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

Happiness and Studying, Do They Coexist?
By: James M. Karle Jr.

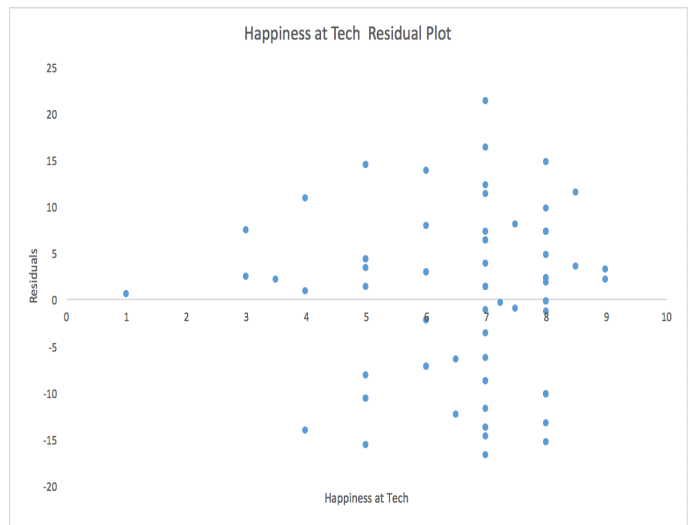
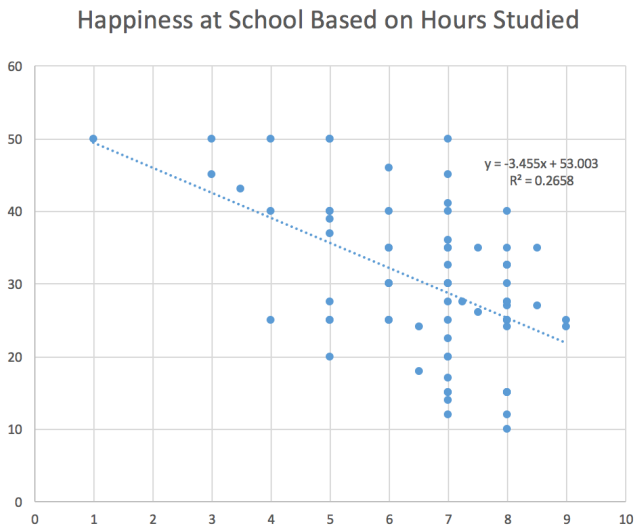
I originally set out to solve a question that had bothered me since my first semester here at Georgia Tech. I wanted to see if there was a correlation between overall happiness at Georgia Tech and the number of hours spent studying for scholarship each week. Is there a positive or negative correlation between the two, would more studying equal higher grades and therefore a better college experience, or is less studying with more free time and potentially lower grades make up happiness at Georgia Tech?

My assumption was that although there would be a happy medium, the average student would be far happier with higher grades because of more study hours than the people with a lower number of study hours. I would test this hypothesis with a linear regression of 74 different students at Georgia Tech. This regression would tell me how linearly positive or negative the relation between my two variables of: happiness at GT on a 1-10 scale and number of hours studied each week.

My method of data collection was a google doc survey that I sent out to all of my friends and acquaintances at GT. The survey had two questions on it worded as such: How many hours for scholarship each week do you study, and how happy on a scale from 1-10 are you with your college experience, including: grades, social life, future success and jobs? I received a total of 74 responses, including myself, and I feel that although it potentially is too small to generalize a statement for all of Georgia Tech it would be a very good size for this project. With a grouping of 74 people recorded, that gave me a bucket number of at least 9, and with the potential of 19 different answers between 1-10, including non natural numbers, there would be plenty of spread.

The results were not what I had expected at all, which I found to be extremely interesting. First thing to mention was the average of both numbers and what these averages tell me. The mean number of hours studied per week was a total of 30.23. The survey revealed a mean value of happiness of 6.59. Solely taking these two numbers and not looking at the distribution or spread of the values, I would say that with an average of less than 7.0, a large number number of people that attend Georgia Tech aren't happy with less than 30 hours a week studying.

Upon looking at the scatter plot of the data, I created a linear regression for the data and found the r^2 value. The equation for the linear regression is $y = -3.455x + 53.003$ and an r^2 value of .2658. Although the r^2 value is so low, looking at the spread of the data makes sense. The variance of the data on the regression line turned out to be 9.009, which provides a big issue for me. There is a high variance with such a defined y equation, this is due to the relatively low number of people that I surveyed. I think that if the whole school was to be questioned we would have a much stronger correlation. When looking at the slope and intercept values of the regression equation, you notice that it has a negative slope with a high intercept, representing the downward angle of the correlation. The problem with this low r^2 value and high variance, is that it makes it difficult to create assumptions and make conclusions. The statistical significance of this value is 2.64857E-06, which unfortunately is not very significant. Another aspect to mention is the spread of my residuals plot, which looks much like a fan. When this is the case, it means that the linear regression is heteroskedastic. Heteroskedasticity refers to the circumstance in which the variability of a variable is unequal across the range of values of a second variable that predicts it. Typically, this means that there are a number of subgroups within the data having varying variances, which makes placing any regression, specifically the one dimensional linear regression, in a tough situation. Some reasons for my high variance, and specifically the heteroskedasticity would be variances in intelligence and personality type at Georgia Tech. The likelihood of intelligence being a large factor is low, due to the selectivity of Georgia Tech, but personality types, especially in regards to study habits, can vary widely.



I would hazard to say that my sample distribution is a little bit biased. Many of the people in my friend group are also apart of the same fraternity, but we are one of the highest ranked Fraternities in regards to GPA on campus. In this situation I would say that the population of guys that I questioned would be similar in one regard, but that being said, I think that it does present a claim that I had previously not realized or anticipated. The mean study time was just over 30 hours a week of studying, this is out

of class work and study for the classes we attend, which is almost equal to working 40 hours a week including attending class, which is at least 12 hours in order to be a full time student. I think that it would be interesting to compare some of the numbers that I attained to seemingly equivalent schools such as the University of Georgia. I think that you would find that at least 42 hours (including classes) of studying per week would be on the high end of their spectrum and that we should realize that our mean of 30 is already exhausting as it stands.

In order to improve my result there are a few things that I would like to change about my testing parameters: 1. Question more people, 2. Add more parameters such as GPA and number of classes being taken, 3. Collect data from another school such as MIT or Cal-Tech. I think that if I increased the number of people that I surveyed, the r^2 value would be higher, but I also think that there would be less of a negative slope on the linear regression.

This information went against what I had hypothesized. The data was showing me that typically, the students at tech have a better experience, are "happier" with their school, if they study less and have more free time. Although the data displayed this mentality, I am not sure I would conclude that it is in fact similar to the actual trend at Tech, whether it be that there is no correlation or that this sample of people happened to study less on average and be more content with worse grades. But I believe that if Georgia Tech could do this survey themselves, they would be able to conclude some interesting things about their students and their happiness throughout their duration here at Georgia Tech.