Jeremy Neiman

# R Challenge #2

## Statement 1:

The null hypothesis is that the median of trip durations is equal to 15 minutes (900 seconds). And the alternative hypothesis is that the median is less than that. If the median is less than 15 minutes, that means that the majority of trips are shorter than 15 minutes. To test the median I used Wilcoxon signed-rank test with a significance level of .05. The resulting P-Value is 0 (or so close to 0 that R can't represent it). Therefore we can reject the null hypothesis that the median is 15 minutes and accept the alternative that it is less. Therefore the majority of trips are indeed trips lasting no more than 15 minutes.

## Statement 2:

First I subseted the data to include only trips greater than or equal to 45 minutes. Then I used the same test as for statement 1, this time with a null hypothesis of 2 hours (120 minutes, 7200 seconds) and an alternative hypothesis that the median is not equal to 120 minutes. The resulting P-Value is $3.566584 \times 10^{-19}$, so we can reject the hypothesis that the median is 120 minutes. The sample size of overtime rides is rather small (only about 1% of the total dataset was overtime) but even so we can confidentially reject the hypothesis that the median is 120 minutes because there is very little evidence in the data we have to support such a claim.

## Statement 3:

I started the same overtime subset of data used in statement 2. The overtime fees are determined by the amount of time over 45 minutes the ride is. The overtime fee structure isn't strictly linear with respect to duration, but using the trip duration as is should be a close enough approximation. To start with, we need to determine whether there is a statistically significant difference between the means of each gender. Using ANOVA we can test a null hypothesis that the means for each gender are equal. We find a P-Value of 0.853, much higher than our significance level of .05. Therefore we can't reject the null hypothesis that the means are unequal and we can't come to a conclusion that men incur more overtime fees.