Team Banana  
- Haeryoung Kim, Ohsang Yoo, Taewoo Kim, Soyeon Kim, Hyeongeun Lee

Professor Jevin West

Data Science and Management

July 09th, 2015

**Preliminary Analysis and Summary**

1. About Observations

The dataset we are going to analyze is from ‘2013 Graduates Occupational Mobility Survey in Korea' researched by Korea Employment Information Service. To find out the relationship, we choose five factors with 5273 observations among hundreds of variables in the raw data; GENDER, AGE, GPA, ENG\_SCORE and INTERN\_DUMMY.

Factor 'GENDER' is one of what we want to seek. In Korea, sexual discrimination in employment field is severe. We assume that so called 'Glass Ceiling' can affect when getting a job in Korea. Therefore, we pick 1 as male, and 0 as female in gender variables. ‘GPA’ is students' grades so it is essential source for entering company. 'ENG\_SCORE' variable is Certified English Test including TOEIC, TOEFL PBT·CBT·IBT, and TEPS which is converted by standard criteria. ‘AGE’ variable is Korean birth standard. This factor influences employment field. Because we consider that internship experience can affects employment. Factor 'INTERN\_DUMMY' is also included in our dataset. The value '1' means the graduate student who has internship experience, and the other is opposite. From now with these five variables, we will analyze the effects to job employment.

1. Descriptive Statistics and Plots

EMPLOYED variable mean is 0.6554144, standard deviation is 0.475278. GENDER variable mean is 0.4699412, standard deviation is 0.499143. ENG\_SCORE variable mean is 214.2737, standard deviation is 343.2686, median is 0, minimum is 0, maximum is 990. AGE variable mean is 25.76882, standard deviation is 2.532085, median is 26, minimum is 20 , maximum is 39. GPA variable mean is 3.661739, GPA variable standard deviation is 0.4187876, median is 3.7, minimum is 1.5, maximum is 4.5. INTERN\_DUMMY variable mean is 0.2545041 , standard deviation is 0.4356234.

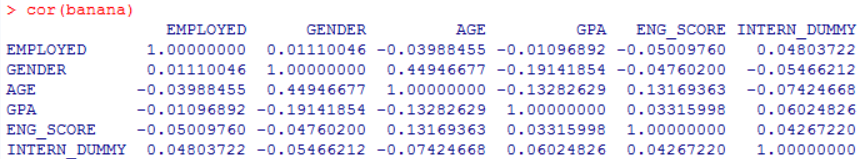
Histogram of each variable

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |

Some of the plots showing ‘The relation of variables’

|  |  |  |
| --- | --- | --- |
|  |  |  |

Correlation values between variables



3. Employment vs. Unemployment

As the table shown below, among dependent variables, Age, English Score, and Internship Experience have low P-value and significant effects to response variable, Employment. On the other hand Gender, and GPA are not significant variables.

Result of T test

|  |  |  |
| --- | --- | --- |
| Variable | P-value |  |
| Gender | 0.4201 | Cannot Reject |
| Age | 0.00446 | Reject |
| GPA | 0.4252 | Cannot Reject |
| English Score | 0.0003407 | Reject |
| Internship Experience | 0.0003796 | Reject |

4. Plan of Analysis

To find what we are looking for, we chose several statistical methods for the data. First of all, logistics model, also known as logistics analysis can be used. Our data contains categorical(binary) response 'Employed'. This variable is not appropriate for original regression model, so we had to take another model, logistics model instead. The binary logistic model is used to predict a binary response based on one or more predictor variables. That is, it is used in estimating the parameters of a qualitative response model. We will find the correlations between job employment and all independent variables including categorical variables. The logistic regression model will provide coefficients of each predictors that explains a possibility of the event, 'getting a job' occurs. In final analysis, our team will start fitting process. And then, we can go on post analysis to find our analysis was appropriate. This progress includes several statistical methods such as, goodness of fit and chi-square test. Finally, we will get conclusion of what we analyzed. Final report will contain relationship between employment and several independent factors like GPA score, age, gender, etc. by clarifying coefficients of regression. We hope it can solve our curiosity about factors which are effective when we get a job after graduation.

**Team Members**

|  |  |
| --- | --- |
| **C:\Users\LHG\AppData\Local\Microsoft\Windows\INetCache\Content.Word\현근.jpg** | Hyeongeun Lee  Department of Statistics |
| C:\Users\김소연\Desktop\P20150613_174832000_4ADC19DB-8E51-4F10-AC0C-0CD3B1EDE663.JPG | Soyeon Kim  Department of Statistics |
|  | Haeryoung Kim  Department of Statistics |
|  | Taewoo Kim  Department of Computer Education |
|  | Ohsang Yoo  Department of Korean Literature |