Exam II

Frank Woodling

STAT 3480

University of Virginia

4/12/2016

There are benefits of hiring a firm specializing in nonparametric statistics. Nonparametric methods require minimal assumptions about the distribution of the population. (Higgins p7) With a traditional t-test a few assumptions must be made. The data must be made up of independent observations, it must be normally distributed, and the variances of the populations must be the same. We can safely assume the first, but we must run more tests in order to check the distribution and variance.

You have requested our services in examining two problems. The first task is to analyze the differences between states regarding test scores. The second task is to determine if the *We Come to You* instructors time spent ultimately makes a difference on test scores. We will begin with a broad overview of the data analyzing summary statistics and a few initial plots.

We will first look at the summary statistics compared the “On-site” and “We Come to You” dataset. It is not difficult to see that the out of county testing scores are worse. Looking at the summary statistics for the state testing data it seems like most are roughly the same, except for the test scores for state 4. State 4 test results appear to be nearly 10 lower than other states. It is worth noting that there are many observations missing from the data set, so we do not have population means.

We will next check for normality using histograms and Q-Q plots in figures 3 and 4. None of the Q-Q plots seem normal. The tails of each of them curve away from the Q-Q line. The state test data definitely does not seem normal with many outliers in the Q-Q plot. The histograms do not seem normal either. The state data is heavily left-skewed, and the other two data sets seem more uniform than normal. All three datasets are good examples of where nonparametric testing is beneficial, or perhaps even required.

The boxplot are a visual representation of the same summary statistics. The boxplots of the onsite and “We Come To You” testing data are much clearer than the table of statistics. The on-site testing is much higher on average. In the state testing data we can see that one box is much different from the others. The internal testing data is all between 81 and 83, so it is fairly close.

|  |  |  |  |
| --- | --- | --- | --- |
|  | On-site | WCTY | Overall |
| Min. | 84.05 | 79.20 | 79.20 |
| 1st Quart. | 84.49 | 80.10 | 81.46 |
| Median | 85.43 | 81.36 | 84.08 |
| Mean | 85.42 | 80.79 | 83.44 |
| 3rd Quart. | 86.35 | 81.47 | 85.48 |
| Max. | 86.94 | 81.65 | 86.94 |
| *Table 1. Table of summary statistics for on-site and “We Come To You” testing data* | | | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Our State | State 1 | State 2 | State 3 | State 4 | State 5 | Overall |
| Min. | 83.08 | 83.16 | 83.37 | 83.48 | 74.62 | 83.41 | 74.62 |
| 1st Quart. | 83.90 | 83.52 | 84.72 | 83.79 | 75.07 | 83.71 | 83.38 |
| Median | 84.70 | 84.81 | 84.88 | 84.79 | 76.06 | 84.33 | 84.07 |
| Mean | 84.78 | 84.69 | 85.09 | 84.90 | 75.99 | 84.61 | 83.20 |
| 3rd Quart. | 86.00 | 85.69 | 86.09 | 85.91 | 76.35 | 85.18 | 85.61 |
| Max. | 86.24 | 86.91 | 86.72 | 86.99 | 78.46 | 86.88 | 86.99 |
| *Table 2. Table of summary statistics for state testing data* | | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | < 30 min | 60 min | 90 min | 120 min | 180 min | Combined |
| Min. | 81.00 | 80.05 | 80.24 | 80.15 | 80.37 | 80.05 |
| 1st Quart. | 82.34 | 80.79 | 81.58 | 80.80 | 81.80 | 81.15 |
| Median | 82.93 | 81.45 | 82.19 | 81.00 | 82.51 | 82.15 |
| Mean | 82.82 | 81.76 | 82.16 | 81.50 | 82.32 | 82.11 |
| 3rd Quart. | 83.59 | 82.92 | 82.77 | 82.38 | 83.03 | 83.06 |
| Max. | 83.98 | 83.84 | 83.89 | 83.29 | 83.71 | 83.98 |
| *Table 3. Table of summary statistics for state testing data* | | | | | | |

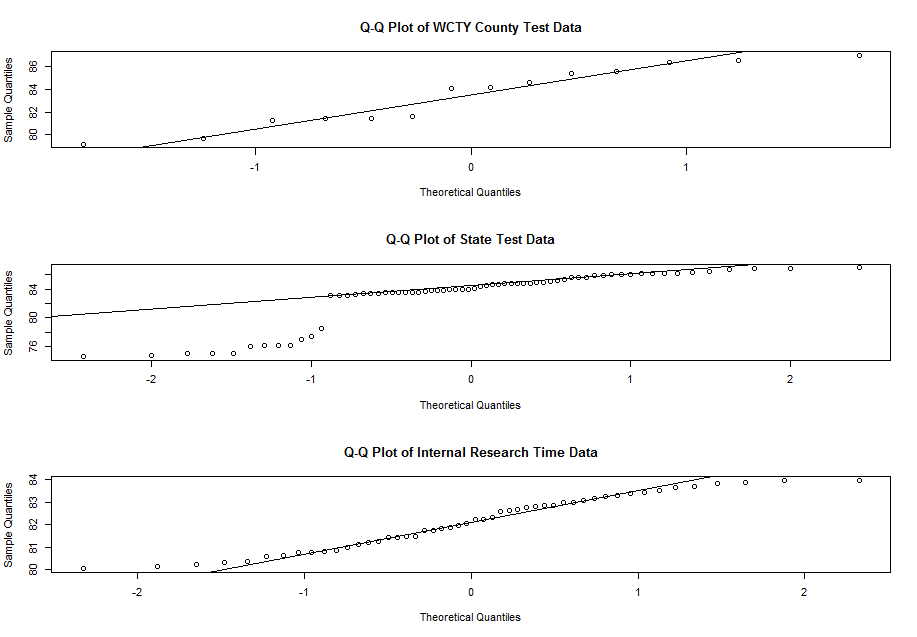


Figure 3: Q-Q plots of datasets

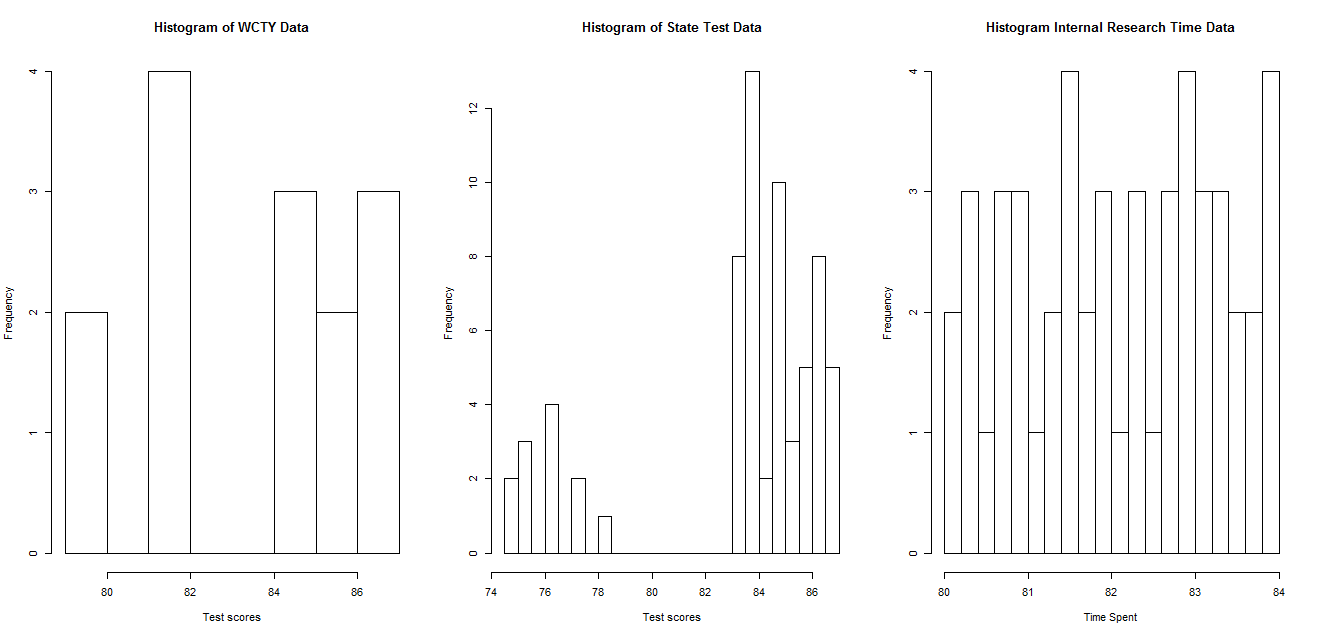


Figure 2: Histogram of datasets

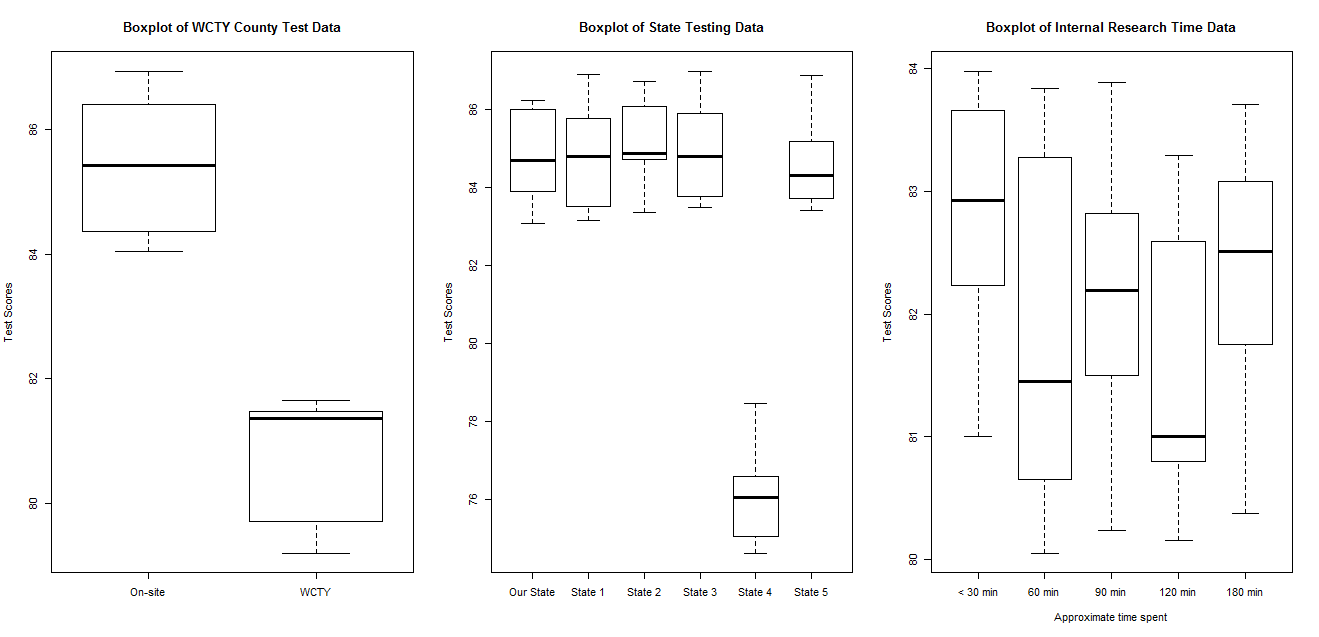


Figure 3: Boxplots of datasets

The first tests we must run are to compare the scale and variance of the distributions. This must be done since the other tests have certain assumptions that must be met in order to be able to use them. Since we do not know the means we cannot use the Siegel-Tukey or Ansari-Bradley tests to check for variance. In this case we will use an RMD test in order to determine whether the groups have equal variances. As mentioned in the introduction nonparametric is useful when we are not analyzing a normal distribution. Since we have gathered that neither distribution is close to normal we know that a nonparametric test is required to achieve an accurate result.

After checking variances we can begin to run k-sample permutation tests. The Kruskal-Wallis permutation test will be used. First we will rank the data from the smallest to largest scores. Using our original data we can find a test statistic and average rank for each different treatment. The data will then be sampled to try many possible combinations performing the same test. At the end we can compare each permutation to our original test statistic and find a p-value. The null hypothesis for this test is that each treatment is the same, while the alternative is that at least one treatment is different. We must also use an adjustment since there are multiple samples in order to account for type 1 error. The adjustment to be used is the Bonferroni adjustment, which is one of the more conservative adjustments.

|  |  |
| --- | --- |
| Dataset | P-value |
| County |  |
| Deviances | 0.31 |
| Wilcoxon | 0.000333 |
|  |  |
| States |  |
| Deviances | 0.23 |
| Kruskal-Wallis | 0.00001573 |
| Pairwise Differences | Table below |
|  |  |
| Internal |  |
| Deviances | 0.21 |
| Kruskal-Wallis | 0.1092 |
| *Table 4: Test results* | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Our State | State 1 | State 2 | State 3 | State 4 | State 5 |
| Our State | - | 0.9 | 0.76 | 1 | 0 | 0.81 |
| State 1 |  | - | 0.67 | 1 | 0 | 0.62 |
| State 2 |  |  | - | 0.78 | 0 | 0.28 |
| State 3 |  |  |  | - | 0 | 0.69 |
| State 4 |  |  |  |  | - | 0 |
| State 5 |  |  |  |  |  | - |

Figure 4. Pairwise comparisons of state data

We ran a Kruskal-Wallis permutation test on each data set. The p-value for the state test data was 0.00001573. Using the Bonferroni adjustment we would need a p-value of 0.003333333 in order to reject the null hypothesis. For the internal data we found a p-value of 0.1092, while we would need a p-value of 0.005 in order to reject the null hypothesis.

To interpret the results from these tests we decide whether or not we were able to reject the null hypothesis (H0 =all treatments are equal). For the county data (on-site vs “We Come to You” testing) we are able to reject the null hypothesis. We can conclude that the two test methods are not equal.

For the state data we rejected the null hypothesis since the p-value was less than 0.00333333. In this case we can conclude that at least one of the treatments is not equal to the other treatments. The test does not tell us which one, however, from the previous graphs it would appear state 4 did not perform as well as the others. To confirm this we can perform pairwise differences tests. In figure 4 we can see the results for this statistic. In this case we must use a Bonferroni adjustment since we are running multiple tests at once. We would need a p-value of 0.05/10 or 0.005 in order to reject the null hypothesis. We can see that all the state 4 tests are rejected. We can conclude that state 4 is different from all other states.

For the internal data we are unable to reject the null hypothesis. We found a p-value of 0.1092. We can conclude that the treatments are equal to each other within a significance level of 0.05.

In conclusion we can say that your assumption is correct in that there is a difference between states. Judging from the summary statistics and the permutation test results we can say that at least one other state is not performing the same as the others. Your state seems to be performing as well as the other 4 states. Regarding possible discrepancies in the *We Come to You* training program we can’t say that there is a difference between the groups. It does not appear to make a difference how much time the instructors spent with clients; all test results are similar.

Code is attached.

Software: R, RStudio

Works Cited:

Higgins, James J. *Introduction to Modern Nonparametric Statistics.* Belmont: Brooks/Cole Cengage Learning. 2004.