. Prepare a written description of the “metadata” or “codebook” for the dataset, so that someone unfamiliar with the dataset would know where these numbers came from and how they are to be understood.
7. Email the data and a text file of the metadata to your TA.