proposal_1

Christina Lyu

10/2/2018

Group Members:

Christina Lyu, Aoi, Ahlam

Title:

Relationship between Salaries after Graduation and Race, Major or Gender

Purpose:

As college students, we work hard for our education and it is true to almost everyone that we would want to get a job one time or another. Therefore, we want to look at the factors that would affect ones salary. For example, it's very common that the people with computer science degrees might get jobs easier than people with art degrees. It is also common that people with higher degrees get higher salaries than people in the same field but with lower degrees. But majors aren't the only one with an influence. Despite the fact that we try to build a world with no discriminations, there still exist in the society different treatments for different people so we are also interested to see how races, gender would make change to the salaries.

Data:

1.Our data was found on this website. Each row represent a specific major in a specific year, with columns representing different factors including teaching, student, male, white and so on. The numbers are the average yearly salaries for that factor. There lines with all 0s means that major was not created that year. https://think.cs.vt.edu/corgis/csv/graduates/graduates.csv?forcedownload=1

2.The data has multiple categories that could have played a role in the salary of the individual, rather than just the major i.e. year, minorities, gender, etc.

https://catalog.data.gov/dataset/current-employee-names-salaries-and-position-titles-840f7/resource/2b09fa08-c85e-4bbe-a90d-cdc4ae39b0cc

3.The data compares the salary for people who live in the city in Chicago. A flaw in us using this data is that this data is catered to Chicago, so we should be more specific if we want data for a specific city or just the national average.

https://cew.georgetown.edu/cew-reports/valueofcollegemajors/#explore-data

4.This data is in response to the previous data presented. This specific data shows earning by state and includes the percentiles for the earnings. I think if we wanted to do a box-and-whisker plot for each state to show the data, it would be great to use the information from the data.

https://www.forbes.com/sites/susanadams/2015/07/02/the-college-majors-with-the-highest-starting-salaries/#ed527cb35024

5.This data shows that our hypothesis was correct in which STEM majors make more than Humanities majors. http://online.wsj.com/public/resources/documents/info-Degrees_that_Pay_you_Back-sort.html

6.This data shows a more extensive version of college majors, and it also shows the different percentiles for the salaries. With these percentiles, we can definitely compare the different majors and see if there are any significant differences.

https://www.glassdoor.com/blog/50-highest-paying-college-majors/

7.This data only focuses on the median salary, which is different than the previous dataset. This would show more about the middle 50% of salaries, which would be good to not have any extreme values.

https://www.cnbc.com/2016/10/19/the-50-highest-paying-college-majors.html

8.The data shows the top 50 highest paying college majors, which would not include every major. This may or may not include certain categories in our hypothesis.

9.The data below shows the expected salaries for undergraduates who majored in humanities and arts. It listed philosophy as the top paid major in the class of 2016 out of the other humanities and arts majors.

http://www.naceweb.org/job-market/compensation/philosophy-projected-as-top-paid-class-of-2016-humanities-major/

With all these differing data, I think we should narrow down the results to see if we want it to be by state or nationwide.

Population:

The population of the study is everyone with a college degree and a job. Since there are columns called "Bachelors" and "Masters", which means we are taking samples with someone with at least a bachelors degree. The values under the columns represent the salaries one get so our sample must have jobs. The observation units are majors in different years. I'll try to generalize the predicted salary for people after they graduate based on their gender, race, major, degree and so on. There are 517 rows of data in the data frame, but I'm sure there are majors that are not covered in the data.

Response Variable:

The response variable is the salary measured in units Dollar and has an approximate range of 1000 to 200,000.

Explanatory Variables:

The explanatory variables can be many possible combinations including Student, Teaching, Whites, Male, Major, Bachelors and so on. We don't know which of them are statistically significant yet. They are all categorical variables so the actual values are 0 or 1. But in the data set, each element will have a corresponding salary as the mean salary for someone with the major that year who has a value of 1 for that factor.