ISyE 2028 – Basic Statistical Methods - Fall 2015
Bonus Project: "Big" Data Analytics
Final Report: Time spent on social media

Abstract:

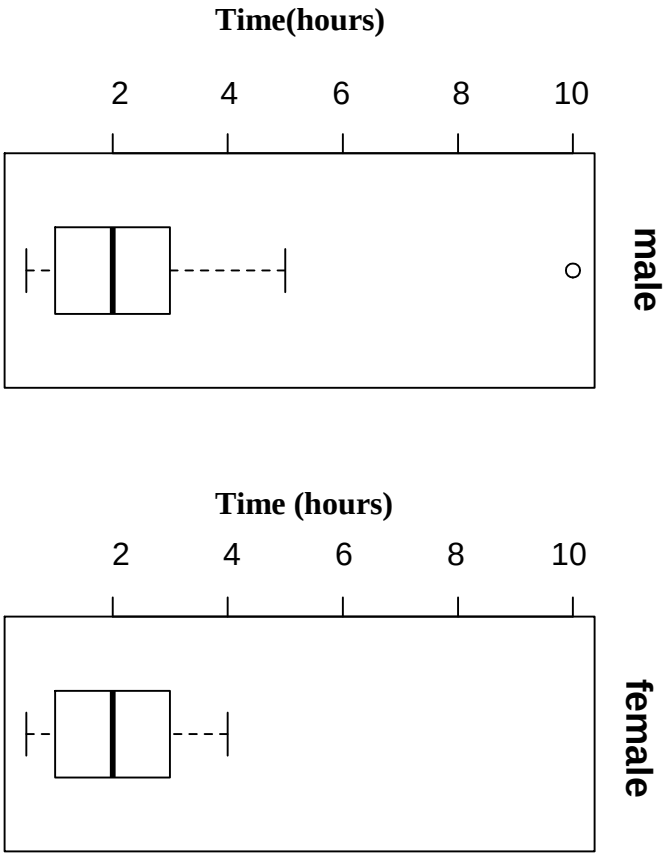
The growth of social media is astounding and part of that success was because everyone is participating especially college students. Most Georgia Tech held events such as networking, tailgating, sports games and much more are constantly updated on social media. Social media usage is a vital evidence of how important social media is. I am particularly interested on time spent on social media for Georgia tech students. I want to investigate on whether there is an expected difference on time spent on social media for female and male population. Among my friends, I have heard a lot of them told me that female students spent more time on social media compared to male students. I am looking to find the truth by using hypothesis testing analysis to see whether female students spend more time on social media compared to male students.

Method used:

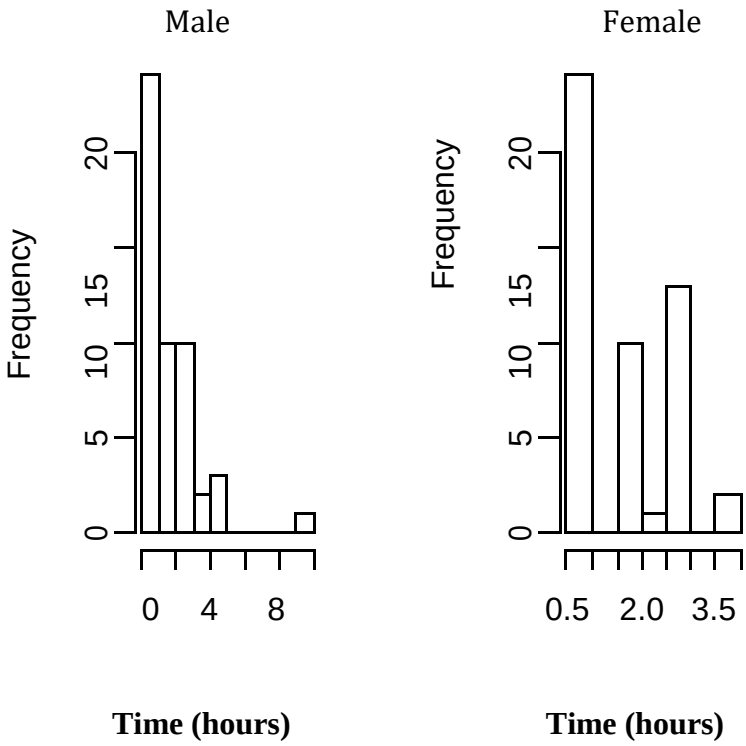
- Collection of raw data by using the website "SurveyMonkey" and ask students on campus personally for their input. (100 samples of male students and female students were collected). Question asked on survey: Time spent on social media and Male/Female.
- Box plots and histograms to show distributions.
- 5 numerical summaries analysis using R software.
- 95 % confidence interval for each population (male students, female students) and both as whole population using R software.
- Hypothesis testing using R software :
    - H0 : μ (female) - μ(male) = 0
    - H1: μ (female) > μ (male)

Results:

Box Plot



Histogram



5 numerical summaries:

- Male students:
  - Minimum = 0.50, 1st Quartile = 1.00, Median = 2.00, Mean = 2.04, 3rd Quartile = 3.00, Max = 10.00
- Female students:
  - Minimum = 0.50, 1st Quartile = 1.00, Median = 2.00, Mean = 1.77, 3rd Quartile = 3.00, Max = 4.00

95 % Confidence Interval:

- Male Students:
  - μ = [1.550409, 2.529591]
- Female students:

- μ = [1.468618, 2.529591]

Expected mean difference on time spent on social media:

μ(male) - μ(female) = 2.04 – 1.77 = 0.27 hours

Welch two sample t-test:

data: femaleSocial and maleSocial (refer to appendix for more information)
t = -0.94376, df = 81.473, p-value = 0.826
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
-0.745987 Inf
sample estimates:
mean of x mean of y
1.77 2.04

Wilcoxon rank test for two population:

data: femaleSocial and maleSocial (refer to appendix for more information)
W = 1209, p-value = 0.6154
alternative hypothesis: true location shift is greater than 0

These are the results analyzed using methods described from “Methods Used” section and used to analyze expected difference on time spent on social media and hypothesis testing.

Refer to appendices for concrete data.

Conclusion:

With analysis that I have conducted, there is no evidence that female students spend more time on social media compared to their male counterparts. My hypothesis clearly does not match my findings. I used Welch two sample t-test and Wilcoxon rank test to test my hypothesis. It turned out the p-value was quite large therefore fail to reject null analysis (accept null analysis and reject alternative hypothesis). If we look at both populations separately, we can see from 95% confidence interval that I constructed, on average male students spent around 1.55 hours to 2.52 hours compared to female students that spent around 1.46 hours to 2.07 hours. The expected mean difference on time spent on social is 0.27 hours. Overall, additional research needs to be conducted in order to answer questions posed from the hypothesis accurately. Possible error could come from small sample (100 samples per population). For better statistical analysis, I would collect more data and do more detailed analysis and use more advanced statistical methods.

Appendices:

- Raw data from the survey can be found on this link:
    - Raw Data
    - R programming source
    - R programming source 1
    - Note: maleSocial is from samples of male students, femaleSocial is from samples of female students.