An analysis of the difference between STEM major and non STEM majors in the proportion of graduates who get a job that requires a college degree

Megan Grabill, Riyad Alam, SeBeom Lee

Introduction

This paper looks at employment outcomes for recent college graduates. More specifically, our group looked at the proportion of recent college grads working in jobs that require a college for both STEM and non-STEM majors. Given the intense speculation about the value of a college degree in a flailing job market, we decided to explore the oft-heard "wisdom" that STEM degrees are better or more valuable than non-STEM degrees. This can be measured in a number of different ways; therefore, we decided to focus our research question on the proportion of recent grads currently working in jobs that require a college degree. While this is a general measure, and does not account for the specificity of employment; after all, working for Boeing, working as a consultant, and working as a middle school teacher all require college degrees, but might directly "use" knowledge from major course of study to varying degrees or require more or less specific undergraduate study. This measure also fails to analyze "value added" in terms of higher wages or earning potential. However, measuring the proportion of jobs that require a college degree allows us to discern a "value added" purely in terms of access to certain types of jobs.

Our data comes from the American Community Survey 2010-2012 Public Use Microdata Series, and looks at employment outcomes (as well as earnings) across 173 distinct majors, which have been further categorized by domain (ie. Engineering, Social Sciences, Health). In order to distinguish between STEM and non-STEM, we used the list of 77 majors referenced in the "Women in STEM" file also included in our data package. Since our data gives the breakdown of employment outcomes by major, rather than on an individual basis, a "major" serves as our case. Our explanatory variable is the type of major, defined broadly as "STEM" and "non-STEM", and our response variable is the mean of the proportion of jobs requiring a college degree divided by total jobs held by recent graduates in the major [College_jobs / (College_jobs + Non_college_jobs)].

Data and Methods

While our research question concerns job outcomes for all recent college graduates in the United States of America, our data comes from the 2010-2012 American Community Survey (ACS), which is facilitated and distributed by the American Census Bureau and is publicly available for use. In a document on the Census website, the ACS is described as "a nationwide survey designed to provide communities with reliable and timely social, economic, housing, and demographic data every year. A separate annual survey, called the Puerto Rico Community Survey (PRCS), collects similar data about the population and housing units in Puerto Rico." [1] Because the survey is conducted by the federal government, we are able to draw from a large sample size– reportedly, "an annual sample size of about 3.5 million addresses, with survey information collected nearly every day of the year." This consistent information collection is averaged out, allowing us to understand trends over time, rather than just every ten years. However, 3.5 million addresses is not as big as one might think; the PUMS (Public Use Microdata) files are published every year and every 5 years, and measure about 1 percent and 5 percent of the total population respectively.[1] Given that this dataset is an aggregate of three years, we can estimate that the recent grads data measures, at best, three percent of the American population– and this survey is conducted by household, not of specifically college students, so it is likely to measure less than three percent of the actual population of “recent graduates”. Furthermore, the data for PRCS is released separately, so this data does not include outcomes for Puerto Rican college graduates. Therefore, we could encounter some sampling bias, in part by potentially not getting a representative sample across the board, and by explicitly excluding Puerto Rico, whose liminal status as a US commonwealth might potentially impact the educational and vocational opportunities available to its citizens– for example, it might have lower employment in jobs requiring bachelor’s degrees across the board, or recent graduates might have to leave the island for more “desirable” types of work.

Measurement bias might also be present; the dataset defines “recent” grads as college graduates younger than 28. This excludes older students who have just completed college, potentially in order to advance in their careers or re-enter the workforce after time away. The data does not specify what type of “college degree” is needed; there is a big difference between a bachelor’s degree and a PhD, and it is not uncommon for mid-career professionals to return to school for a masters degree– could they not also be termed “recent graduates”? This potential measurement bias skews younger, and is likely to capture data from young, early career professionals with little specialization. Particularly given that we are interested in the divide between STEM and non-STEM, this has potential to tip the scale in STEM’s favor– a twenty three year old with a BA in Mechanical Engineering can be, well, an engineer– but it will be hard for a twenty three year old with a BA in Comparative Literature to be a… comparatist?

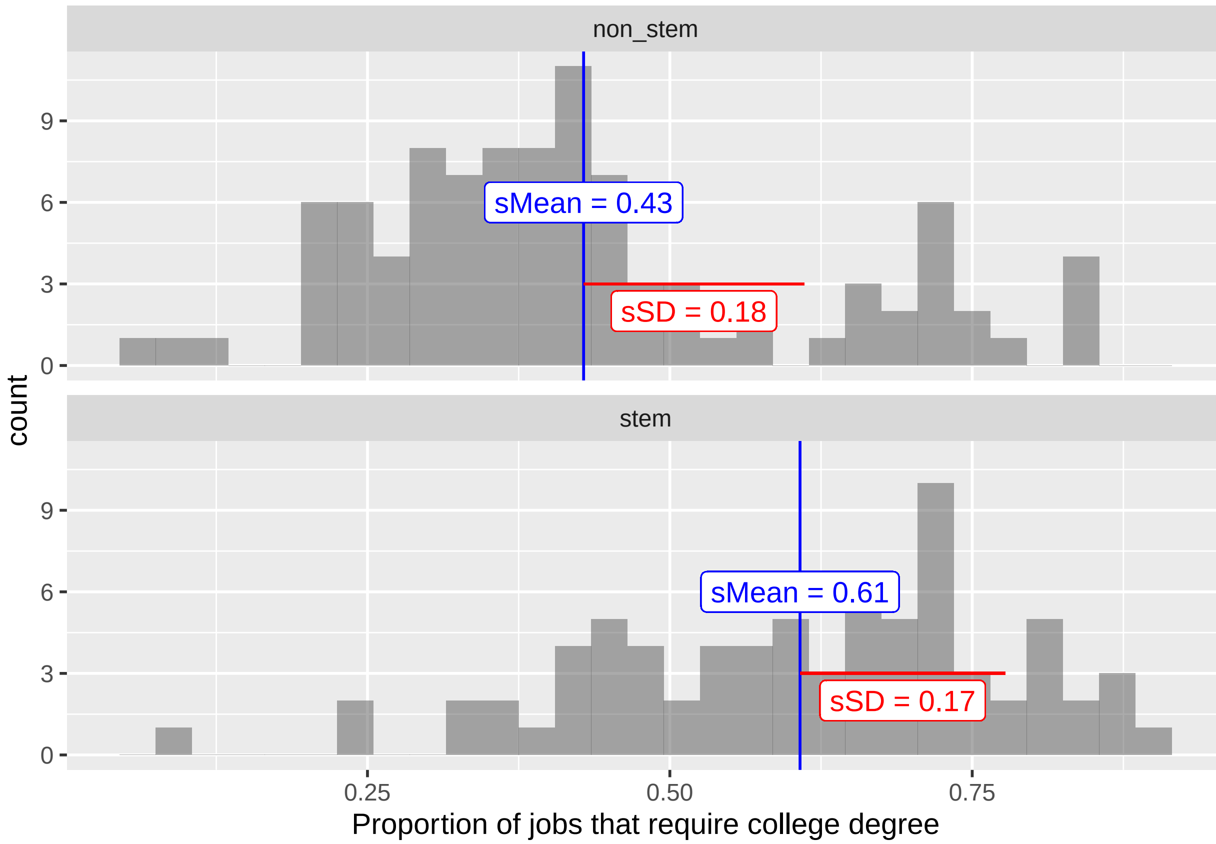
Results

Parameters and Hypotheses

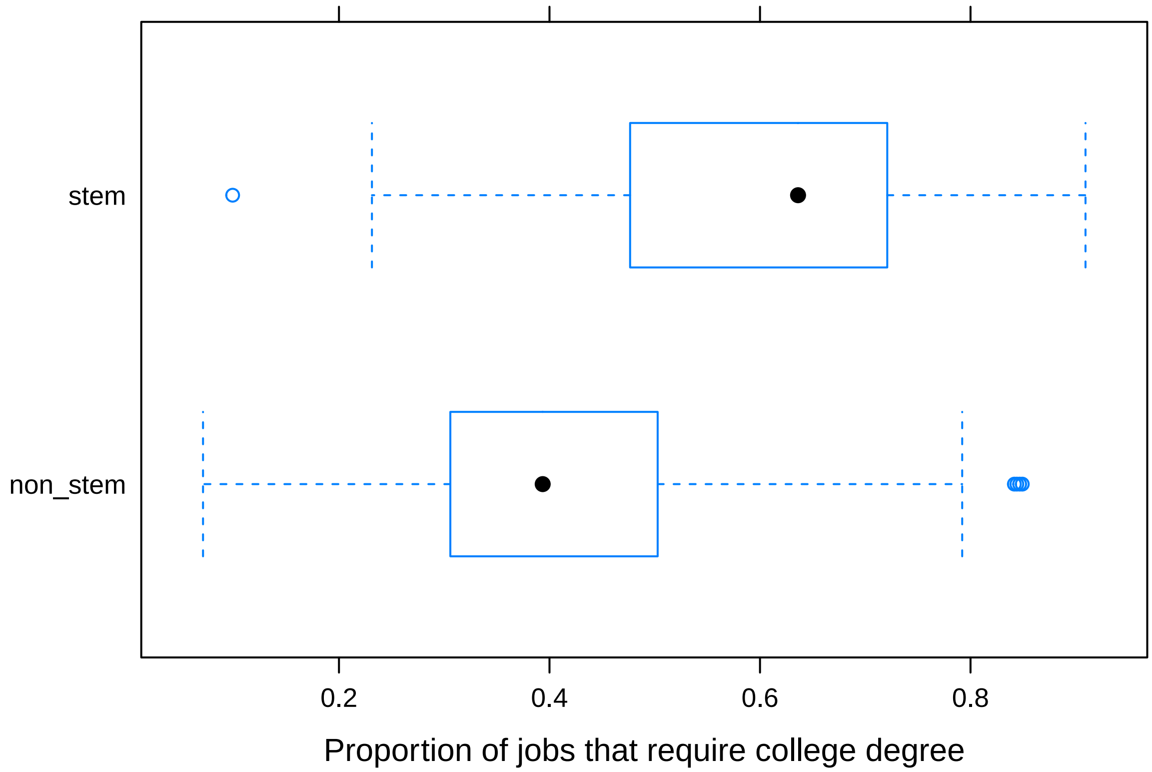
Because the cases that we received from this dataset [2] are college majors and not individualized people, we are looking at the mean difference of proportions between STEM majors and non-STEM majors. Our parameter of interest is trying to learn what the true difference of means is between the proportion of STEM majors who require a college degree to do their job in comparison to the proportion of non-STEM majors who require a college degree for their career. Our null hypothesis is that there is no difference in mean rate of employment in jobs requiring a college degree between STEM and non-STEM majors, while our alternative hypothesis is that there is a positive difference in the mean proportion of employment in jobs requiring a college degree between STEM and non-STEM majors. From these hypotheses, we know that we will need to conduct a right-tailed test to determine our p-value in order to either reject or fail to reject the null hypothesis.

Descriptive Analysis

For the analysis, we divided the dataset into two groups. The group “STEM” has 76 cases, and the other group, “Non-stem” has 97 cases in it. We estimated the levels of the variable by calculating the proportion of the jobs that require a college degree out of the jobs that graduates with each major have. Figure below shows the distribution of the variable in two groups.



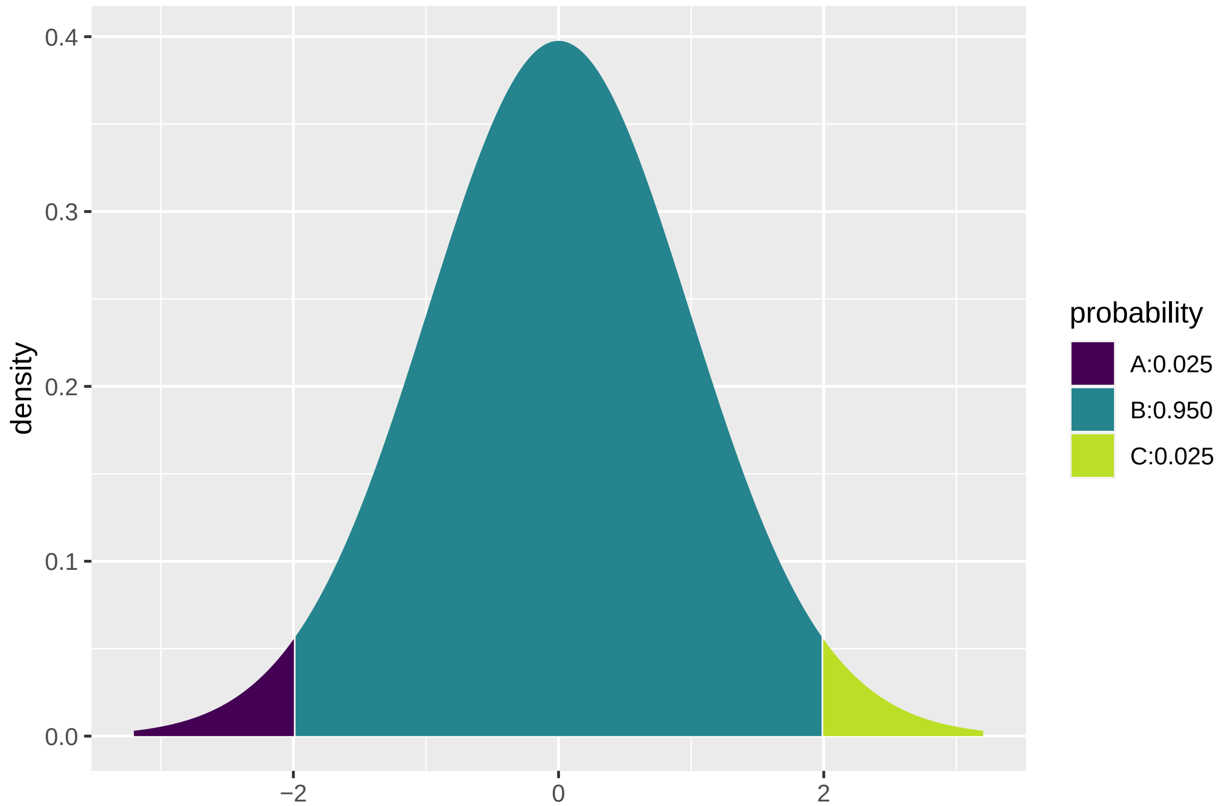
Looking at the initial graphs of the proportion of jobs that require a degree, it definitely appears that, on average, STEM majors end up in jobs that require a degree more than their non-STEM counterparts. While there are many other confounding variables that we cannot see from these graphs, such as age, gender, or race, we can calculate that the mean proportion of STEM students whose work requires a degree is 0.61 with a standard deviation of 0.17 while the mean proportion of non-STEM students whose job does not require a degree is 0.43 with a standard deviation of 0.18. Simply based on looking at these graphs, we would be inclined to expect that there is a difference in the mean proportion of degree requiring jobs based on the field of a graduate's major since the peak for STEM degrees that require a degree is around 0.7 and the peak for non-STEM degrees is around 0.4, showing us that there is a greater proportion of non-STEM majors that do not require a college degree for their post-graduation careers. When looking at the proportions of college related jobs that recent graduates had, the STEM majors were slightly right-skewed with possible outliers towards the lower proportions. The non-STEM majors were fairly normally distributed with a slight right skew or bimodal distribution as the peak is centered around 0.4, but there is another smaller peak around 0.72 and 0.8. The 0.8 proportion may be a slight outlier for the non-STEM majors, but it is important to recognize that the skewing between the two majors is in opposite directions.



When we plot the box-whisker plot, we can identify one outlier in STEM group and 4 outliers in non-STEM group, but considering that the difference between the mean and the median is fairly small (stem: -0.029, non-stem: 0.035), we decided that the effects of outliers is not extreme, and mean can be used as a representative statistic for each group.

Our statistic of interest $x_{stem} - x_{non}$ is 0.18. The difference in mean is positive, which means that the mean of STEM group is bigger than that of non-STEM group.

Inferential Analysis



In order for us to do an analytic approximation there are two checks that we need to look at for a difference of means. One of the tests to see whether analytic approximation is reasonable is whether the data is normally distributed or whether both groups have a sample size greater than or equal to 27. We wouldn't say that the distribution of these two datasets is entirely normally distributed because there is some slight skewing, but we definitely do have the condition of sample sizes of 27+ satisfied as there are 76 majors that are included in the STEM category and 96 in the non-STEM category. Since we are doing a t-distribution when doing an analytical approximation for difference of means, we would first calculate our standardized endpoints from a t-distribution with 75 degrees of freedom since that is the smaller of the two sample sizes, and we would then find that our t-endpoints for a 95% confidence interval are $t_{end} = \pm$ 1.99. Our sample statistic is the mean proportion of STEM majors minus the mean proportion of non-STEM majors who are in a job that require a college degree, which is $x_{stem} - x_{non}$ is 0.18.

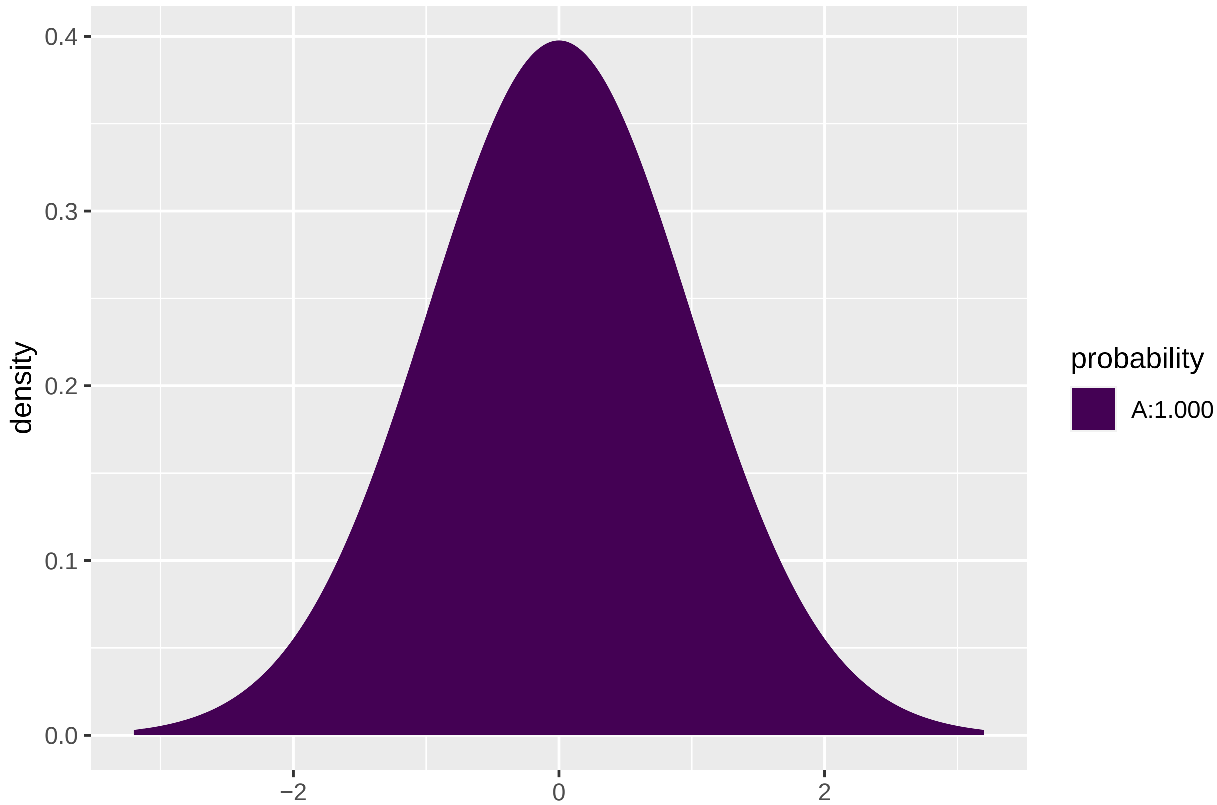
To estimate the standard error, we used an equation $\hat{SE} = \sqrt{\frac{s^2_{stem}}{n_{stem}} + \frac{s^2_{non}}{n_{non}}}$. From the given sample, we could estimate the standard error of $\hat{SE} = 0.03$.

With the sample statistic $x_{stem} - x_{non} = 0.18$ and standard error of $\hat{SE} = 0.03$, we calculated the 95% confidence interval using an equation of $CI = SampleStat. \pm t_{end} \cdot \hat{SE}$.

We could obtain a 95% confidence interval for the difference of means from 0.13 to 0.23, which means that we are 95% confident that the percentage of graduates who get a job that requires a college degree with STEM majors is from 13 to 23 percentage points higher than that of graduates with non-STEM majors. One implication of the fact that the confidence interval does not include zero is that we are 95% confident that the difference between two groups is not zero.

Our null hypothesis is that $H_0 : \mu_{stem} - \mu_{non} = 0$ and the alternative hypothesis is $H_1 : \mu_{stem} - \mu_{non} > 0$.

To test the hypothesis we estimate how likely it is to observe the result as extreme as our observed sample. Using the sample statistic and the standard error estimated above, we set up a test statistic using an equation $TestStat. = \frac{ObservedDiff. - NullDiff.}{SE}$. With the observed difference of 0.18, null difference of 0, and standard error of 0.03, we calculated the test statistic of $TestStat. = 6.64$. Then, we estimated the P-value by a right-tailed test on a t-distribution with degree of freedom 75.



We obtained a P-value of 2×10^{-9}, which is extremely smaller than $\alpha = 0.05$, which means that if there were no difference between STEM group and Non-STEM group, the probability of observing result as extreme as our sample is almost 0. Therefore, we can conclude that there exists a statistically significant difference between two groups and reject the null hypothesis. Aligning with our alternative hypothesis, we conclude that the percentage of STEM major students who get a job that requires a college degree is significantly higher than that of non-STEM major students.

In addition, we could obtain a Cohen's d of 1.01 which can be considered relatively large.

Discussion

We began this project by asking ourselves whether there was a difference in means between the proportion of jobs that require a college degree between recent college grads who either majored in STEM fields or majored in the humanities? By just looking at the data that was provided we already had a sense that STEM majors had a seemingly higher proportion of majors that resulted in students working in jobs that required a college degree as opposed to all humanities majors. When we conducted our hypothesis test there was no surprise that we did find a significant result that allowed us to reject the null hypothesis that there is no difference in the mean proportion of degree required jobs.

From this study we can take away that of the sample that was interviewed, there was a difference in the mean proportion of jobs that needed a college degree, but we cannot directly draw a conclusion because this is an observational study that does not account for confounding variables or test for controls. As we had discussed earlier, there is definitely a measurement bias because of how this sampling defines recent graduates as being under the age of 28, seemingly erasing an entire population of graduates who have chosen to continue their education after their undergraduate experience. There are also many confounding variables that are not directly addressed, specifically thinking about race and gender. It is very possible that certain majors attract more men, and certain careers may also view educated men as more competent than their female counterparts. This study erases the idea that certain majors are male dominated, as well as certain fields only being interested in upholding these patriarchal values. In order to draw a more convincing conclusion, it would be helpful to have data on individual outcomes, thus allowing us to better account for the effects of race, gender, age, and degree type on employment outcomes. While the ACS is facilitated through the federal government, it also only samples a very small population and does not specifically target college graduates; therefore, using data from a survey dedicated to job outcomes for recent college graduates might allow us to capture a more representative sample.

References

[1] Dillingham, Steven, et al. "Understanding and Using the American Community Survey Public Use Microdata Sample Files." U.S. Census Bureau, Apr. 2020.

[2] Mehta, Dhrumil. https://Github.com/Fivethirtyeight/Data/Tree/Master/College-Majors, 2014.

"We affirm that we have adhered to the Honor Code in this project report"