James Rohan Gangavarapu

james.reckon@gmail.com

Georgia State Univiersity

Analysis of MBA Salaries

All variables seem to be normally distributed according to the Scatter Plot Matrix.

Categorical data seem to be skewed or have multiple normal curves directly proportional to the

number of sub categories for each category. Ex: There are 4 normal curves for the Satisfaction Distribution

Very few linear relationships esist according to the Scatter plot Matrix. Most of those that do exist have

very weak correlation i.e. close to zero. The strongest correlation seems to between age and work experience

which intuitively does make a lot of sense.

Very few males and females who got placed and whose first language is not English.

Men Exhibit higher levels of satisfaction in the 5,6,7 categories relative

to Woman in the same category, but it is due to the fact that there are more Men than

Woman in the data set who got placed.

Men have a much higher average starting salary than Woman.

Men have a higher average GMAT quantitative percentile

Both Men and Woman have about the same average GMAT Total Scores, Men having a slightly higher

score.

Both Men and Woman have about the same average GMAT Total Percentile, Men having a slightly

higher percentile score.

Both Men and Woman have about the same GMAT Verbal Percentile, Woman having a slightly higher

score.

Both Men and Woman have a about the same average Spring scores, Woman having a slightly higher

score.

Both Men and Woman have a about the same average Fall scores, Woman having a slightly higher

score.

The proportion of 1,2,3,4 quartiles among males and females is about the same.

Men have a higher average Work Years’ experience.

Second Linear Model(StartSalLM\_2) is based on the findings from CrossTable and aggregate functions.

Since the Second Linear Model(StartSalLM\_2) has the largest P-Value (0.0725) it is the most accurate. The first

linear model with a very small P-Value is the least accurate. Also, it

has been constructed on the earlier findings. Hence the dependent variables have been chosen

as per the relavant previous findings.

The Second Linear Model(StartSalLM\_2) also has the smallest mean error value of the Three linear models.

There is strong evidence to suggest sex, gmat\_qpc, gmat\_tot, gmat\_tpc, gmat\_vpc,

quarter, work\_yrs, frstlang are important predictors for the starting salary.

The higher wages for Men can be explained by the fact that men in this data set on average are

older than woman and as result also have higher average work experience, all else equal(almost equal)

men end up earning higher due to their higher average work experience.

According to the t-Tests on lines 201-203 there is evidence to suggest, the means of salary, age and work

experience differ between men and woman.

Comparing the box plots of work experience between men and woman between those who were placed

and thos who were not, one can concluded work experience is not a that big of a determining factor.

This data set does not allow us to investigate what role the first language plays since an overwhelming

majority in this data set have English as their first language. Hence this is data set is biased.

According to Chi Square Tests on lines 262-268 we reject the Null Hypothesis that sex is an

independent factor for both being placed and not being placed.