Midterm 1 Study Guide

This is meant to be a representative sampling of the key concepts you will need to know, and it is not meant to be exhaustive. You should make sure that you are comfortable with Quizzes 1-5 and Homework Assignments 1-8, as well as the recommended problems from the book.

1. In what way does the data we use in statistics differ from just a bunch of numbers?
2. What are the main types of variables? What distinguishes them?
3. What types of graphs do we have available? When do we use which graph? What are advantages/disadvantages of each type of graph compared to other graphs for the same variable?
4. What terms do we use to describe a scalar variable? Demonstrate with examples.
5. What are the various measures of center? What are the advantages and disadvantages of each?
6. The IQR and the standard deviation both measure spread, but they do so in totally different ways. Explain in what way they attempt to measure "spread".
7. What happens to the shape, center and spread of a distribution of a variable when it undergoes a linear transformation?
8. How does the "suspected outlier test" work?
9. How is the (modified) boxplot drawn?
10. Suppose a distribution is skewed to the right. How will that show in the boxplot?
11. What does it mean to say that a measure is *robust*?
12. In terms of the z-values, what are the first and third quartile for the normal distribution?
13. What percent of values in a normal distribution would be classified as outliers?
14. How do we find where the middle 40% of data lies in a normal distribution?
15. How do we go back and forth between p, z and x in a normal distribution?
16. What graphs are appropriate when we want to examine the relationship between two variables? What types of variables does each apply to?
17. What terms do we use to describe the relationship between two scalar variables?
18. What does the correlation coefficient r measure?

19. When we have a regression line fitted to some data, explain the following terms: predicted value, actual value, residual, sum of squared residuals.

20. What is the key property that makes the "*least squares regression line*" special?

21. What is the meaning of "r-squared" in the context of the least squares regression line?

22. How does the *residual plot* work? What do we expect from it if we have a suitable fit?

23. Outliers far in the x direction and only far in the y direction affect the least squares regression line in very different ways. Explain.