Analyzing the Correlation between Degree Percentage and Salary

Overview:

India has arguably one of the toughest higher education systems. Bloomberg's article titled: **India's IIM-A, the World's Toughest B-School to Get Into** emphasizes how India's top business school accepted less than 1% of applicants in 2014-2016 academic years. The article also sheds light on Stanford University's' acceptance rate at 7% acceptance rate and Harvard University's' being at 13%. The minimum GMAT score for students in IIM-A's postgraduate program in management, equivalent to a two-year MBA program, was 770, compared with an average score of 730 at Stanford's Graduate School of Business, according to data reported by 200 global B-schools to QS Quacquarelli Symonds, a London-based researcher. Based on these statistics, we have decided to analyze data on the Indian higher education system.

The dataset we chose consists of 14 variables. It includes gender, placement data of students at a university, the secondary (Grade 10), higher secondary (Grade 12) final percentages, MBA percentages, entrance test percentages, which boards of education they studied for. It also includes the undergraduate degree specialization, the percentage they received, their work experience, their status, and their salary. Based on this data, we're going to analyze the correlation between the degree percentage of students and their salary.

Assumptions:

We assumed that the gender, Secondary Education percentage (10th Grade), Boards of Education, Higher Secondary Education percentage (12th Grade), specialization in Higher Secondary Education, degree type, work experience, MBA percentage, and status had no influence on a student's salary earned after graduating.

Data Analysis:

We used R and Excel to find the relationship between the degree percentage and salary earned by students after earning their degree. The table below shows the central tendencies that were calculated for this data analysis.

	Degree_p (%)	**Salary (INR)**
Minimum	50	200000
Mean	66.37	288655.41
Mode	65	300000
Median	66	265000
Q1	61	240000
Q3	72	300000
IQR	11	6000
Maximum	91	940000

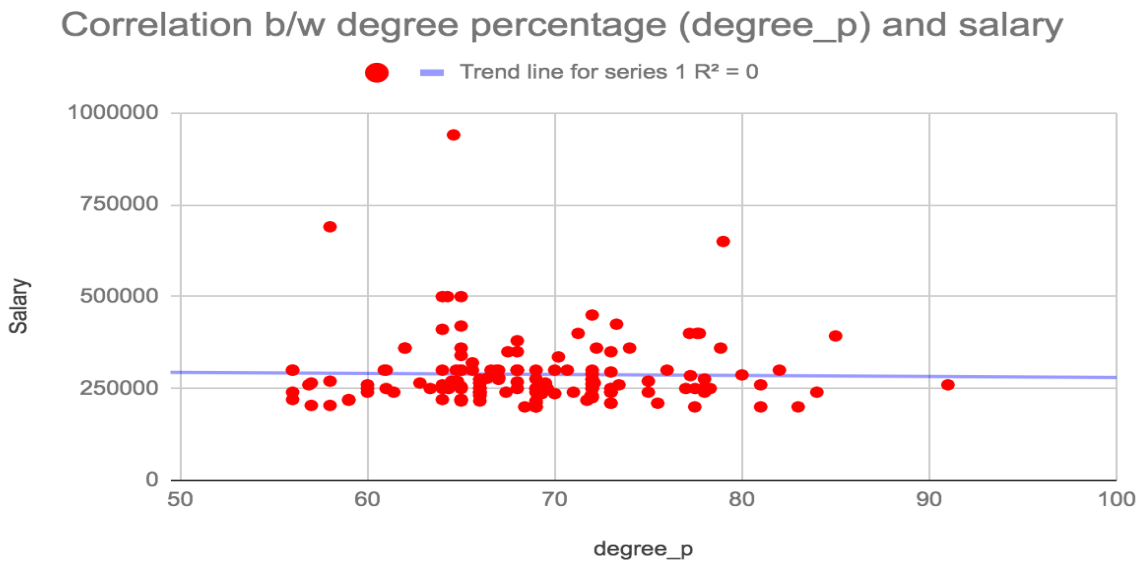
Degree_p has a minimum of 50 and a maximum of 91 on a scale of 0-100. On the other hand, Salary is nominal data and has a minimum of 200,000 and a maximum of 940,000. Q1 shows the lowest 25% of degree percentage and salary. Q3 shows the highest 75% of degree percentage and salary.

Note: Salary is in INR. 1 INR = 0.014 USD (as at March 18,2021)

Data Visualization:

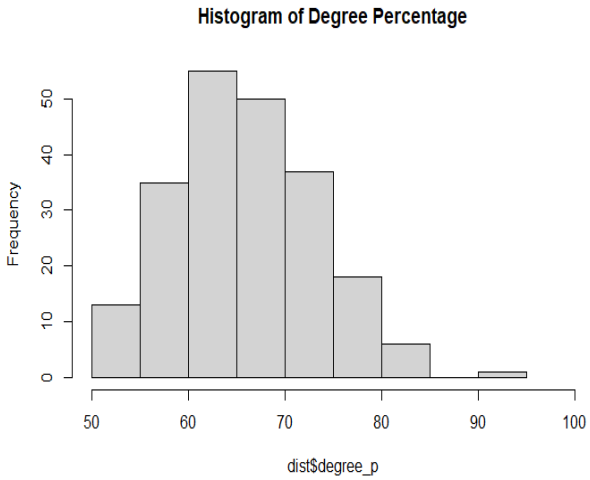
The data visualization has been done using both R and Excel. A scatter plot was used to illustrate the correlation between degree percentage and salary.

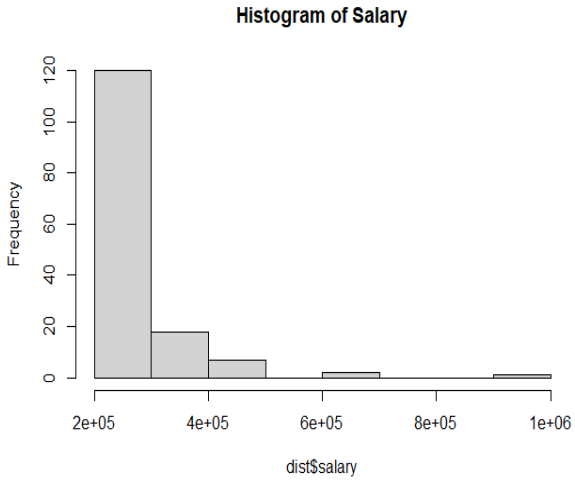
- *Using Excel*

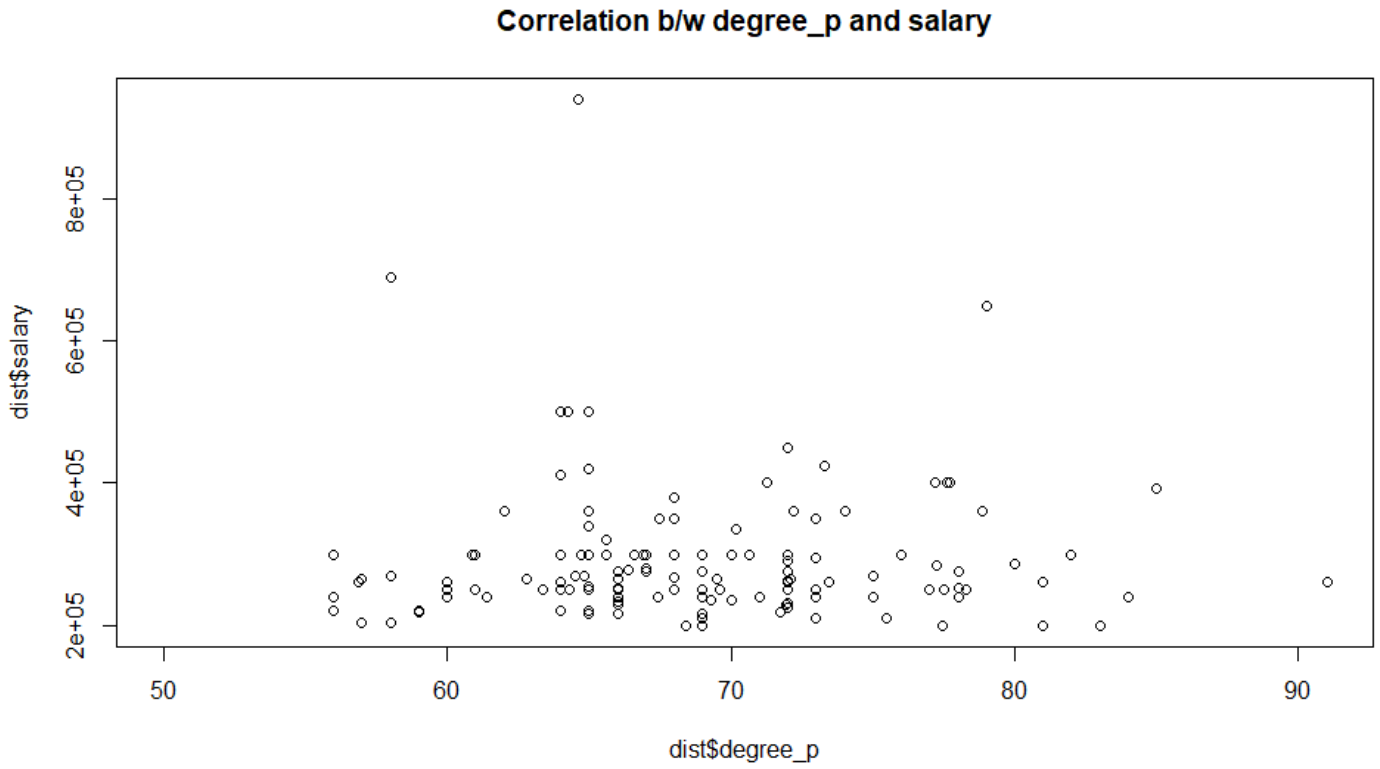


Based on the scatter plot designed in Excel, it can be identified that there is no correlation between degree percentage and salary. The trend line represented in blue, shows where most of the data is clustered. The correlation coefficient was also calculated to be 0.0 proving that there is no relationship between degree percentage and salary.

- *Using R*


Histogram of Degree Percentage
Frequency
0
10
20
30
40
50
50
60
70
80
90
100
dist$degree_p


Histogram of Salary
Frequency
0
20
40
60
80
100
120
2e+05
4e+05
6e+05
8e+05
1e+06
dist$salary


Correlation b/w degree_p and salary
dist$salary
2e+05
4e+05
6e+05
8e+05
50
60
70
80
90
dist$degree_p

Conclusion:

Based on our analysis, we found out that there is no correlation between the degree percentage and the salary earned by a student. The degree percentage earned by a student does not determine the salary that the latter earns after graduating.

During the process, we also realized that this was not the best dataset as it only includes surveys from people who specialized in either Marketing & Human Resources or Marketing & Finance. This limits our analysis to only this population but as we know students from Higher Secondary School disperse into various different fields of education and even in undergrad, there are various subjects that are taught. In our opinion, a better dataset would have included surveys from different fields of specialization.

Overall, we really enjoyed working on this project. It was an enriching experience putting into practice what we learned in this class. We used R to analyze the relationship between different variables and used different data visualization techniques like scatter plots and histograms to visually analyze the data.

Appendix:

campus_recruitment.R

```r
#The next line reads in the csv data
#please replace the csv location as given when run file.choose() and choosing Placement_Data_Full_Class.csv

file.choose()
dist<-data.frame(read.csv("C:\\Users\\hoang\\Documents\\STAT201\\Placement_Data_Full_Class.csv"))

nrow(dist) #number of row in data set
summary(dist)
#Here's a way to output the descriptives in a format that's a little easier to import for your report.
write.csv(data.frame(summary(dist)),"C:\\Users\\hoang\\Documents\\STAT201\\Placement_Data_Summary.csv")

#Look at the distributions of the Degree Percentage and Salary variables.
#degree percentage histogram
hist(dist$degree_p,
     main = paste("Histogram of Degree Percentage"),
     xlim = range(50,100)) #xlim range we chose 50 because it is the minimum
#salary histogram
hist(dist$salary,
     main = paste("Histogram of Salary")) #change label if possible

plot(dist$degree_p,dist$salary,
     main = paste("Correlation b/w degree_p and salary"))

#cor(dist$degree_p, dist$salary) #supposedly find correlation between degree_p and salary
```

Citations

https://www.bloomberg.com/news/articles/2013-04-18/indias-iim-a-the-worlds-toughest-b-school-to-get-into

https://www.kaggle.com/benroshan/factors-affecting-campus-placement