Quiz 2 - Review

Descriptive Statistics

Be able to interpret:

-Box Plots and Histograms

-Mean, Median, Standard Deviation, and Percentiles.

Correlation

Correlation is a summary statistic measuring the association between two variables.

- The correlation coefficient is always between 1 (perfect positive correlation) and -1 (perfect negative correlation)
- A positive correlation means that as one variable increases, the other one tends to increase as well
  A negative correlation means that as one variable increases, the other one tends to decrease

Example #1: Using Correlation

- In the United States the correlation between income and education is 0.44, the respective standard deviations for each are $20,000 and 3 years.
- How would the education level of a person with an income $30,000 higher than average compare to the average person's education?

Linear Regression Modeling

The Equation:

$$Y = \alpha + \beta X$$

Where alpha is the intercept and beta is the slope.

Note: Beta is related to r (the correlation coefficient).

-A 1 unit increase in X corresponds with a Beta unit increase in Y.

-A 1 standard deviation increase in X corresponds with an increase of r standard deviations in Y.

Example #2: Using Regression

- For the 2,649 adult women in the NHANES data set, the regression intercept is -137 pounds and the slope is 4.8 pounds/inch where X = height, Y = weight

Part A) predict the weight of a woman who is 5’6’’

Part B) how will an increase in height of 1 inch effect the predicted weight? How about an increase of 2 inches?

Part C) can you predict the height of a woman who weighs 170 lbs using this model?

Probability

- Law of total probability (addition rule)
- Complement rule
- Independence
- Conditional probability
- Bayes Theorem

Example #3 – Basic Probability

Suppose 10% of the patients taking drug A experience gastrointestinal (GI) side effects, whereas 5% of patients taking drug B experience such side effects.

If the side effects of the two drugs are independent events, then what would be the probability that a person taking both drugs will experience GI side effects (i.e., GI side effects from at least one of the two drugs)?

Example #4 – Basic Probability

Suppose that, on a given day, 60% of UI students take the bus to their first class, while the other 40% walk. Suppose that if you take the bus, you have a 5% chance of being late, while you have a 10% chance of being late if you walk.

Part A: What is the probability that a student will take the bus and be late to class?

Part B: What is the probability that a student will walk and be late to class?

Part C: What is the probability of being late to class?

Part D: A student arrives to class late. What is the probability that he/she took the bus that morning?

Example #5 – Bayes Theorem

A medical research team wishes to assess the usefulness of a certain symptom (call it S) in the diagnosis of a particular disease. In a random sample of 800 patients with the disease, 780 reported having the symptom. In an independent random sample of 1500 subjects without the disease, 50 reported that they had the symptom.

Part A: Find the Sensitivity and Specificity of using the presence of this symptom as a screening test.

Part B: Find the predictive value positive of the test, assuming that the prevalence is 0.01