|  |  |
| --- | --- |
|  | 🞂Analysis of spending  Comparison between female and male candidates |
|  |  |
|  | **Ye In Jeon** 🞂STAT 2183W 🞂 2015-10-01 |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# Introduction

Candidates who are running public offices report their spending during their campaign. It is found through political scientists that female candidates tend to spend less amount of money than male candidates. Generally, it is because female candidates are hard to raise funds than male candidates. To confirm if female candidates spend less money in their campaigns for public office than male candidates, the data of campaign expenditures is collected.

## Analysis and Discussion

Before starting the analysis, the null hypothesis has to be set. It is that there is no difference between female and male candidates’ spending. Also, the alternative hypothesis has to be determined as the spending of each gender candidates are different.

This can be rewritten with statistic terms as followings:

H0: Difference=0 Ha: Difference0

T -test is used to test if there is a difference between two genders’ expenditure. Pooled t-test has to be used because both groups of variables are independent. Also, the equality of two groups’ variances has to be checked. If the two variances are equal, pooled method has to be used. If not, Satterthwaite method has to be used.

Before t-test, the summary statistics give information about two groups.

Table 1. All variable summary statistics

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Analysis Variable : spend** | | | | | | | | |
| **N Obs** | **N** | **Mean** | **Median** | **Std Dev** | **Lower 95% CL for Mean** | **Upper 95% CL for Mean** | **Minimum** | **Maximum** |
| 40 | 40 | 298.15 | 296.00 | 77.77 | 273.28 | 323.02 | 126.00 | 458.00 |

Through Table 1, the mean, median, standard deviation, confidence interval, minimum and maximum values of all candidates can be found.

Table 2. Summary statistics by gender

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Analysis Variable : spend** | | | | | | | | | |
| **NAME OF FORMER VARIABLE** | **N Obs** | **N** | **Mean** | **Median** | **Std Dev** | **Lower 95% CL for Mean** | **Upper 95% CL for Mean** | **Minimum** | **Maximum** |
| Female | 20 | 20 | 245.30 | 254.00 | 51.95 | 220.98 | 269.62 | 126.00 | 318.00 |
| Male | 20 | 20 | 351.00 | 355.50 | 61.92 | 322.02 | 379.98 | 209.00 | 458.00 |

In Table 2, it can be seen that female group have lower value of mean, median, confidence interval, minimum, and maximum than male group.

Table 3. Summary statistics by gender and difference

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **gender** | **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| **Female** | 20 | 245.3 | 51.9545 | 11.6174 | 126.0 | 318.0 |
| **Male** | 20 | 351.0 | 61.9210 | 13.8460 | 209.0 | 458.0 |
| **Diff (1-2)** |  | -105.7 | 57.1554 | 18.0741 |  |  |

The difference noted in the Table 3 cannot be used directly to the t-test because the variance equal are not checked.

Table 4. Summary statistics of Pooled and Satterthwaite

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **gender** | **Method** | **Mean** | **95% CL Mean** | | **Std Dev** | **95% CL Std Dev** | |
| **Female** |  | 245.3 | 221.0 | 269.6 | 51.9545 | 39.5109 | 75.8833 |
| **Male** |  | 351.0 | 322.0 | 380.0 | 61.9210 | 47.0903 | 90.4401 |
| **Diff (1-2)** | **Pooled** | -105.7 | -142.3 | -69.1108 | 57.1554 | 46.7100 | 73.6607 |
| **Diff (1-2)** | **Satterthwaite** | -105.7 | -142.3 | -69.0745 |  |  |  |

As in Table 4, two different methods, Pooled and Satterthwaite, are used instead of subtraction between female groups and male groups’ spending.

Table 5. F test table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Equality of Variances** | | | | |
| **Method** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **Folded F** | 19 | 19 | 1.42 | 0.4514 |

With Table 5, the equality of variances which are tested through Folded F-test can be seen. P-value, 0.4514, is bigger than the value of alpha, 0.05. Therefore, it can be said that variances are equal. So, the Pooled method section of Table 6 is appropriate for the analysis.

Table 6. Variance Equality

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Method** | **Variances** | **DF** | **t Value** | **Pr > |t|** |
| **Pooled** | Equal | 38 | -5.85 | <.0001 |
| **Satterthwaite** | Unequal | 36.887 | -5.85 | <.0001 |

In the Table 6, it is shown that Pooled method’s p-value, 0.001, is smaller than the value of alpha, 0.05. Therefore, it is significant that the null hypothesis, there is no difference between two groups, is rejected.

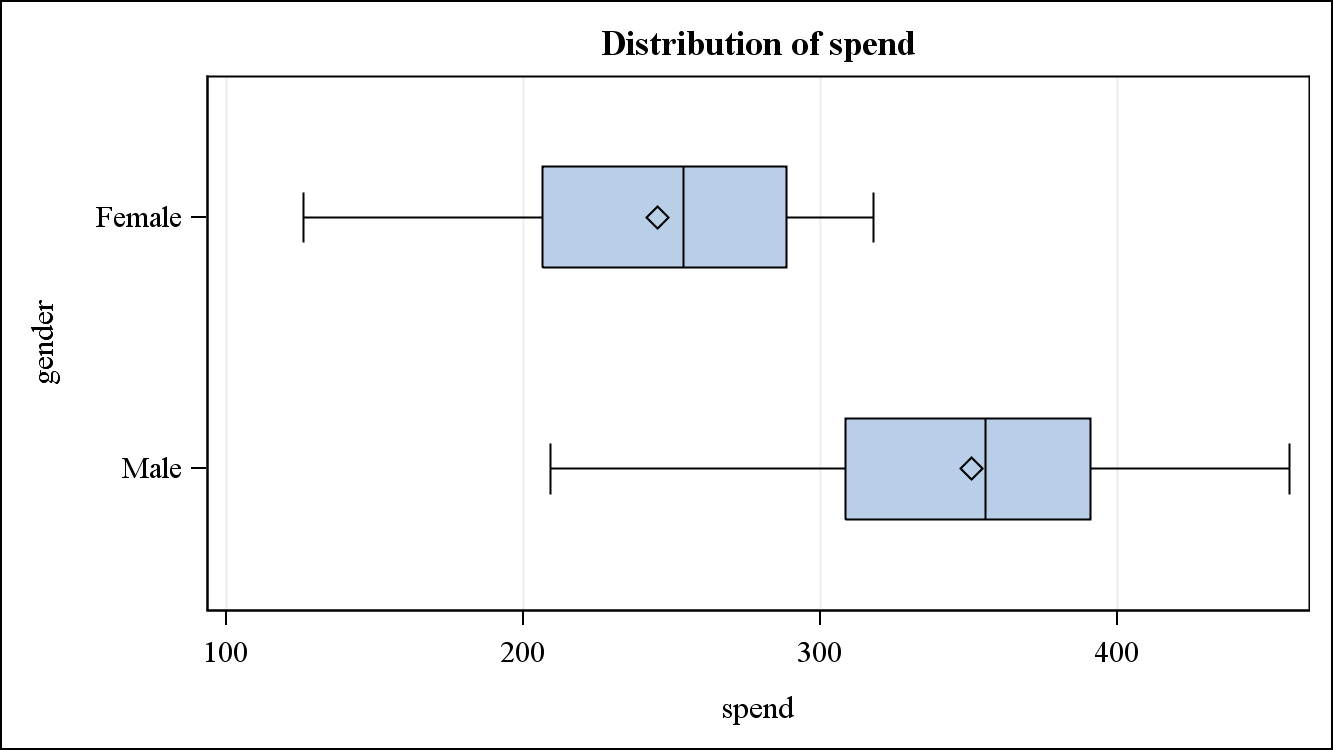


Figure 1. Box plot

Figure 1 shows box plots which give distributions of female and male candidates’ expenditures. It can be seen that the mean of male group is smaller than female group. Also, female group’s outliers have smaller value than those of male group.

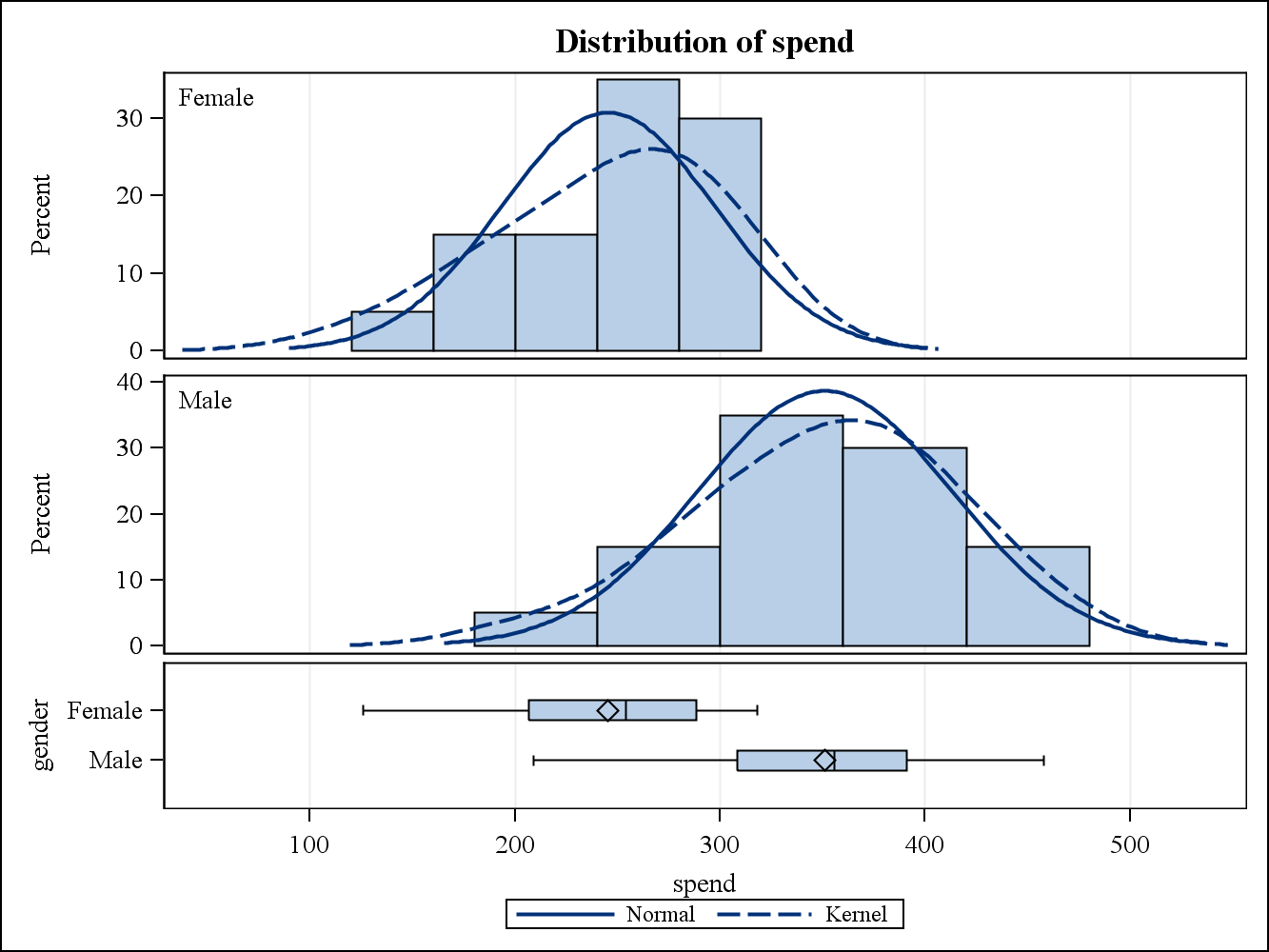


Figure 2. T-test graphs

T-test graphs are shown in Figure 2. Like Figure 1, these graphs show that female candidates spend less than male candidates. Also, the lines are similar to normal distribution graph. Therefore, it can be assumed that they are normally distributed.

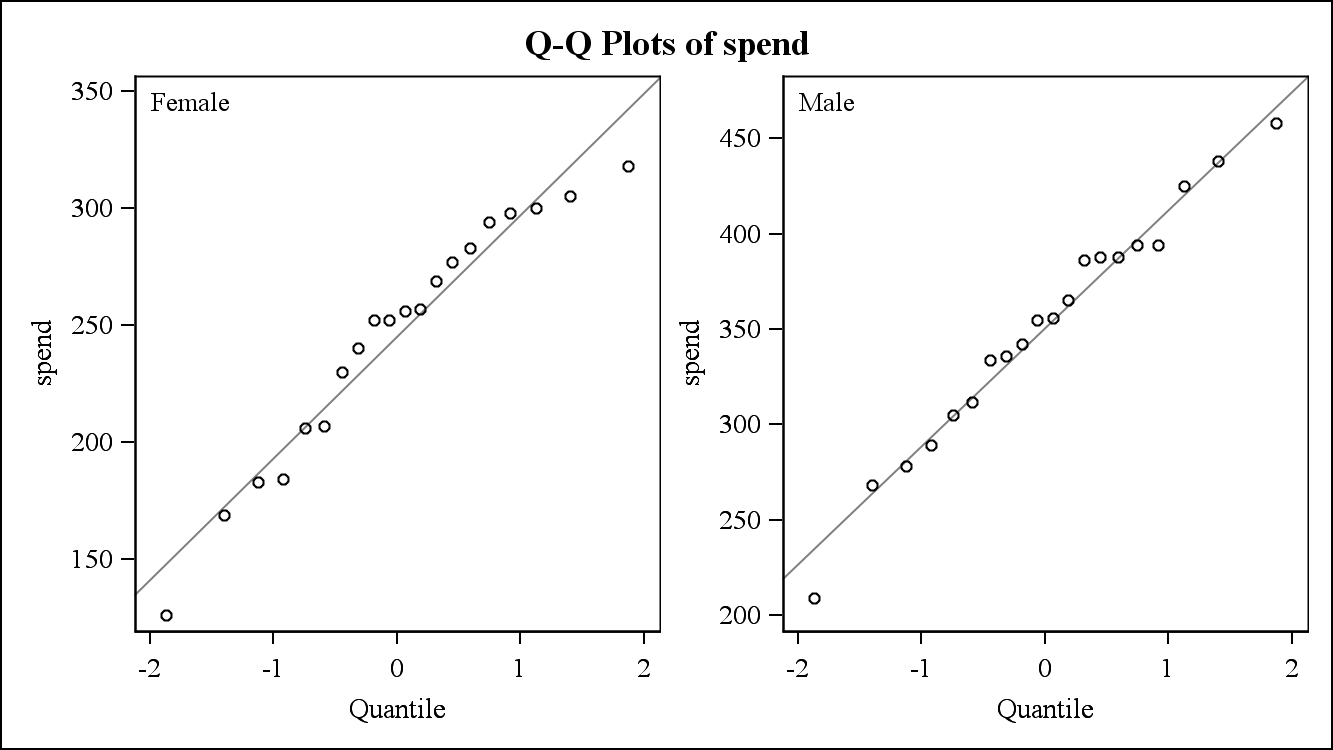


Figure 3. Q-Q plots

Figure 3 gives Q-Q plots of each gender. Since variables are close to the line, it can be assumed that two groups have normality.

### Conclusion

The purpose of this analysis is to check if female candidates spend less than male candidates during their campaign in their public offices. The data of expenditures is collected from two groups, which are separated by gender. With this data, it is statistically proved that there is a difference between male and female candidates’ spending. By analyzing this test result with the summary of data distribution, it can be said that female candidates tend to spend less money than male candidates. As mentioned, the one of the reason that female candidates spend the relatively small amount of money is difficulties in raising money. Therefore, it would be recommended to find a way to give female candidates more chances to raise their money which they can spend for their public offices. Helping female candidates to hire skilled professionals in fundraising strategy can be one example. For further studies, it will be helpful to gain data of candidates’ regions. It is because candidates with low spending can be densely populated in specific states.

#### Appendix

ods rtf file="Project1.rtf";

**proc** **means** data=prob6\_55 mean median std CLM N Min Max;

var male female;

**run**;

**data** temp (keep=Female Male);

set prob6\_55;

**run**;

**Proc** **transpose** data =temp out=tprob6\_55 (rename = (\_Name\_=gender));

**run**;

**proc** **sort** data=tprob6\_55;

by gender;

**run**;

**proc** **transpose** data=tprob6\_55 out=finalprob (rename=(col1=spend) drop=\_name\_);

by gender;

**run**;

**proc** **means** data=finalprob N mean median std clm min max PRINTALLTYPES MAXDEC=**2**;

class gender;

var spend;

**run**;

ods select boxplot qqplot;

**proc** **ttest** data=finalprob sides=**2** alpha=**0.05** plot=(box qq);

class gender;

var spend;

**run**;

ods rtf close;