Statistics Homework 4

Mr. Grant

July 31, 2017

Due Date: Tuesday, July 18th

This homework can be done either by hand or on the computer. It will be **much** faster if you do it on the computer–trust me. If you do it on the computer, turn in a word document with the graphs pasted into it. If you do it by hand, you will have to draw the graphs neatly by hand. In either case,

- Be neat

- Put your name, the date, and the name of this class on your homework

- Number each problem clearly

- Show your work

These questions all use the data from *class_data_complete.csv*.

1. Make boxplots of the heights of the girls and the boys in the class (use the Shoe Gender variable to determine gender). Put the boxplots on the same graph, so that you can compare the boys to the girls easily. We will learn how to do this in R on Wednesday (or you can Google it!).

2. What do the boxplots tell you about the height of the average woman in the class versus the height of the average man? Why? Which gender has more variability in height, and how can you tell from the boxplot? We already know the answers to these questions from class, so your explanation of how you can tell from the boxplot is what matters.

3. Suppose Booker randomly picks a girl in the class to go on a date with. What is the probability that she is 5'4" or shorter? What is the probability that she is taller than 5'8"? **Given that he has picked a girl who is 5'5" or taller**, what is the probability that Booker chose Idara? What is the probability that Booker goes on a date with a Hawaiian who is exactly five feet tall?

4. Consider the boxplot you made in question 1 and answer the following true/false questions. With each question, give a 1-sentence explanation of why you know your answer is correct.

- The tallest $\frac{3}{4}$ of the boys in the class are taller than the shortest $\frac{3}{4}$ of the girls.
- If Idara were a boy, she would be shorter than Q3.
- If Pearson were a girl, he would be an outlier.

5. Read this article and answer the following questions:

- Which is more likely–a white student getting a degree from the college they started at in 6 years, or a black student getting a degree from any university in 6 years?
- Approximately how many–not what percent, but how many–eligible Asian students did not enroll in college in 2010 (the number of students of each race is shown at the bottom of the graph; for example, there were 322,205 Hispanic students in this study)?
- Do you think the percentages in this study are likely to apply to STEMPrep students? Why or why not?
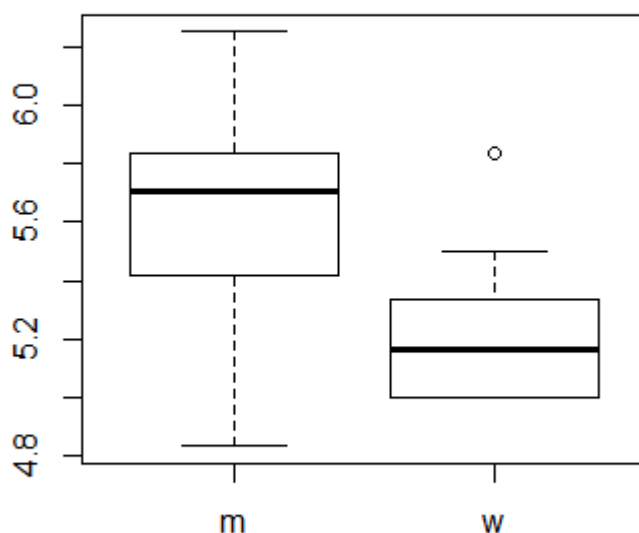
1.

2. We can from the bold black line that the median height of the boys is around 5.7 ft, whereas for girls it is about 5.2 ft. Furthermore, the first quartile of boys' heights is taller than the third quartile of girls' heights. These both indicate that boys are taller than girls on average. The boys' heights appear to be more variable, as well. The range of heights is much larger for boys: their heights run from about 4.8 ft to 6.3 ft, whereas the girls' heights only run from 5 ft to 5.8 ft. The "box" in the boxplot is a bit bigger for boys than girls, additional evidence that boys' variability is higher.

3. See attached code. The probability of a girl 5'4" or shorter is the number of girls 5'4" or shorter divided by the total number of girls, which is .764. The probability of a girl taller than 5'8" is, similarly (see how these first two questions are basically the same question because they require the exact same kind of reasoning), .058. The third part is a conditional probability: we count the number of Idaras (just one, sadly) and divide by the condition: the number of girls 5'5" or taller. That gives us 0.25. For the final part, we need to count the number of people who are girls, Hawaiian, and exactly 5 feet. We can use 3 conditions and put them together with "and" symbols (see code) to get a probability of .176.

4. (a) Yes–the whole boys' box is taller than the whole girls' box.

   (b) It looks like Idara would be right at the third quartile of the boys, so she would technically not be shorter than Q3 (we can use the R quantile function to verify this).

   (c) It appears that Pearson will be an outlier (makes sense). Remember, we can calculate the "interquartile range" as $1.5(Q3 - Q1)$ and add that to Q3 to see what the largest non-outlier would be. The attached R code does this. Turns out that if Idara were any taller, she'd be an outlier too!

5. Looking at the bar graphs in the article, we can see that 47.5% of white students finish at the school they started at in 6 years, whereas 38% of black students finish at any school within 6 years.

6. About 27,260–we simply need 20.0% of the total 136,309 Asian people.

7. These percentages are unlikely to apply to STEMprep students because y'all are not a representative sample of the population as a whole. If you had been randomly selected, these percentages might be close, but because you had to apply, take an entrance exam, and pay for the program, you are not a random sample. Most likely, a higher percentage of you will graduate college within 6 years, because you're academically motivated students!