## Report on MBA Salaries Analysis

### Pratyusha Maiti

#### Email: [pratyusha.maiti@gmail.com](mailto:pratyusha.maiti@gmail.com)

#### Jadavpur University

## Dependance of starting salary

### 1. Boxplot analysis:

The boxplot studies two factors: **sex** and **first language** and plots its effect on starting salary. Clear conclusions are:

* Males have a higher median starting salary
* Higher starting salary is observed for candidates with English as first language

### 2. Scatterplot analysis:

Jitter plots are drawn for studying dependence of starting salary on various factors. Conclusions drawn are:

* There is no immediately visible dependence between GMAT scores and starting salary.
* A bulk of placed students have first language being English.
* Most well-paid candidates lie in 25-30 years range.
* More work experience led to higher starting salaries for females but not the same for males.

### 3. Chi-squared and T tests:

Chi-squared and T tests performed on the variables showed the following dependence:

* Variables like **age, gmat scores,** and **work experience** have a strong effect on the starting salary of the candidates. A very low p-value of the order 10^-5 removes null hypothesis.
* However, variables like **spring** and **fall MBA average** have no such dependence. The p-value obtained are 0.9524 and 0.2518 respectively and being >0.05, null hypothesis cannot be neglected.

### 4. Regression models:

Initially a model was tested with no interaction terms. The resulting output showed a high RSE and low R-squared value. Hence, interaction terms were added to improve the model.

In increasing order of accuracy, the models tested were:

* MODEL 1:

salary ~ age + sex + age:sex + gmat\_tot + quarter + work\_yrs + frstlang

The above model generated an RSE of 13390 on 95 degrees of freedom, with Multiple R-squared: 0.4773 and Adjusted R-squared: 0.4387.

* MODEL 2:

salary ~ age + sex + age:sex + gmat\_tot + age:gmat\_tot + quarter + work\_yrs + frstlang

The above model generated an RSE of 13160 on 94 degrees of freedom, with Multiple R-squared: 0.5002 and Adjusted R-squared: 0.4576.

* MODEL 3:

salary ~ age + sex + age:sex + gmat\_tot + age:gmat\_tot + age:sex:gmat\_tot + quarter + work\_yrs + frstlang

The above model generated an RSE of 12870 on 93 degrees of freedom, with Multiple R-squared: 0.5272 and Adjusted R-squared: 0.4815.

* MODEL 4:

salary ~ age + sex + age:sex + gmat\_tot + age:gmat\_tot + age:sex:gmat\_tot + quarter + age:quarter:sex + work\_yrs + age:work\_yrs + frstlang

The above model generated an RSE of 12330 on 91 degrees of freedom, with Multiple R-squared: 0.5755 and Adjusted R-squared: 0.5241.

* MODEL 5:

salary ~ age + sex + age:sex + gmat\_tot + age:gmat\_tot + age:sex:gmat\_tot + quarter + age:quarter:sex + work\_yrs + age:work\_yrs + age:work\_yrs:sex + frstlang

The above model generated an RSE of 11930 on 90 degrees of freedom, with Multiple R-squared: 0.6068 and Adjusted R-squared: 0.5543.

Finally, the model that produced best results was Model 5.