# Week 7 Assignment – Non-Parametric Data

By: Zach Adair

Regis University

MSDS 660 – Statistical Methods and Experimental Design

## Introduction

For this week’s assignment we are going to be analyzing non-parametric data of food rated at a new chain restaurant named “Souperb”. I will be using the sign test to test the median rating and determining if the median is at least 3 given a confidence interval of 95%. The next part of the assignment I will be taking two different operating systems (M, W) which are rated from 1-10 and testing to see if the operating systems have the same distribution, I will need to determine the statistical test I am going to use and give a hypothesis with a confidence level of 95% and significance level of 0.05.

## Part 1 – Souperb Ratings Analysis

### Load the Data

Since the Souperb data is just a single vector I will just input it in manually.



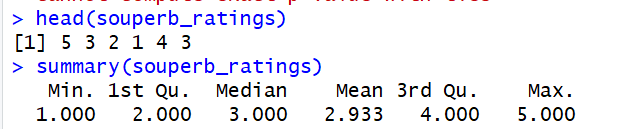
### Set the Hypothesis

Null hypothesis: The median rating of the data will be equal to 3.

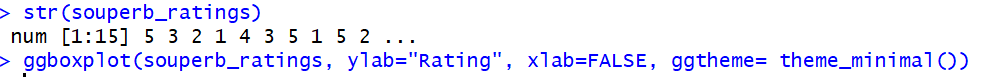
Alternative hypothesis: The median rating of the data will not be equal to 3.

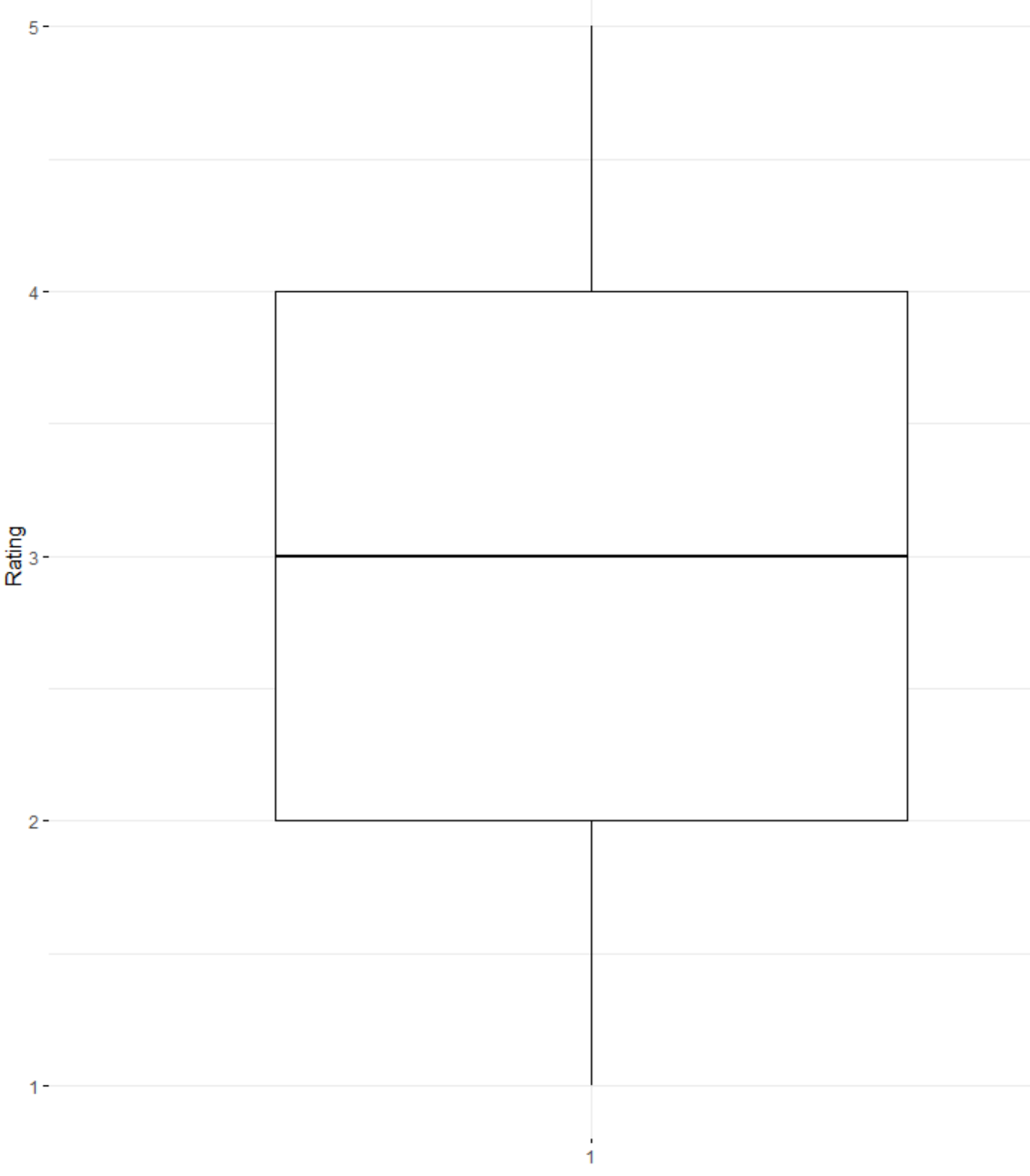
### Explore the Data

The dataset isn’t very big but it can still be explored, in this section I will explore the data and look to see if I can pull meaningful information from it just by doing some simple data exploration.



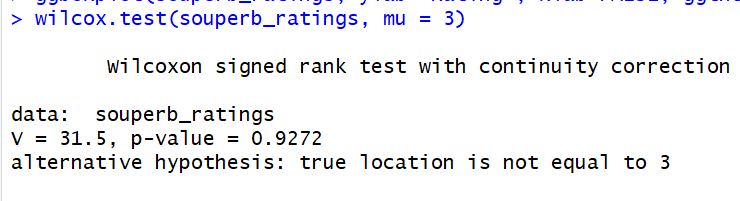
From looking at the summary of our data, I notice right away that our median and mean are either 3 or really close to 3. Which means our null hypothesis could be correct based on what it states about the median being equal to 3.



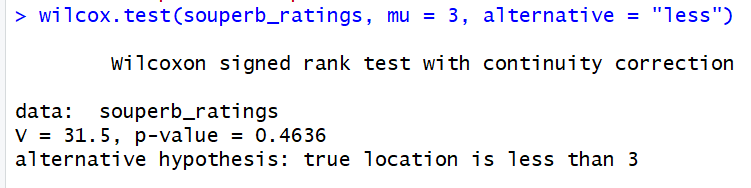


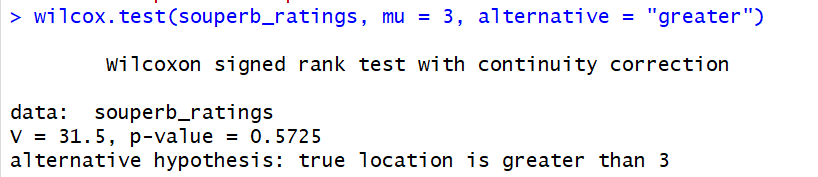
From our boxplot it would appear that our data has a median equal to 3 while the data looks normally distributed. Now this could be because the data points can only be 1 to 5 so we need to use a true non-parametric test to try and find out if the null hypothesis should be rejected or not. The test we are going to use to determine to either reject or accept the null hypothesis is the Wilcoxon Test.

### Wilcoxon Signed Rank Test



After running the Wilcoxon Sign Test on the Souperb ratings that was given by the customers has a median of 3. I was able to determine this based on the Wilcoxon Signed Rank Test which concluded that the p-value from the test was extremely high at 0.9272, drastically above the significance level of 0.05 which means we do not reject the null hypothesis. I wanted to test the data again, this time based on either the alternative being less than or greater than 3.





When both greater and less than alternative hypothesis, the Souperb ratings were deemed by the p-value as not able to reject because they both were higher than the significance level of 0.05, which brings further evidence to the credence that the median of the Souperb ratings is exactly 3.

## Part 2 – Two different Operating Systems

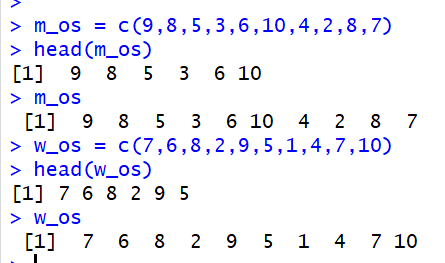
The second half of this assignment is fairly similar to the first one but instead of a rating from 1-5 we are measuring a scale from 1-10 and we are comparing two operating systems to see if they have the same distribution. The two operating systems are M and W.

### Stating the Hypothesis

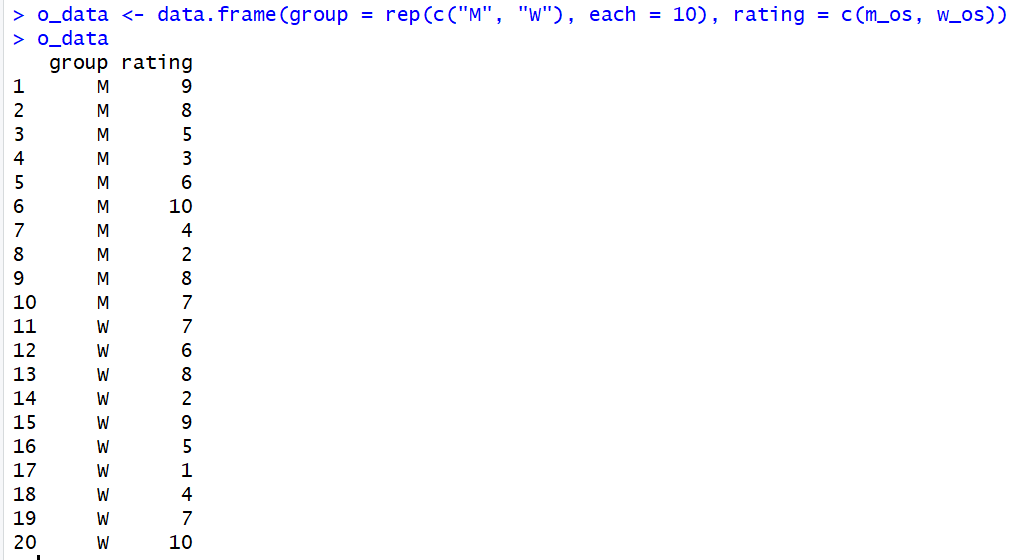
Null Hypothesis: The operating systems M and W have a mean of paired difference that are equal to zero.

Alternative Hypothesis: The operating systems M and W do NOT have a mean of paired differences that are equal to zero.

## Load the Data

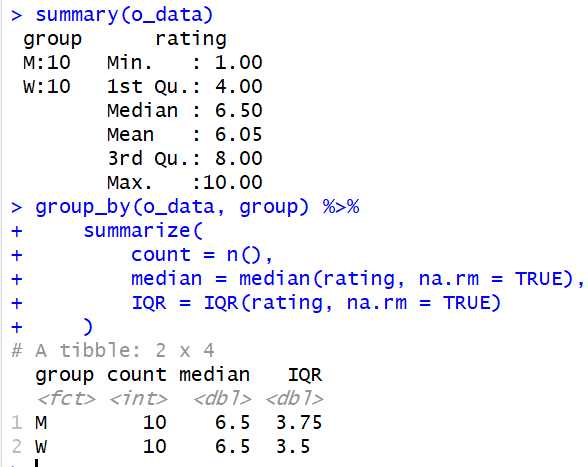


I want to put this data into a data frame so that is the next step here with my data now loaded.



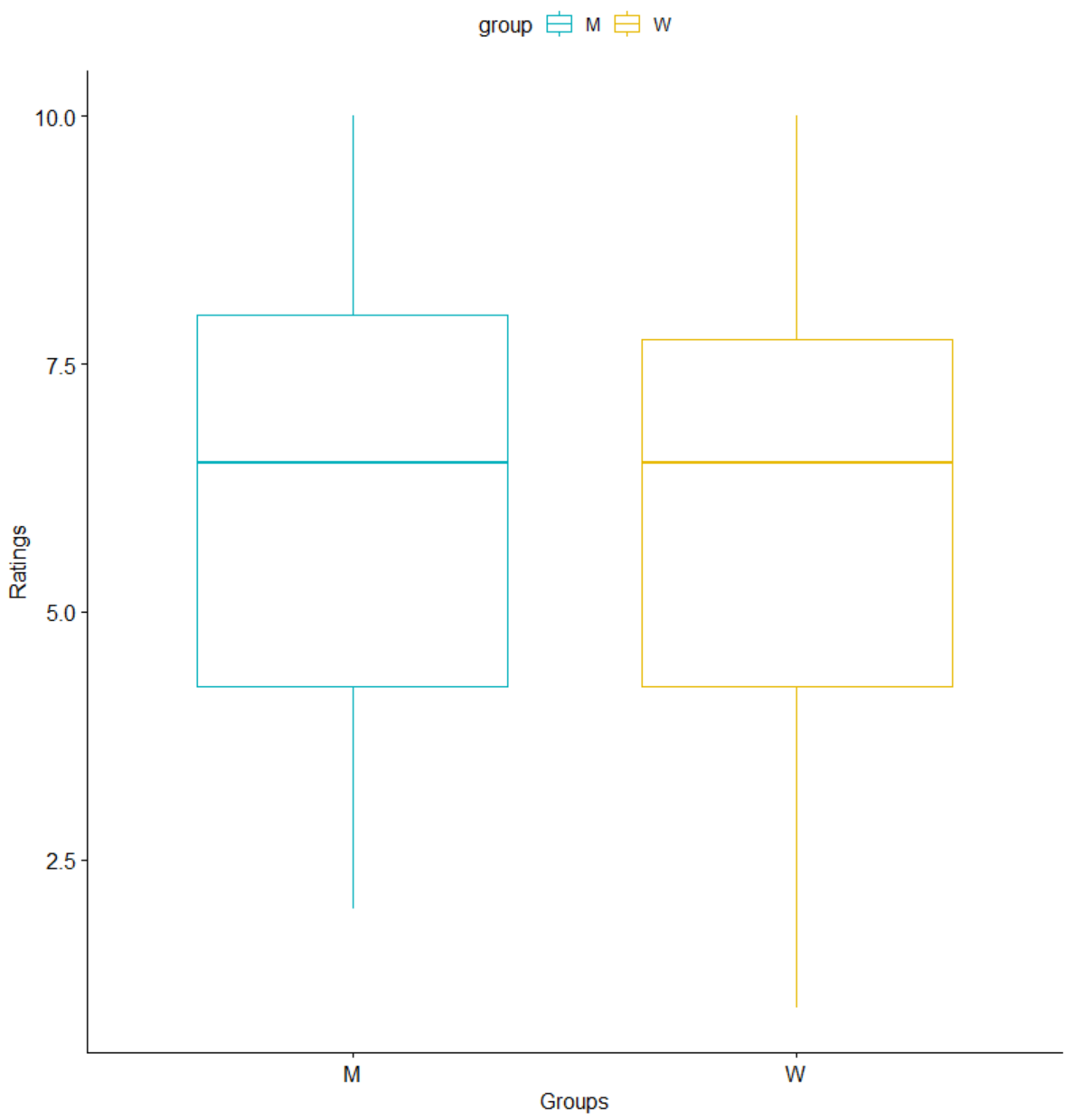
### Data Exploration

I will now do a data exploration, having the two sets in their own data frame is convenient because now I can do data exploration with them together instead of separately.



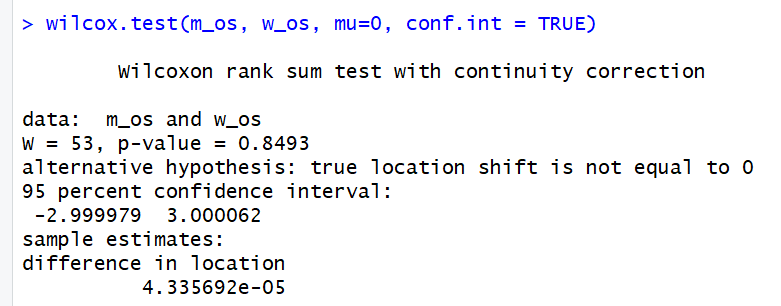
From our summary statistics we can see that our median is at 6.5, and when we group by and look at the median and interquartile range of the data we see that both M and W have a median of 6.5 and an interquartile range difference of only 0.25. Now I will plot the data based and compare the two distribution systems with a boxplot.





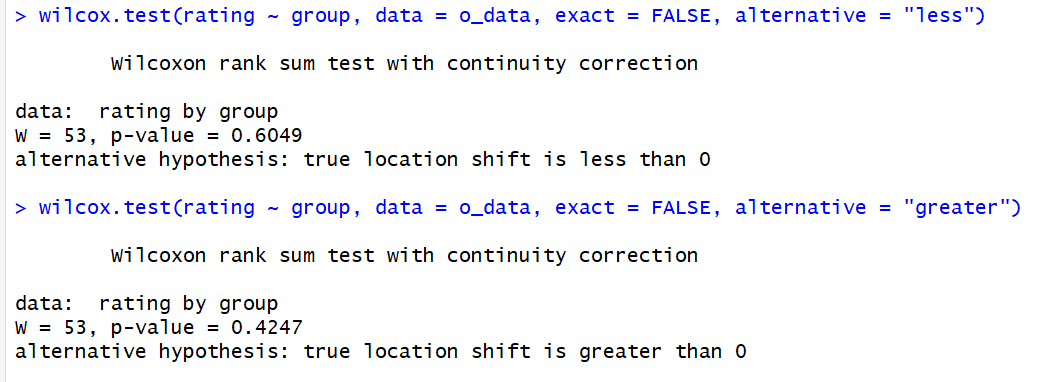
### Wilcoxon Sign Test

Now we run the Wilcoxon Sign Test will be run on the two distribution systems.



From our Wilcoxon Sign Test we can determine that the null hypothesis should not be rejected, that is because the p-value is greater than the significance level of 0.05 which when greater than 0.05 we would not want to reject the null hypothesis. Thus, we can also state the distribution of the data is indeed the same.

We can also test the data here too by either less than or greater than alternative hypotheses.



From testing each side of the distribution, less than or greater than, it backs up our previous notion that the distributions are the same. Both p-values are greater than 0.05 and thus we can determine that our distribution for the M and W operating systems are the same.

## Summary

From this assignment I learned about non-parametric testing in RStudio and got to test it out in a one-tail and two-tail scenario. The test I used to test my non-parametric data was the Wilcoxon Sign Test and found it simple yet effective when trying to determine where my distributions are related to the median. I also felt like I learned a lot about the topic of non-parametric testing and how it can be useful for researchers when their data is deemed not normal and a lot of tests become useless because of that.