

---

ISyE 2028 – Basic Statistical Methods - Fall 2015  
Bonus Project: “Big” Data Analytics  
Final Report

---

Chrystelle Nare

STEM vs. non STEM majors academic performance at Georgia Tech

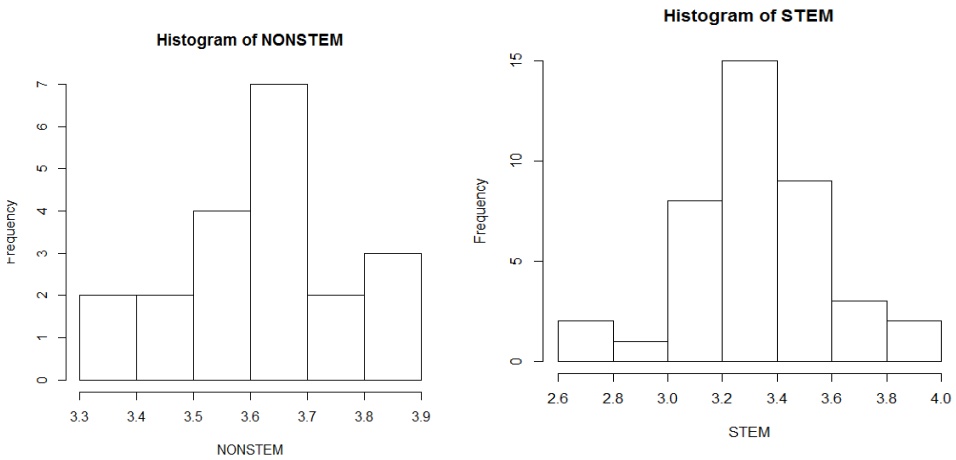
Throughout college, and after changing majors from non-STEM to STEM; I always wondered if non STEM majors really performed better in school; so I decided to do try to compare the average GPAs that these two categories of majors were getting. In order to do so, I used data from the Georgia Tech Institutional Research and Planning website, and looked for the grades awarded to each majors for the spring and summer semesters 2015.

The non STEM majors (Group 1) used include Business , Public Policy, modern Languages, Literature, Media & Communication , International Affairs; History , technology , and society ; Economics, Public policy, Industrial design , City Planning, and Music; the STEM majors (Group 2) used include : Architecture, Building Construction , Computer Science (College of Computing), Aerospace Engineering , Biomedical Engineering , Chemical & Biomolecular Engineering , Civil & Environmental Engineering, Electrical & Computer engineering , Industrial & Systems Engineering , Materials Science & Engineering, Mechanical Engineering , Polymer & Fiber Engineering , applied physiology , Biology, Chemical & Biomolecular engineering , Computer Science (college of Science) , Earth & Atmospheric Science , Mathematics, Physics , and Psychology.

The data given on the website is a collection of letter grades given to students from each majors during a specific term. I chose used A,B,C,D and ignored the S, U, I, W , and V because they're not usually taken into account when calculating the final GPA. In order to test and compare the data from both groups, I converted them into useful numbers. To calculate the average GPA of each majors, I used the GPA scale and assigned numbers to the letters obtained by students: A = 4 , B = 3 , C = 2 , D = 1 , F = 0 . For each majors, I multiplied the number of students that received an A by 4, the number of students that received a B by 3; a C by 2 , a D by 1 ,an F by 0, and added up all the numbers. I then divided this sum by the total number of students that received these grades to have a GPA for each majors.



After obtaining the GPAs for both group, I used R to obtain a graph to find out if the samples were normally distributed and I concluded that they were judging by the shapes or the graphs.



My first step in the comparison process was to use R to find the five number summary for both groups and construct box plots.

GROUP 1 (non STEM):

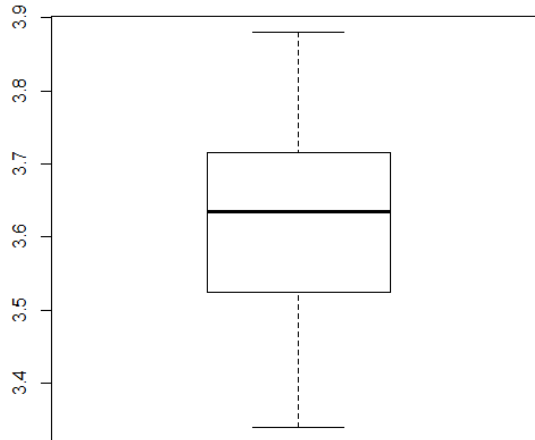
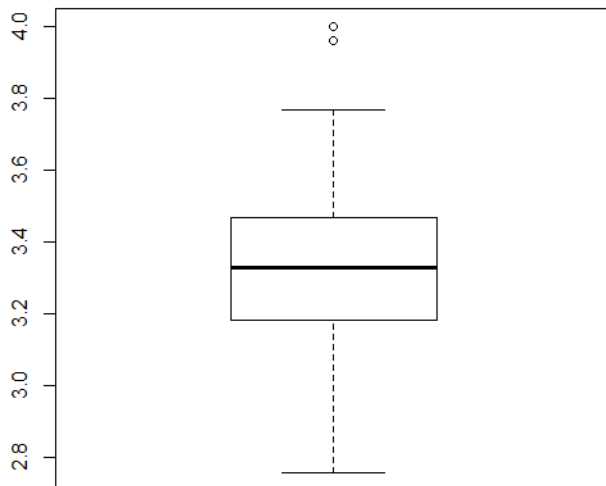
Min. 1st Qu. Median Mean 3rd Qu. Max.  
 3.340 3.528 3.635 3.618 3.708 3.880

Group 2 (STEM):

Min. 1st Qu. Median Mean 3rd Qu. Max.  
 2.760 3.188 3.330 3.338 3.465 4.000

STEM BOXPLOT

NONSTEM BOXPLOT



From the boxplots, I noticed that even though the highest and lowest GPAs are found in STEM majors, the mean of STEM majors, the First quartile, median, and third quartile of the GPAs are all less than the ones found in non STEM majors. So I deduced that the GPAs tend to be lower in stem majors.

My next step was to do a 95% confidence interval for the mean of Stem majors. To obtain the needed data such as variance and standard deviation , I used R .

$$\alpha = 0.05$$

$$\bar{x} = 3.338$$

$$n = 40$$

$$\sigma = 0.272$$

$$Z_{0.05/2} = 1.96$$

$$\text{Lower bound} = \bar{x} - Z_{0.05/2} * \frac{\sigma}{\sqrt{n}} = 3.338 - 1.96 * \frac{0.272}{\sqrt{40}} = 3.2537$$

$$\text{Upper bound} = \bar{x} + Z_{0.05/2} * \frac{\sigma}{\sqrt{n}} = 3.338 + 1.96 * \frac{0.272}{\sqrt{40}} = 3.4223$$

**CI for STEM mean= [3.2537,3.4223]**

Then I repeated the same process for non STEM majors :

$$\alpha = 0.05$$

$$\bar{x} = 3.618$$

$$n = 20$$

$$\sigma = 0.152$$

$$Z_{0.05/2} = 1.96$$

$$\text{Lower bound} = \bar{x} - Z_{0.05/2} * \frac{\sigma}{\sqrt{n}} = 3.618 - 1.96 * \frac{0.152}{\sqrt{20}} = 3.549$$

$$\text{Upper bound} = \bar{x} + Z_{0.05/2} * \frac{\sigma}{\sqrt{n}} = 3.618 + 1.96 * \frac{0.152}{\sqrt{20}} = 3.687$$

**CI for non-STEM mean= [3.549,3.687]**

After obtaining this data, I conducted a two population hypothesis test.

1. Parameter : mean
2. Hypothesis :  $\mu_1$  (non STEM) ,  $\mu_2$  STEM  
 $H_0 : \mu_1 = \mu_2$   
 $H_1 : \mu_1 > \mu_2$
3. Test statistics:

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

4. Reject  $H_0$  when :

$$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > Z_\alpha$$

5. Calculate :

$$\bar{x}_1 = 3.618$$

$$n_1 = 20$$

$$\sigma_1^2 = 0.0248$$

$$\bar{x}_2 = 3.338$$

$$n_2 = 40$$

$$\sigma_2^2 = 0.0742$$

$$Z = \frac{0.28}{0.0556} = 5.03$$

6. Decision:  $5.03 > 1.65$  ; So, we reject  $H_0$ .

7. Calculate p-value :

$$P(Z > z_0) = 2.45 * 10^{-7}$$

$$0.05 > 2.45 * 10^{-7}$$

So we reject  $H_0$ .

Because the Z value found is large, the probability that  $H_0$  is true is very small , so therefore we reject  $H_0$ . We can conclude that non STEM majors tend to have a higher GPA , and therefore perform better than STEM majors.