**Report On MBA Starting Salaries case study HBR**

By

Name: Ankit Kashyap

[ankitkashyap.te15@rvce.edu.in](mailto:ankitkashyap.te15@rvce.edu.in)

Phone: 9582854497 College: R.V College Of Engineering

Introduction :This is the analysis of MBA Starting Salaries HBR Case study. The data set with which we are working over here is a classification data set of MBA Students and their salaries, GMAT Score, GMAT Percentile, Age, Sex e.t.c.

From my analysis of the data I have concluded the following inferences. These inferences are made by executing the R code snipped one after the other for easier understanding:

1. The MBA field is dominated by men over women by a ratio of 3:1.
2. The mean GMAT score is 619.5 and the median is 620.
3. The mean spring average is 3.025 and mean fall average is 3.062
4. The mean work experience is 3.872yrs
5. The mean percentile of the students is 84.52 and the median is 87.
6. The number of placed students to those who weren’t is almost in a 1:1 ratio.
7. There is a positive correlation between Work experience, salary and age.
8. Positive correlations are seen between, quarter and s\_avg, f\_avg, gmat\_tpc, gmat\_vpc, gmat\_qpc and gmat\_tot.
9. Most important predictor for price Salary is work\_yrs, gmat\_tpc and age..
10. Hypothesis test
    1. Observations from Contingency Table and Chi square test.
       1. We cannot predict the sex of the MBA student from work experience(mytable)
       2. We can predict the satisfaction level of the MBA program from work experience.(mytable1)
       3. We cannot predict the sex of the MBA student from first language(mytable2)
       4. We can predict the work experience of the MBA program from first language.(mytable3)
    2. Observtion from T test
       1. We can’t say that average salary of males is greater than females because p value is more than 0.05
       2. We can’t say that average salary of people whose first language is English is greater than average salary of other language speakers As our p value after conducting t test is more than 0.05 so we can’t reject our null hypothesis.
       3. We can’t say that Average GMAT percentile of male is greater than that of female. As our p value after conducting t test is more than 0.05 so we can’t reject our null hypothesis.
11. Multiple Regression Model1:
    1. The first model has Salary as response variable
    2. Predictor variables are work\_yrs + gmat\_tot.
    3. The R square value is 0.9712 which is a very good model. The model’s, p-value: < 2.2e-16 is also lower than the statistical significance level of 0.05, this indicates that we can safely reject the null hypothesis that the value for the coefficient is zero (or in other words, the predictor variable has no explanatory relationship with the response variable).
12. Multiple Regression Model2 :
    1. The first model has Salary as response variable
    2. Predictor variables are work\_yrs+age\*frstlang+gmat\_tot+sex. The R square value is 0.9785 which is a very good model. The model’s, p-value: < 2.2e-16 is also lower than the statistical significance level of 0.05, this indicates that we can safely reject the null hypothesis that the value for the coefficient is zero (or in other words, the predictor variable has no explanatory relationship with the response variable).
13. Multiple Regression Model3:
    1. The first model has Salary as response variable
    2. Predictor variables are work\_yrs+age
    3. The R square value is 0.2506 which is average model. The model’s, p-value: < 5.438e-07is also lower than the statistical significance level of 0.05, this indicates that we can safely reject the null hypothesis that the value for the coefficient is zero (or in other words, the predictor variable has no explanatory relationship with the response variable).
14. In logistic regression the dependent variable can only be categorical so we selected sex as the dependent variable
15. The model is prepared for both who have a starting salary(i.e. a job) and those who don’t.
16. The accuracy of both the models came out to be very less and almost equal to approximately 4.5%