# ISyE 2028 – Basic Statistical Methods - Fall 2015
# Bonus Project: "Big" Data Analytics
# Proposal (or Final Report)

Rebeccah Sharpe

ISYE 2028

Professor Yao Xie

Bonus Project

November 29, 2015

Women in engineering is not a term most often heard until recent years. In fact, only recently has there been a movement to encourage girls to follow an engineering or STEM path in life. As a female engineering student, I find that I am surrounded by men in classes, around campus, and in the workplace. I find that women think and approach ideas and problems in a different way than men usually do. If more women join the STEM professions, we may as a human kind be able to excel ideas that could cure disease or develop new technologies to change the development of humans forever. The main problem is to examine colleges around the United States and compare the amount of men and women enrolled in engineering programs and compare the most popular degrees received.

The Data was collected from a website that had up to 378 colleges report their enrollment statistics and was organized into main lists included how many women in each college received degrees, how many total students enrolled in engineering programs, and how many students enrolled in each engineering program. This layout allows for multiple perspectives in evaluating the data presented.

The first test on the data I performed was a hypothesis test that tested whether an acceptable estimate for the amount of women in engineering programs in 2011 was 10%. The sample proportion I used was the percentage of women from all colleges reported across the United States. This sample proportion is directly from the data. The result of the test would conclude whether or not the amount of women in engineering programs across the United States would be either 10% or less than 10%. After performing the hypothesis test, I accepted the null hypothesis that 10% is a good estimation of the amount of women in engineering in 2011. I also found a confidence interval for this test which gave the same conclusion of accepting the null hypothesis. This test gives a broad outlook on how many women are actually in engineering, which is very few. The hypothesis test may be found on page 5.

The next test on the data I performed was a hypothesis test on the percentage of women represented in the top ten engineering schools ranked by the amount of enrolled students. I estimated the amount of women in

these schools to be 10% or less than 10%. After performing the test, I found that I must reject the null hypothesis and accept that the estimation must be less than 10%. Furthermore, I also found a confidence interval to back up my original finding. The confidence interval also resulted in a rejection of the null hypothesis. By looking at the most popular engineering schools in the country, we can gain a better understanding of the representation of women in engineering because the populations of students enrolled are larger, so intuitively there should be more women enrolled in those programs. By analyzing that assumption, there were more women at those top ten schools than the other schools included in the data set. The hypothesis test for this data can be found on page 6.

Continually, I formed two box plots for the women and men enrolled in the top ten schools for engineering. Initially, I combined the two plots for a better comparison; however, the scale for the women was too small to evaluate any idea from the plot. Both data sets are evenly spread for the amount of students in each category. However, looking at the scales of each plot there is an obvious uneven distribution of men and women. The men amounts are in the thousands whereas the women are in the hundreds. This is a result I foresaw just because of the initial observation I made from my own experiences in classes and throughout my current college experience. These plots can be found on page 7.

To evaluate the data further, I collected the data given for just the Georgia Institute of Technology. I looked at the distribution of engineering majors within the school as well as compared the degrees awarded to Ga Tech students to the degrees awarded to students across the United States. I found the most popular programs are ISYE and Mechanical. These two programs also have the most amount of women enrolled. Also, comparing the two populations of students, there was no correlation between the two seen by a line of best fin on the graph, found on page 8, and also shown by the $R^2$ value being .04 which is not close to a value of one. This evaluation does not give much insight into the amount of women involved in the two populations. This

could be from a sample size or a need for more in depth statistics for the amount of women enrolled in every program offered at each school.

Lastly, I further looked into the top ten schools to see any other trends involving women. I found the top programs of students enrolled in each type of engineering program. I found that the two programs that had the most enrolled students at each college were mechanical and civil. I then found the average amount of women in the engineering schools was about 3%.I also found the percentage difference from the average women enrollment across the United States. This data can be found on page 9. I further analyzed the differences by using the first graph on page 10 which allows for a pictorial view of how far each school is from the average. Each school has about the same percentage of women; however, Ga Tech is the furthest away from the average meaning that school has the most women enrolled out of the other top ten school in relation to the total amount of engineering students. The second graph on page 10 evaluates whether the differences are linearly related to the average across the United States. The data perfectly fits a line and the $R^2$ value is equal to one meaning this relationship is relevant to the goal of evaluating women enrolled in engineering.

Origonaly I hypothesized that biomedical or chemical engineering would have the most amount of women involved. However, after the tests performed I found that about 3% of enginerers are women and that the most popular program is mechanical followed by civil engineering. By analyzing this data, colleges will be inspired to further analyze why women are not joining STEM professions and will be able to find ways of promoting these important and ever changing professions.

Hypothesis Test for Percentage of Women in Engineering across USA (2011)

1. Parameter of interest: $\hat{p}$

2. Test Hypothesis:

   $H_o$: P= .1
   $H_1$: P< .1

3. Test Statistic

$$Z_o = \frac{p_o - \hat{p}}{\sqrt{\frac{p_o(1-p_o)}{n}}} = \frac{.21-.1}{\sqrt{\frac{.1(.9)}{1868}}} = 15.847$$

$\hat{p}$= 387/1868

4. Rejection Region

   α=.05
   $Z_ə$=1.65

   Reject $H_o$ if $|Z_o| < -Z_ə$
   15.847<-1.65

5. We Accept $H_o$, so the Total amount of women in engineering that received degrees in 2011 was about 10%.

6. Confidence Interval:

$$\text{CI: } (-\infty, \hat{p} + Z_ə\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ ] = (-\infty, .21 + 1.65\sqrt{\frac{.21(.79)}{1868}}\ ] = (-\infty, .266)$$

   This further concludes that the hypothesis was correct within a reasonable estimate.

Hypothesis Test for percentage of Women in Engineering from Top Ten School (2011)

1. Parameter of interest: $\hat{p}$

2. Test Hypothesis:

   $H_o$: P=.1
   $H_1$: P<.1

3. Test Statistic

   $Z_o = \frac{p_o - \hat{p}}{\sqrt{\frac{p_o(1-p_o)}{n}}} = \frac{.03-.1}{\sqrt{\frac{.1(.9)}{70769}}} = -62.07$

   $\hat{p}$= 2154/70769

4. Rejection Region
   α=.05
   $Z_ə$=1.65

   Reject $H_o$ if $|Z_o| < -Z_ə$
   -62.07<-1.65

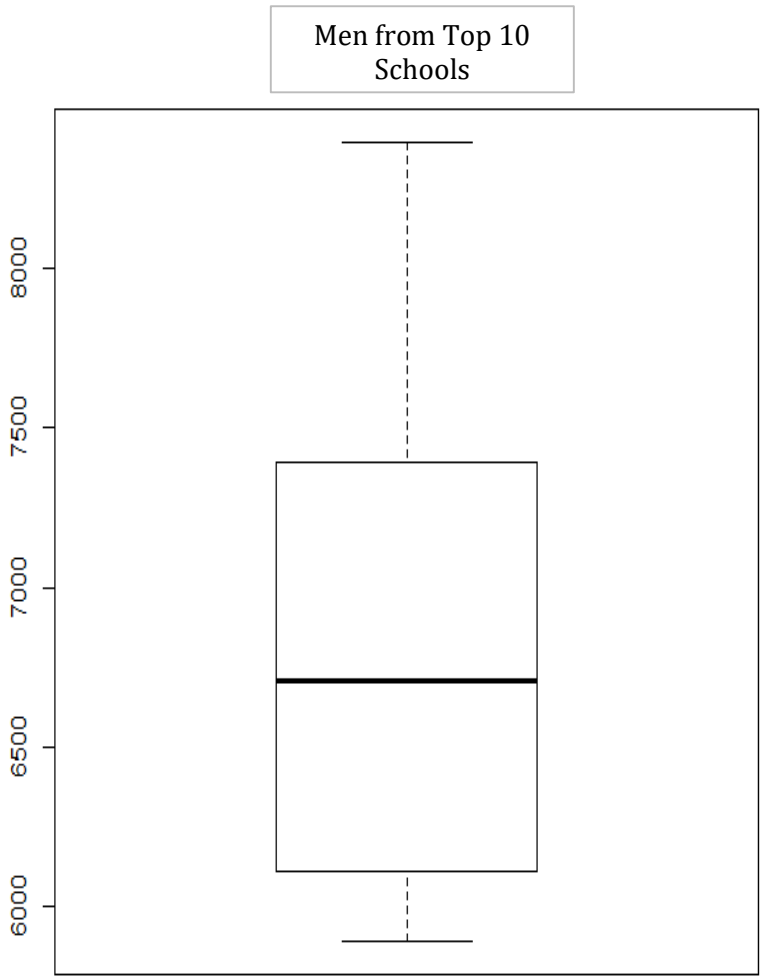
5. We reject $H_o$, so the total percentage of women receiving engineering degrees form the top ten engineering schools in 2011 is less than 10%.

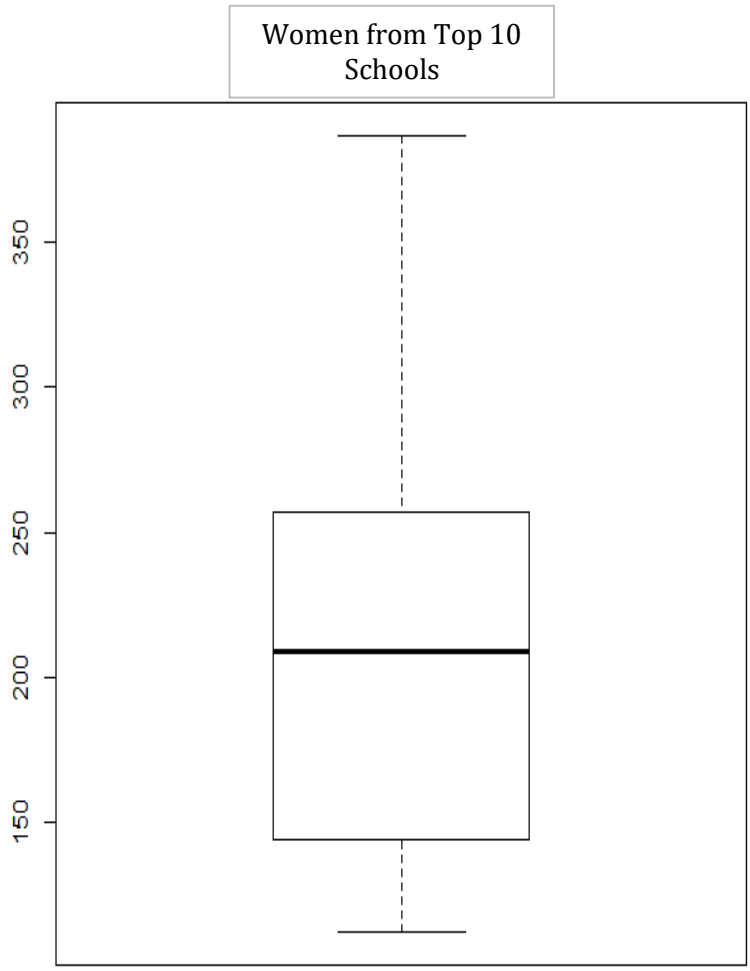
6. Confidence Interval:

   CI: $(-\infty, \hat{p} + Z_ə\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\ ] = (-\infty, .03 + 1.65\sqrt{\frac{.03(.97)}{70769}}\ ] = (-\infty, .031)$

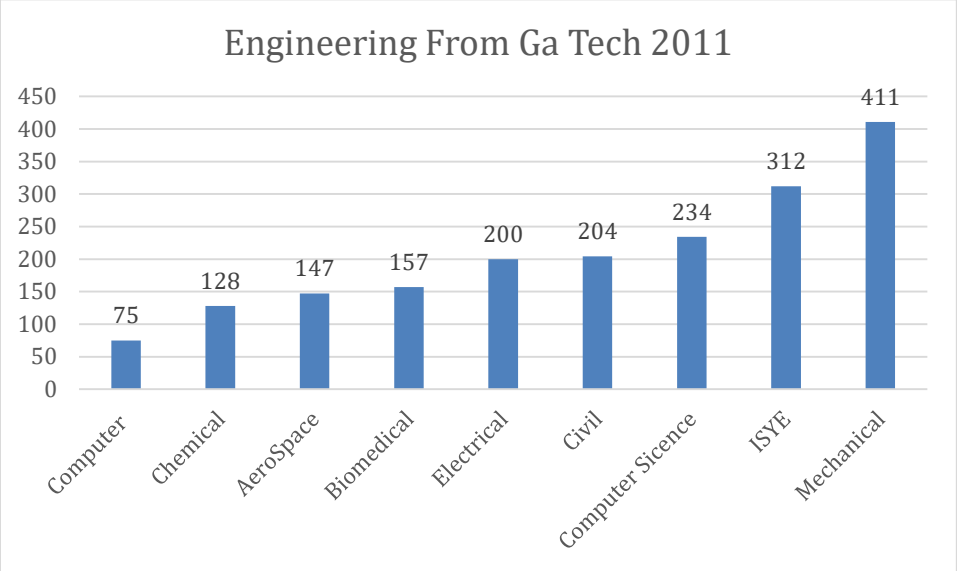
   $P_o$ is not within this confidence interval, so we our initial conclusion of rejecting the null hypothesis is correct.

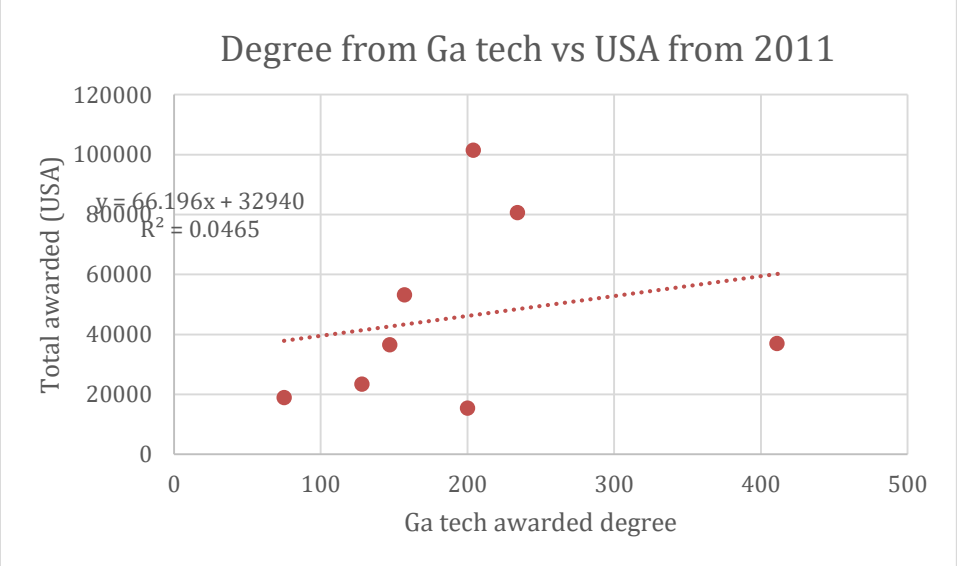
# Men Vs Women Box Plot



# Georgia Tech Engineering Spread (Men and Women)



# Top Ten Schools

| enrollment | total enrolled | women | men | top program | Total in Program |
|---|---|---|---|---|---|
| **Top ten schools** | | | | | |
| Penn State | 8593 | 195 | 8398 | ME | 233 |
| Ga tech | 8297 | 387 | 7910 | ME | 411 |
| Texas A&M | 7628 | 235 | 7393 | Civil | 263 |
| Perdue | 7321 | 270 | 7051 | ME | 264 |
| Univ Illinois | 7181 | 257 | 6924 | ME | 206 |
| Ohio State | 6629 | 144 | 6485 | ME | 247 |
| Virginia tech | 6576 | 216 | 6360 | ME | 296 |
| North Carolina State | 6311 | 202 | 6109 | Civil | 217 |
| Univ Central FL | 6208 | 112 | 6096 | Civil | 213 |
| Iowa State | 6025 | 136 | 5889 | ME | 224 |
| Total | 70769 | 2154 | 68615 | | 2574 |

| **Top ten schools** | percentage of females | Difference from average |
|---|---|---|
| Penn State | 0.02269289 | -0.77% |
| Ga tech | 0.046643365 | 1.62% |
| Texas A&M | 0.030807551 | 0.04% |
| Perdue | 0.036880208 | 0.64% |
| Univ Illinois | 0.035788887 | 0.54% |
| Ohio State | 0.021722733 | -0.87% |
| Virginia tech | 0.032846715 | 0.24% |
| North Carolina | 0.032007606 | 0.16% |
| Univ Central FL | 0.018041237 | -1.24% |
| Iowa State | 0.022572614 | -0.79% |
| Total | 0.030437056 | |

# Representation of Top Ten Engineering Schools Compared to Country