Policy Research Report 4

I. Introduction

This evaluation study examines data regarding the first salary paid to engineering graduates in India, just after graduation, with a specific focus on understanding the role of gender in affecting the salary amount. The engineering discipline has long been a sought after academic field of study in India. However, the field has historically been dominated by men and remains so. At present, the gender distribution of Indian engineering undergraduates stands at a 3:1 ratio of males to females[1]. Indian society is also known to be deeply patriarchal, having poorer human development indicators for women versus men[2]. Women, thus, face discrimination and inequality which is perceived to translate into wage gaps due to gender bias. However, Indian society is evolving and a new wave of feminism and conversation regarding gender equality is on the horizon. In this light, it is interesting to understand whether the female youth of today are facing gender bias in terms of their employment outcomes.

This study aims to answer the question of whether there is gender bias in the first salary amounts paid to fresh engineering graduates in India. More specifically, it seeks to examine whether gender affects performance indicators like cognitive tests and academic examinations for female graduates, and consequently affects their Salaries to understand the complete effect of Gender on Salaries.

II. Methods

Data & Sample

Aspiring Minds is an Indian employment assessment company that annually conducts the Aspiring Minds Computer Adaptive Test (AMCAT). AMCAT is one of the largest employability tests in India, taken by thousands of job seekers every year. In 2015, Aspiring Minds released the cross-sectional Aspiring Minds Employment Outcomes, 2015 (AMEO, 2015) dataset consisting of the AMCAT test results, biodata information and employment outcomes for 3700 engineering graduates that utilized the organization's services.

AMEO, 2015 has since been used to analyse factors that influence the job prospects of engineering graduates, as it brings together traditional parameters like student scores, college reputation, area of specialization, with unique data on job readiness through the results of the AMCAT. Gender being a parameter captured by this dataset, it is uniquely suited to conduct this study on gender bias in first salary amounts. By capturing crucial performance measures like 10th Board examination results, 12th Board examination results, College GPA and AMCAT

---

1 Girls constitute only 26% of total students at engg colleges. (2017, April 17). Retrieved from https://www.dnaindia.com/india/report-girls-constitute-only-26-of-total-students-at-engg-colleges-2404527

2 Gendering Human Development Indices. Ministry of Women and Child Development, India. (2009), http://www.undp.org/content/dam/india/docs/gendering_human_development_indices_summary_report.pdf

results, this dataset allows for insights on whether any performance gap exists between male and female graduates.

The dataset contains information on both postgraduate and undergraduate students. For the purpose of this study, only data on undergraduate students is considered, to remove the differences that may be caused by uncaptured professional or research experience amongst postgraduate students. The AMEO 2015 reports AMCAT performance in English, Logic ability, Quantitative ability, and Domain specific ability (depending upon the area of specialization of a student), along with scores on a personality assessment. Using backward and forward selection techniques, a few specific performance measures have been retained in the study, while others have been discarded. Personality assessment results were not found to be strongly significant in determining salary outcomes. This may be a reflection of the difficulty in gauging qualitative strengths while recruiting and assessing candidates, which may translate into over-emphasis on quantitative performance measures. The performance of graduates in the 10^{th} and 12^{th} Board examinations were found to be strongly correlated, with a correlation of 0.63. Further, both the 10^{th} and 12^{th} Board examination results were found to be correlated with the College GPA of a graduate. Keeping in mind that the College GPA, being the most recent indicator of academic ability, is most reported on resumes and has greatest influence on recruitment decisions, it has been retained in the model. The 10^{th} Board Examination and 12^{th} Board Examination results have been dropped from the model.

While the AMCAT data in the model was collected from the computer adaptive test, minimizing the chance or errors in data collection, significant biodata information regarding graduates was collected through a voluntary survey. This led to a high number of outliers being detected in the data which were removed before implementing the analysis.

Variables

- ***Dependent Variable*** – The dependent variable in the study is the Salary outcome that measures the annual cost-to-company (CTC) offered to an engineering undergraduate student upon graduation, as reported in the AMEO, 2015 dataset. This variable is operationalized as a continuous variable that must have a positive value, and is measured in Indian Rupees. For the purposes of this study, these figures have been converted into US dollar values at the exchange rate of $1 USD = Indian Rs.70/-. The minimum and maximum Salaries reported in the dataset are $500 and $25714.29 respectively.

- ***Independent* Variables** – The primary independent variable considered in the sample is the Gender of the engineering graduate. Only male and female genders are reported in the dataset. Gender has been operationalized as a categorical variable, taking the value 0 for males and 1 for females. To evaluate the indirect effects of gender on Salary, gender has also been modelled as a dependent variable against performance indicators that are introduced as exogenous controls in the larger study. Thus, gender is present as an endogenous variable in the model.

- ***Mediator Variables*** – Performance indicators like academic examination results and cognitive ability tests are present in the model as mediators, to understand the effect of gender on performance and consequently on Salaries.

As an academic performance indicator, the aggregate Grade Point Average (GPA) of a graduate in college is introduced as a mediator. The College GPA is operationalized as a continuous variable reported as a percentage value on a scale of 0 to 100, taking a minimum and maximum value of 6.45% and 98.4% respectively in the dataset.

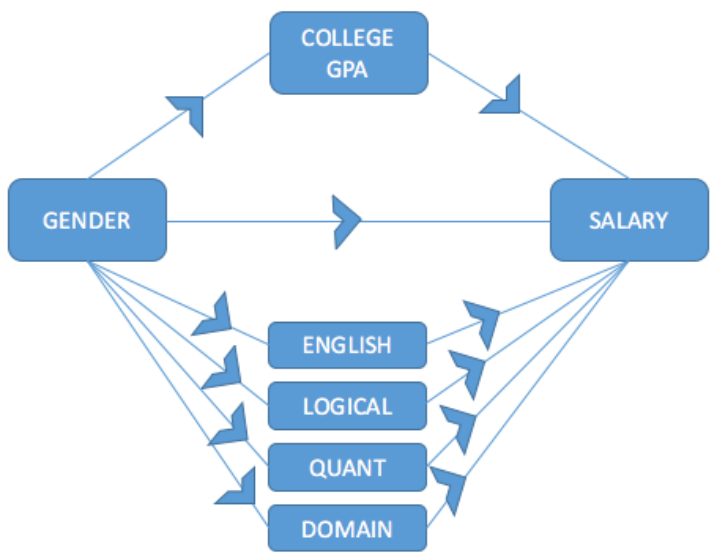
Other performance indicators introduced as mediators in the model are a graduate's scores on the AMCAT. English, Logical ability and Quantitative ability scores, that are cognitive ability measures, are operationalized as continuous variables reported on a scale of 100-900. Domain score is the score of the graduate, on a test specific to their area of specialization or domain. This is a technical proficiency measure and is operationalized as a continuous variable reported on a scale of 0 to 1 after being standardized across candidates.

- ***Control Variables*** – There are important variables considered in the study as controls. The tier of a graduate's college is introduced as a control. College Tier is operationalized as a categorical variable that takes the value of 1 for Tier 1 colleges (that are considered superior for producing graduates with better skills), and which takes a value of 2 for Tier 2 colleges. The year of graduation is also introduced as a control, since Salaries were seen to show a decreasing trend when plotted against graduation year. The years from 2009 to 2017 are included as a categorical variable in the model. Finally, the program Specialization of the graduate is also introduced as a control and is operationalized as a categorical variable.

Analysis

The analysis was conducted using Ordinary Least-Squares Regression (OLS) with the Statistical Analysis tool R. As a first step, assumption diagnostics and corrections were carried out in order to ensure that assumptions for OLS regression were satisfied. Next, to estimate the effects of Gender on the first Salary paid to engineering graduates, a path analysis model was conducted. The analysis attempts to understand the total effect of gender on Salary, by accounting for the indirect effect of Gender (if any) via performance indicators like academic achievement, cognitive ability and technical ability. Figure 1 depicts the path model that was used for the analysis.

Figure 1: Path Analysis Model – Total Effect of Gender on Salary



The complete regression model used in this analysis is represented by the equation:

Salary = B0 + B1* Gender + B2*College GPA + B3*English Score + B4*Logical Score + B5*Quant Score + B6*Domain Score + B7* College Tier + B8*Graduation Year + B9*Specialization + error

III. Results

<u>Descriptive Statistics</u>

Sample descriptive statistics for all the continuous variables included in our study, are laid out in Table 1. For categorical variables like Graduation Year and Specialization, descriptive statistics have not been included in the table for ease of reporting. Within our sample, the average Salary was found to be $4183.744 with a standard deviation of $1882.393. Within the sample, the dependent variable measuring Salary was present in all observations.

Examining our independent variable, Gender, 75.7% of the sample consisted of male graduates, with only 24.3% of the sample comprising female graduates. This, however, is not a cause for worry as the sample gender distribution is in line with the population gender distribution of engineering graduates. Data on gender was available for the complete sample, with no missing data.

Amongst the mediators, the average College GPA was found to be 71.454%, with a standard deviation of 7.2425% and 0% missingness. Examining measures of cognitive ability, the average English score for the sample was 505.08 points with a standard deviation of 104.47 points. The average Logical score for the sample was 503.06 points with a standard deviation of 85.49 points. The average Quant score for the sample was 515.76 points, with a standard deviation of 121.84 points. Data was complete for each of these measures. When examined by gender, females were found to have a lower average Quant score than males, with a difference of 28.87 points. No such difference in average scores by gender was found for English and Logical ability scores. Finally, the technical ability measure Domain score was found to have an average value of 0.6030 points, with a standard deviation of 0.2735 points. 6.22% of the sample opted out from giving the Domain specific test, leading to missing data under this variable head.

Amongst the control variables, descriptive statistics were as follows. College Tier is an artificially constructed variable that uses the average AMCAT scores of students at different colleges, to categorize them as Tier 1 or Tier 2. 93.36% of the sample consists of observations from Tier 2 college graduates, with only 6.64% consisting of Tier 1 graduates. Graduation Year data was present for 9 years ranging from 2009 to 2017. The sample also covered graduates from across 18 different engineering specializations. Graduation Year and Specialization were introduced in the model as controls. Due to the number of categories included under each, for ease of reading and legibility, they are not reported in detail in this paper. (Table 1)

Table 1: Descriptive Statistics for Continuous Variables

Variable	Mean	Std. Deviation	Missingness

Dependent Variable			
Salary	$4183.744	$1882.393	0%
Independent Variables			
Gender	0.243	0.184	0%
Mediator Variables			
College GPA	71.454	7.2425	0%
English Score	505.08	104.47	0%
Logical Score	503.06	85.49	0%
Quant Score	515.76	121.84	0%
Domain Score	0.6030	0.2735	6.22%

Sample Size – 3,700 observations

Regression Results

I. Assumption Diagnostics and Corrections

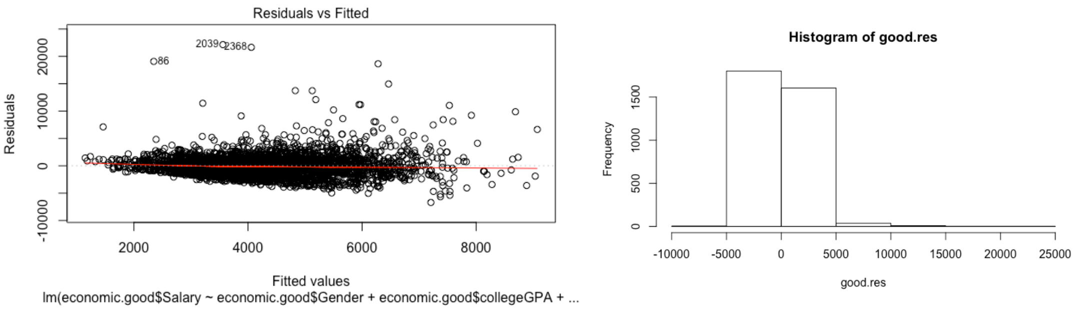
Before carrying out the Regression analysis, assumption diagnostics for Ordinary Least Squares Regression was carried out. Based on results, the simple linear regression model was appropriately re-specified.

The mediator variables College GPA, Logical Score, Quant Score and Domain Score were all found to have non-linear quadratic relationships with the dependent variable, Salary. In order to account for this non-linear relationship, the variables were included along with an additional squared term each, so that it could account for the non-linear effects of the variable on Salary.

Multicollinearity was checked for, by examining the Variance Inflation Factor for each independent, mediator and control variable in the model. No multicollinearity was detected amongst the final variables included, after backward and forward selection was already carried out.

The model was then diagnosed for homoscedasticity and normality of residuals. Figure 2 shows that the model did not satisfy either of the assumptions, with a high number of outliers seeming to cause the effect.

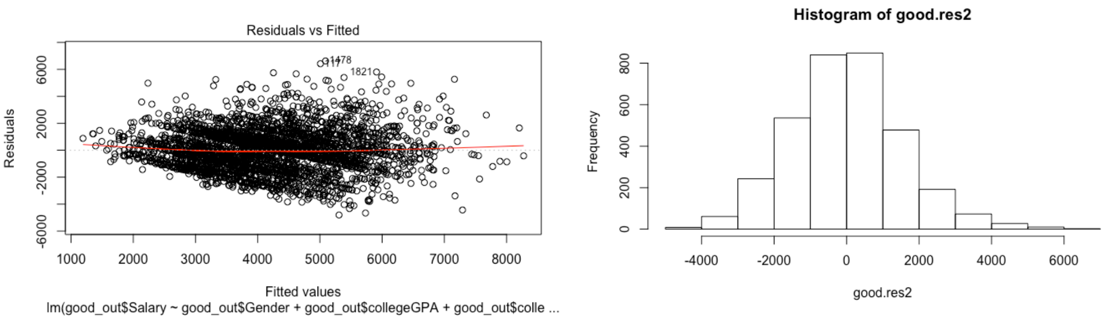
Figure 2: Homoscedasticity and Normality of Residuals: Before Re-specification



The model was then diagnosed for outliers by investigating the influence of observations using the measure Cook's D. 161 observations were omitted from the re-specified model due to their high degree of influence that was limiting the normality of the data. This reduced the size of the sample to 3539 observations, still a satisfactory number of observations to continue with the analysis.

After the omission of outliers and appropriate transformation of non-linear variables, the model was re-diagnosed for homoscedasticity and normality of residuals. Figure 3 shows that the corrected model satisfied both assumptions.

Figure 3: Homoscedasticity and Normality of Residuals: After Re-specification



Data for the variable Domain Score was found to be missing in case of 6.22% of the sample. However, this data was not missing at random (MAR) or missing completely at random (MCAR). Instead, certain individuals chose to opt out of giving the Domain specific technical test and so their scores were unavailable. When imputation of scores for this variable was attempted, it added noise to the dataset, and once again the model did not satisfy the OLS assumptions. Since the missing values did not satisfy either the MAR or MCAR conditions, and the Domain score was not the primary independent variable under consideration, the incomplete data was not included further in the analysis and list-wise deletion was adopted.

II. <u>Model Results</u>

Regression results are available for three models considered in the analysis. Model 1 examines the full sample, whereas Model 2 and 3 exclude Gender as the independent variable and look at the effects of other variables on Salary for the male subset and female subset of the sample, respectively.

Table 2: Regression Results

Variable	Model 1	Model 2	Model 3
Constant	-4162 (267.5)	-2901 (290.2)	-5956 (470.8)
Independent Variables			
Gender	-271* (65.77)	-	-
Mediator Variables			
College GPA	125.8* (53.13)	78.39 (60.58)	240* (122.7)
College GPA (Squared)	-0.472 (0.367)	-0.115 (0.4213)	-1.35 (0.83)
English Score	3.063* (0.2981)	3.082* (0.344)	2.81* (0.606)
Logical Score	-7.514* (2.925)	-6.626* (3.32)	-10.11 (6.479)
Logical Score (Squared)	0.0086* (0.0029)	0.0075* (0.0033)	0.012* (0.006)
Quant Score	3.533* (1.429)	4.499* (1.653)	0.5516 (3.079)
Quant Score (Squared)	-0.0007 (0.0014)	-0.0012 (0.0015)	0.0023 (0.0031)
Domain Score	-563.3 (443.6)	-654.8 (516.2)	-187.9 (882)
Domain Score (Squared)	884.9* (390.6)	1038* (453.1)	347.6 (782)
Control Variables			
College Tier 2	-559.2* (114.8)	-589.2* (127.9)	-339.5 (273)
Graduation Year	###	###	###

Specialization (Computer Application)	-3678* (1744)	-	-2776 (1737)
Observations	3319	2506	813
Residual Std. Error	1551	1579	1470
Adjusted R^2	0.3219	0.3315	0.2715

Std. Errors are included in parenthesis
*p < .05

- Results found to not be statistically significant. For ease of reporting, results have been shown in the appendix as part of output itself.

a. Model 1

Model 1 reveals the regression results for the full sample. It shows that Gender, College GPA, English Score, Logical Score and Quant Score are all significant variables in the model. All other variables are found to not be statistically significant. The model is seen to have an adjusted R^2 of 0.3219, which means that it explains approximately 32.19% of the variation in the dependent variable, Salary.

Gender is found to be significant with the coefficient suggesting that female graduates receive, on average, an annual salary of $271 less than male graduates. This reveals the direct effect of Gender on Salary and at first glance seems to support the assumption of there being gender bias in salaries offered. The indirect effects of Gender were also examined, by looking at the effect of Gender on performance indicators like College GPA, English Score, Logical Score, Quant Score and Domain Score. The Indirect Effect was found to be significant only via the mediating variables College GPA and Quant Score. Table 3 reveals the results of the indirect effects found. Gender is mediated via College GPA such that the effect of GPA on Salary for females results in an, on average, higher salary by $418.64. However, the mediating effect of Quantitative scores is negative such that the Salary for females is, on average, lower by $102.02 via their Quant Score. This results in a direct effect of Gender on Salary of females earning, on average, $45.62 higher than that for males. The perceived gender bias of the full model that did not account for mediating effects of gender, disappears when these effects are included.

Table 2: Indirect Effects of Gender on Salary

Mediating Variable	**Indirect Effect**
College GPA	418.64*
English Score	-0.269

Logical Score	20.70
Quant Score	-102.02*
Domain Score	7.319

----

*p < .05

Looking at other significant variables in the model we see that the College Tier is an important indicator affecting Salary. Tier 2 college graduates receive, on average, a salary of $559.2 less than graduates from Tier 1 colleges. College GPA is a significant variable but its non-linear effect is not significant. Thus, a unit increase in GPA is found to result in, on average, an increase in Salary by $125.8. A unit increase in English Score is found to result in, on average, a small increase in Salary by $3.063. A unit increase in Quant Score is found to result in, on average, a small increase in Salary by $3.533. A unit increase in Logical Score is found to result in, on average, a small decrease in Salary by $7.5. However, the non-linear effect of Logical Score means that with every unit increase in the Logical Score, its negative effect on Salary is reduced by $0.0086.

While Model 1 examines the entire sample, Model 2 and 3 look at the male and female subsets of the sample. This provides some insight on what factors affect the Salary received by the two genders. Model 2 shows that English score, Logical Score and Quant score and College Tier are the only significant variables affecting Salary for males. The model explains 33.15% of the variation in Salary for males. The model reveals that the effect of English scores and Quant scores on Salary is more pronounced for Males than for the overall sample results. College GPA does not come out to be a significant factor affecting Salary for males. The negative effect of the Logical score on Salary is less pronounced for males than the overall sample, while the negative effect of tier 2 colleges on Salary is more pronounced for males than the overall sample.

Model 3 shows that English score and College GPA are the only significant variables affecting Salary for females. The model explains 27.15% of the variation in Salary for females. The model reveals that the effect of English scores on Salary is less pronounced for females than for the overall sample results. However, the college GPA seems to be an extremely significant factor for female Salaries with a unit increase in GPA leading to an, on average, increase of $240 in Salary for females. Interestingly, the College Tier does not seem to be a significant factor in determining female Salaries.

The results seem to suggest while there does not seem to be enough evidence of gender bias in salaries, female graduates seem to be valued more based on their academic performance in college, reflected in the College GPA, while males seem to be assessed based on their cognitive ability reflected through Logical and Quant scores. English scores remain important across the spectrum, which is to be expected since Indian commerce and industry is heavily dependent on the country’s English-speaking labor force, with most commercial activities being carried out in English.

Conclusion

The study made use of the AMEO 2015 dataset to answer the question of whether there exists gender bias in the first salary amounts paid to fresh engineering graduates in India. The analysis shows that after accounting for both the direct and indirect effects of gender on Salary, no gender bias is detected in the data. The study also provides insight on other important variables in determining Salary outcome.

All three Models considered in the studyhave very large residual standard errors and low effect sizes in terms of adjusted R^2. This means that the sample and chosen models do not adequately explain the variation in the dependent variable, Salary. In the case of social and human phenomenon like recruitment and salaries, there are umpteen factors that are at play to determine an outcome. Thus, several important variables may have been excluded from the models considered in this study, since they are unavailable in the sample. Given the particular focus of this study to ascertain whether Gender bias exists, and an effort to retain simplicity in the model, a particular set of variables was chosen for the analysis. Thus, this study is simply a first step in the exploration of this research question. After the collection of better quality data, and the examining of different models, the question of gender bias in salaries may be more convincingly answered.

Appendix

```
> ### Policy Report 4 - Engineering Students: Gender Bias in First Salary ###
>
> #Set of libraries used throughout
> library(readxl) #import data
> library(lmSupport)
> library(ggplot2)
> library(lme4)
> library(lattice)
>
> #Clearing Working Environment
> rm(list = ls())
>
> #Setting Working Directory
> setwd("/Users/shwetachopra/OneDrive - PennO365/Applied Linear Modeling/PR4")
>
> #First Step is to Import the Dataset
> good <- read_excel(("train.xlsx"))
>
>
> #Extract data subset with relevant variables
> economic.good <- cbind(good$ID,
+                        good$Salary,
+                        good$Gender,
+                        good$g10percentage,
+                        good$g12percentage,
+                        good$Degree,
+                        good$Specialization,
+                        good$collegeGPA,
+                        good$CollegeTier,
+                        good$CollegeState,
+                        good$GraduationYear,
+                        good$English,
+                        good$Logical,
+                        good$Quant,
+                        good$Domain)
>
> #Set column names
> colnames(economic.good) <- c("ID", "Salary", "Gender", "g10percentage", "g12percentage",
+                              "Degree", "Specialization", "collegeGPA",
+                              "CollegeTier", "CollegeState", "GraduationYear",
+                              "English", "Logical", "Quant", "Domain")
> economic.good <- as.data.frame(economic.good)
> View(economic.good)
>
> #Clean data
> economic.good <- subset(economic.good, Degree == "B.Tech/B.E.")
```

```
> economic.good$Gender <- ifelse(economic.good$Gender == "m", 0, ifelse(economic.good$Gender ==
"f", 1, "NA"))
> economic.good$Domain <- ifelse(economic.good$Domain == -1, NA,
as.character(economic.good$Domain))
>
>
> #Fix data types
> class(economic.good$Salary)
[1] "factor"
> class(economic.good$Domain)
[1] "character"
> economic.good$Salary <- as.numeric(as.character(economic.good$Salary))
> economic.good$Gender <- as.factor(economic.good$Gender)
> economic.good$g10percentage <- as.numeric(as.character(economic.good$g10percentage))
> economic.good$g12percentage <- as.numeric(as.character(economic.good$g12percentage))
> economic.good$collegeGPA <- as.numeric(as.character(economic.good$collegeGPA))
> economic.good$English <- as.numeric(as.character(economic.good$English))
> economic.good$Logical <- as.numeric(as.character(economic.good$Logical))
> economic.good$Quant <- as.numeric(as.character(economic.good$Quant))
> economic.good$Domain <- as.numeric(economic.good$Domain)
>
> #Convert Salary to dollar values at - $1 = Rs.70/- AND REMOVE OUTLIER SALARIES
> economic.good <- subset(economic.good, Salary < 2000000)
> economic.good$Salary <- (economic.good$Salary)/70
> economic.good <- subset(economic.good, as.character(GraduationYear) != 0 ||
as.character(GraduationYear) !=  2007)
>
> #Remove wrong GPA figures
> economic.good <- subset(economic.good, collegeGPA > 10)
> ##PRELIMINARY ASSUMPTION DIAGNOSTICS
> ##Test for homoscedasticity in Simple Linear Model - BEFORE OUTLIER REMOVAL
> lm_eng <- lm(economic.good$Salary ~ economic.good$Gender + economic.good$collegeGPA
+            + factor(economic.good$CollegeTier) + factor(economic.good$GraduationYear) +
economic.good$English +
+             economic.good$Logical + economic.good$Quant + economic.good$Domain +
factor(economic.good$Specialization))
> plot(lm_eng) #Check residuals vs fitted values
Hit <Return> to see next plot:
Hit <Return> to see next plot:
Hit <Return> to see next plot: #
Hit <Return> to see next plot: #
Warning messages:
1: not plotting observations with leverage one:
  499, 1499, 1562, 2310
2: not plotting observations with leverage one:
  499, 1499, 1562, 2310
3: In sqrt(crit * p * (1 - hh)/hh) : NaNs produced
4: In sqrt(crit * p * (1 - hh)/hh) : NaNs produced
> #
```

```
> #
>
> ##Test for Normality of Residuals in Simple Linear Model 1 - BEFORE CORRECTIONS
> lm2 <- lm(economic.good$Salary ~ economic.good$Gender + economic.good$collegeGPA
+         + factor(economic.good$CollegeTier) + factor(economic.good$GraduationYear) +
economic.good$English +
+           economic.good$Logical + economic.good$Quant + economic.good$Domain +
factor(economic.good$Specialization))
> good.res <- resid(lm2)
> hist(good.res,10)
>
> ##Check for Multicollinearity - BEFORE CORRECTIONS
> lm_multi <- lm(economic.good$Salary ~ economic.good$Gender + economic.good$collegeGPA
+             + factor(economic.good$CollegeTier) + factor(economic.good$GraduationYear) +
economic.good$English +
+               economic.good$Logical + economic.good$Quant + economic.good$Domain +
factor(economic.good$Specialization))
> library(car)
> vif(lm_multi)
                                      GVIF Df GVIF^(1/(2*Df))
economic.good$Gender                1.099275  1        1.048463
economic.good$collegeGPA            1.201358  1        1.096065
factor(economic.good$CollegeTier)   1.110056  1        1.053592
factor(economic.good$GraduationYear) 1.474765  9       1.021818
economic.good$English               1.346502  1        1.160389
economic.good$Logical               1.560479  1        1.249191
economic.good$Quant                 1.614541  1        1.270646
economic.good$Domain                1.291016  1        1.136229
factor(economic.good$Specialization) 1.634956 30       1.008227
> ## Analysis of Outliers
> ##------------------------------##
> #Outlier information for model 1, stored in new data frame
> outliers <- na.omit(economic.good) #copy data regarding outliers
>
> lm_out <- lm(outliers$Salary ~ outliers$Gender + outliers$collegeGPA
+           + factor(outliers$CollegeTier) + factor(outliers$GraduationYear) + outliers$English +
+             outliers$Logical + outliers$Quant + outliers$Domain + factor(outliers$Specialization))
>
> outliers$cd <- cooks.distance(lm_out)
> ##Influence - Cook's D
> large_cd <- subset(outliers, cd > (4/3699))
> View(large_cd)
>
> library(Hmisc)

Attaching package: 'Hmisc'

The following objects are masked from 'package:base':
```

```
   format.pval, units

> describe(large_cd$cd)
large_cd$cd
      n  missing distinct     Info     Mean      Gmd      .05      .10      .25
    141        0      141        1 0.006302 0.007807 0.001156 0.001180 0.001531
    .50      .75      .90      .95
0.002696 0.006098 0.012471 0.016330

lowest : 0.001085402 0.001095395 0.001103353 0.001112956 0.001117090
highest: 0.045305503 0.048786101 0.070035027 0.070035027 0.088430547
> hist(large_cd$cd)
> quantile(large_cd$cd, probs = seq(0, 1, 0.05))
         0%          5%         10%         15%         20%         25%         30%
0.001085402 0.001156465 0.001179742 0.001278570 0.001366729 0.001531227 0.001673976
        35%         40%         45%         50%         55%         60%         65%
0.001823637 0.002022186 0.002375355 0.002695574 0.002904076 0.003069225 0.003455449
        70%         75%         80%         85%         90%         95%        100%
0.004540413 0.006098134 0.006797477 0.008323180 0.012470825 0.016329573 0.088430547
>
> ##Mark outliers
> good_out <- merge(economic.good, outliers[,c("ID", "cd")], by = "ID", all.x = TRUE)
> good_out <- subset(good_out, is.na(cd) | cd < 4/3699)
#Extract descriptive statistics
>summary(good_out)
> ##Test for Homoscedasticity and Normality of Residuals in Simple Linear Model after removing outliers
> lm3 <- lm(good_out$Salary ~ good_out$Gender + good_out$collegeGPA
+         + factor(good_out$CollegeTier) + good_out$English +
+           good_out$Logical + good_out$Quant + good_out$Domain + factor(good_out$GraduationYear)
+
+           factor(good_out$Specialization))
> good.res <- resid(lm3)
> hist(good.res,10)
> plot(lm3)
Hit <Return> to see next plot:
Hit <Return> to see next plot:
Hit <Return> to see next plot:
Hit <Return> to see next plot:
Warning messages:
1: not plotting observations with leverage one:
  160, 964, 1521, 2634, 2706, 2749
2: not plotting observations with leverage one:
  160, 964, 1521, 2634, 2706, 2749
3: In sqrt(crit * p * (1 - hh)/hh) : NaNs produced
4: In sqrt(crit * p * (1 - hh)/hh) : NaNs produced
> #Fix non-linear transformations
> good_out$collegeGPA2 <- good_out$collegeGPA ^ 2
> good_out$Quant2 <- good_out$Quant ^ 2
> good_out$Logical2 <- good_out$Logical ^ 2
```

```
> good_out$Domain2 <- good_out$Domain ^ 2
> lm4 <- lm(good_out$Salary ~ good_out$Gender + good_out$collegeGPA +
+           good_out$collegeGPA2 + good_out$Quant2 + factor(good_out$CollegeTier) +
good_out$English +
+           good_out$Logical + good_out$Logical2 + good_out$Quant + good_out$Domain +
good_out$Domain2 +
+           factor(good_out$GraduationYear) +
+           factor(good_out$Specialization))
> summary(lm4)

Call:
lm(formula = good_out$Salary ~ good_out$Gender + good_out$collegeGPA +
    good_out$collegeGPA2 + good_out$Quant2 + factor(good_out$CollegeTier) +
    good_out$English + good_out$Logical + good_out$Logical2 +
    good_out$Quant + good_out$Domain + good_out$Domain2 + factor(good_out$GraduationYear) +
    factor(good_out$Specialization))

Residuals:
    Min      1Q  Median      3Q     Max
-4806.8 -1026.4   -22.7   936.6  6624.5

Coefficients:
                                                              Estimate
(Intercept)                                                 -4.162e+03
good_out$Gender1                                            -2.710e+02
good_out$collegeGPA                                          1.258e+02
good_out$collegeGPA2                                        -4.720e-01
good_out$Quant2                                             -7.120e-04
factor(good_out$CollegeTier)2                               -5.592e+02
good_out$English                                             3.063e+00
good_out$Logical                                            -7.514e+00
good_out$Logical2                                            8.619e-03
good_out$Quant                                               3.533e+00
good_out$Domain                                             -5.633e+02
good_out$Domain2                                             8.849e+02
factor(good_out$GraduationYear)2009                          2.053e+03
factor(good_out$GraduationYear)2010                          2.325e+03
factor(good_out$GraduationYear)2011                          1.665e+03
factor(good_out$GraduationYear)2012                          9.192e+02
factor(good_out$GraduationYear)2013                          3.297e+02
factor(good_out$GraduationYear)2014                          2.370e+02
factor(good_out$GraduationYear)2015                          5.552e+02
factor(good_out$GraduationYear)2016                         -1.960e+01
factor(good_out$GraduationYear)2017                         -4.913e+02
factor(good_out$Specialization)automobile/automotive engineering  -1.054e+03
factor(good_out$Specialization)biotechnology                -6.308e+02
factor(good_out$Specialization)civil engineering             5.931e+02
factor(good_out$Specialization)computer application         -3.678e+03
factor(good_out$Specialization)computer engineering          2.341e+02
```

factor(good_out$Specialization)computer science & engineering	-9.291e+01
factor(good_out$Specialization)computer science and technology	-6.721e+00
factor(good_out$Specialization)electrical engineering	-4.029e+02
factor(good_out$Specialization)electronics & instrumentation eng	-2.015e+02
factor(good_out$Specialization)electronics & telecommunications	1.037e+02
factor(good_out$Specialization)electronics and communication engineering	-1.496e+02
factor(good_out$Specialization)electronics and computer engineering	-1.740e+02
factor(good_out$Specialization)electronics and electrical engineering	-3.986e+02
factor(good_out$Specialization)electronics and instrumentation engineering	-1.514e+02
factor(good_out$Specialization)electronics engineering	8.490e+01
factor(good_out$Specialization)industrial & management engineering	4.258e+02
factor(good_out$Specialization)industrial & production engineering	1.220e+01
factor(good_out$Specialization)industrial engineering	1.566e+03
factor(good_out$Specialization)information & communication technology	9.281e+02
factor(good_out$Specialization)information science engineering	2.412e+02
factor(good_out$Specialization)information technology	6.966e+01
factor(good_out$Specialization)instrumentation and control engineering	7.508e+02
factor(good_out$Specialization)instrumentation engineering	4.331e+01
factor(good_out$Specialization)mechanical and automation	-1.034e+03
factor(good_out$Specialization)mechanical engineering	-1.103e+02
factor(good_out$Specialization)telecommunication engineering	4.166e+02

	Std. Error
(Intercept)	2.675e+03
good_out$Gender1	6.577e+01
good_out$collegeGPA	5.313e+01
good_out$collegeGPA2	3.673e-01
good_out$Quant2	1.368e-03
factor(good_out$CollegeTier)2	1.148e+02
good_out$English	2.981e-01
good_out$Logical	2.925e+00
good_out$Logical2	2.930e-03
good_out$Quant	1.429e+00
good_out$Domain	4.436e+02
good_out$Domain2	3.906e+02
factor(good_out$GraduationYear)2009	1.667e+03
factor(good_out$GraduationYear)2010	1.562e+03
factor(good_out$GraduationYear)2011	1.560e+03
factor(good_out$GraduationYear)2012	1.559e+03
factor(good_out$GraduationYear)2013	1.558e+03
factor(good_out$GraduationYear)2014	1.558e+03
factor(good_out$GraduationYear)2015	1.571e+03
factor(good_out$GraduationYear)2016	1.797e+03
factor(good_out$GraduationYear)2017	1.708e+03
factor(good_out$Specialization)automobile/automotive engineering	1.736e+03
factor(good_out$Specialization)biotechnology	1.102e+03
factor(good_out$Specialization)civil engineering	8.431e+02
factor(good_out$Specialization)computer application	1.744e+03
factor(good_out$Specialization)computer engineering	7.826e+02
factor(good_out$Specialization)computer science & engineering	7.804e+02

```
factor(good_out$Specialization)computer science and technology           1.351e+03
factor(good_out$Specialization)electrical engineering                    7.997e+02
factor(good_out$Specialization)electronics & instrumentation eng         8.437e+02
factor(good_out$Specialization)electronics & telecommunications          7.929e+02
factor(good_out$Specialization)electronics and communication engineering   7.797e+02
factor(good_out$Specialization)electronics and computer engineering       1.738e+03
factor(good_out$Specialization)electronics and electrical engineering     7.865e+02
factor(good_out$Specialization)electronics and instrumentation engineering  8.566e+02
factor(good_out$Specialization)electronics engineering                   9.517e+02
factor(good_out$Specialization)industrial & management engineering        1.738e+03
factor(good_out$Specialization)industrial & production engineering        1.191e+03
factor(good_out$Specialization)industrial engineering                    1.739e+03
factor(good_out$Specialization)information & communication technology      1.347e+03
factor(good_out$Specialization)information science engineering           8.383e+02
factor(good_out$Specialization)information technology                    7.812e+02
factor(good_out$Specialization)instrumentation and control engineering     9.200e+02
factor(good_out$Specialization)instrumentation engineering               1.346e+03
factor(good_out$Specialization)mechanical and automation                 1.351e+03
factor(good_out$Specialization)mechanical engineering                    7.867e+02
factor(good_out$Specialization)telecommunication engineering             1.735e+03
                                                          t value
(Intercept)                                                -1.556
good_out$Gender1                                           -4.120
good_out$collegeGPA                                         2.367
good_out$collegeGPA2                                       -1.285
good_out$Quant2                                            -0.520
factor(good_out$CollegeTier)2                              -4.871
good_out$English                                           10.276
good_out$Logical                                           -2.568
good_out$Logical2                                           2.942
good_out$Quant                                              2.471
good_out$Domain                                            -1.270
good_out$Domain2                                            2.266
factor(good_out$GraduationYear)2009                         1.231
factor(good_out$GraduationYear)2010                         1.488
factor(good_out$GraduationYear)2011                         1.067
factor(good_out$GraduationYear)2012                         0.590
factor(good_out$GraduationYear)2013                         0.212
factor(good_out$GraduationYear)2014                         0.152
factor(good_out$GraduationYear)2015                         0.353
factor(good_out$GraduationYear)2016                        -0.011
factor(good_out$GraduationYear)2017                        -0.288
factor(good_out$Specialization)automobile/automotive engineering          -0.607
factor(good_out$Specialization)biotechnology                             -0.572
factor(good_out$Specialization)civil engineering                          0.703
factor(good_out$Specialization)computer application                      -2.109
factor(good_out$Specialization)computer engineering                       0.299
factor(good_out$Specialization)computer science & engineering             -0.119
factor(good_out$Specialization)computer science and technology            -0.005
```

```
factor(good_out$Specialization)electrical engineering                   -0.504
factor(good_out$Specialization)electronics & instrumentation eng           -0.239
factor(good_out$Specialization)electronics & telecommunications            0.131
factor(good_out$Specialization)electronics and communication engineering    -0.192
factor(good_out$Specialization)electronics and computer engineering        -0.100
factor(good_out$Specialization)electronics and electrical engineering      -0.507
factor(good_out$Specialization)electronics and instrumentation engineering  -0.177
factor(good_out$Specialization)electronics engineering                    0.089
factor(good_out$Specialization)industrial & management engineering          0.245
factor(good_out$Specialization)industrial & production engineering         0.010
factor(good_out$Specialization)industrial engineering                    0.900
factor(good_out$Specialization)information & communication technology       0.689
factor(good_out$Specialization)information science engineering             0.288
factor(good_out$Specialization)information technology                    0.089
factor(good_out$Specialization)instrumentation and control engineering      0.816
factor(good_out$Specialization)instrumentation engineering                0.032
factor(good_out$Specialization)mechanical and automation                 -0.766
factor(good_out$Specialization)mechanical engineering                   -0.140
factor(good_out$Specialization)telecommunication engineering               0.240
                                                           Pr(>|t|)
(Intercept)                                                 0.11977
good_out$Gender1                                            3.88e-05
good_out$collegeGPA                                          0.01798
good_out$collegeGPA2                                         0.19887
good_out$Quant2                                             0.60291
factor(good_out$CollegeTier)2                                 1.16e-06
good_out$English                                           < 2e-16
good_out$Logical                                            0.01026
good_out$Logical2                                           0.00329
good_out$Quant                                              0.01351
good_out$Domain                                             0.20429
good_out$Domain2                                            0.02354
factor(good_out$GraduationYear)2009                           0.21828
factor(good_out$GraduationYear)2010                           0.13679
factor(good_out$GraduationYear)2011                           0.28591
factor(good_out$GraduationYear)2012                           0.55550
factor(good_out$GraduationYear)2013                           0.83245
factor(good_out$GraduationYear)2014                           0.87915
factor(good_out$GraduationYear)2015                           0.72374
factor(good_out$GraduationYear)2016                           0.99130
factor(good_out$GraduationYear)2017                           0.77367
factor(good_out$Specialization)automobile/automotive engineering          0.54382
factor(good_out$Specialization)biotechnology                          0.56713
factor(good_out$Specialization)civil engineering                       0.48182
factor(good_out$Specialization)computer application                    0.03506
factor(good_out$Specialization)computer engineering                    0.76484
factor(good_out$Specialization)computer science & engineering             0.90525
factor(good_out$Specialization)computer science and technology            0.99603
factor(good_out$Specialization)electrical engineering                   0.61444
```

factor(good_out$Specialization)electronics & instrumentation eng 0.81128
factor(good_out$Specialization)electronics & telecommunications 0.89597
factor(good_out$Specialization)electronics and communication engineering 0.84785
factor(good_out$Specialization)electronics and computer engineering 0.92025
factor(good_out$Specialization)electronics and electrical engineering 0.61232
factor(good_out$Specialization)electronics and instrumentation engineering 0.85970
factor(good_out$Specialization)electronics engineering 0.92892
factor(good_out$Specialization)industrial & management engineering 0.80645
factor(good_out$Specialization)industrial & production engineering 0.99183
factor(good_out$Specialization)industrial engineering 0.36796
factor(good_out$Specialization)information & communication technology 0.49082
factor(good_out$Specialization)information science engineering 0.77359
factor(good_out$Specialization)information technology 0.92894
factor(good_out$Specialization)instrumentation and control engineering 0.41452
factor(good_out$Specialization)instrumentation engineering 0.97433
factor(good_out$Specialization)mechanical and automation 0.44388
factor(good_out$Specialization)mechanical engineering 0.88852
factor(good_out$Specialization)telecommunication engineering 0.81023

(Intercept)
good_out$Gender1 ***
good_out$collegeGPA *
good_out$collegeGPA2
good_out$Quant2
factor(good_out$CollegeTier)2 ***
good_out$English ***
good_out$Logical *
good_out$Logical2 **
good_out$Quant *
good_out$Domain
good_out$Domain2 *
factor(good_out$GraduationYear)2009
factor(good_out$GraduationYear)2010
factor(good_out$GraduationYear)2011
factor(good_out$GraduationYear)2012
factor(good_out$GraduationYear)2013
factor(good_out$GraduationYear)2014
factor(good_out$GraduationYear)2015
factor(good_out$GraduationYear)2016
factor(good_out$GraduationYear)2017
factor(good_out$Specialization)automobile/automotive engineering
factor(good_out$Specialization)biotechnology
factor(good_out$Specialization)civil engineering
factor(good_out$Specialization)computer application *
factor(good_out$Specialization)computer engineering
factor(good_out$Specialization)computer science & engineering
factor(good_out$Specialization)computer science and technology
factor(good_out$Specialization)electrical engineering
factor(good_out$Specialization)electronics & instrumentation eng

```
factor(good_out$Specialization)electronics & telecommunications
factor(good_out$Specialization)electronics and communication engineering
factor(good_out$Specialization)electronics and computer engineering
factor(good_out$Specialization)electronics and electrical engineering
factor(good_out$Specialization)electronics and instrumentation engineering
factor(good_out$Specialization)electronics engineering
factor(good_out$Specialization)industrial & management engineering
factor(good_out$Specialization)industrial & production engineering
factor(good_out$Specialization)industrial engineering
factor(good_out$Specialization)information & communication technology
factor(good_out$Specialization)information science engineering
factor(good_out$Specialization)information technology
factor(good_out$Specialization)instrumentation and control engineering
factor(good_out$Specialization)instrumentation engineering
factor(good_out$Specialization)mechanical and automation
factor(good_out$Specialization)mechanical engineering
factor(good_out$Specialization)telecommunication engineering
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1551 on 3272 degrees of freedom
  (220 observations deleted due to missingness)
Multiple R-squared:  0.3313,  Adjusted R-squared:  0.3219
F-statistic: 35.24 on 46 and 3272 DF,  p-value: < 2.2e-16
> #Indirect effects
> lmi1 <- lm(good_out$collegeGPA ~ good_out$Gender)
> summary(lmi1)

Call:
lm(formula = good_out$collegeGPA ~ good_out$Gender)

Residuals:
    Min      1Q  Median      3Q     Max
-21.6146  -4.8481  -0.0846   4.5019  27.7154

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      70.6846     0.1375  514.21   <2e-16 ***
good_out$Gender1   3.1669     0.2789   11.36   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.115 on 3537 degrees of freedom
Multiple R-squared:  0.03518,	Adjusted R-squared:  0.03491
F-statistic:   129 on 1 and 3537 DF,  p-value: < 2.2e-16

>
> lmi3 <- lm(good_out$English ~ good_out$Gender)
> summary(lmi3)
```

```
Call:
lm(formula = good_out$English ~ good_out$Gender)

Residuals:
    Min      1Q  Median      3Q     Max 
-300.10  -75.10   -5.01   69.90  369.90 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)      505.09966    2.01868 250.212   <2e-16 ***
good_out$Gender1  -0.08804    4.09505  -0.021    0.983    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 104.5 on 3537 degrees of freedom
Multiple R-squared:  1.307e-07,	Adjusted R-squared:  -0.0002826 
F-statistic: 0.0004622 on 1 and 3537 DF,  p-value: 0.9828

> 
> lmi4 <- lm(good_out$Logical ~ good_out$Gender)
> summary(lmi4)

Call:
lm(formula = good_out$Logical ~ good_out$Gender)

Residuals:
     Min       1Q   Median       3Q      Max 
-298.726  -58.726    1.274   61.274  291.274 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)       503.726      1.652 304.969   <2e-16 ***
good_out$Gender1   -2.755      3.351  -0.822    0.411    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 85.49 on 3537 degrees of freedom
Multiple R-squared:  0.0001911,	Adjusted R-squared:  -9.161e-05 
F-statistic: 0.6759 on 1 and 3537 DF,  p-value: 0.4111

> 
> lmi5 <- lm(good_out$Quant ~ good_out$Gender)
> summary(lmi5)

Call:
lm(formula = good_out$Quant ~ good_out$Gender)

Residuals:
```

```
    Min      1Q  Median      3Q     Max 
-402.78  -82.78    2.22   82.22  377.22 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)      522.778      2.342 223.210  < 2e-16 ***
good_out$Gender1  -28.876      4.751  -6.078 1.35e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 121.2 on 3537 degrees of freedom
Multiple R-squared:  0.01034,	Adjusted R-squared:  0.01006 
F-statistic: 36.94 on 1 and 3537 DF,  p-value: 1.348e-09

> 
> lmi6 <- lm(good_out$Domain ~ good_out$Gender)
> summary(lmi6)

Call:
lm(formula = good_out$Domain ~ good_out$Gender)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.60344 -0.23014  0.02978  0.23605  0.40681 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)       0.606195   0.005463 110.974   <2e-16 ***
good_out$Gender1 -0.013097   0.011037  -1.187    0.235    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2735 on 3317 degrees of freedom
  (220 observations deleted due to missingness)
Multiple R-squared:  0.0004243,	Adjusted R-squared:  0.000123 
F-statistic: 1.408 on 1 and 3317 DF,  p-value: 0.2355

> #MALE DATASET
> male <- subset(good_out, Gender == 0)
> 
> lm8 <- lm(male$Salary ~  male$collegeGPA +
+           male$collegeGPA2 + male$Quant2 + factor(male$CollegeTier) + male$English +
+           male$Logical + male$Logical2 + male$Quant + male$Domain + male$Domain2 +
+           factor(male$GraduationYear) + factor(male$Specialization))

> #FEMALE DATASET
> female <- subset(good_out, Gender == 1)
> 
> lm9 <- lm(female$Salary ~ female$collegeGPA +
```

```
+           female$collegeGPA2 + female$Quant2 + factor(female$CollegeTier) + female$English +
+           female$Logical + female$Logical2 + female$Quant + female$Domain + female$Domain2 +
+           factor(female$GraduationYear) + factor(female$Specialization))
> summary(lm8)

Call:
lm(formula = male$Salary ~ male$collegeGPA + male$collegeGPA2 +
    male$Quant2 + factor(male$CollegeTier) + male$English + male$Logical +
    male$Logical2 + male$Quant + male$Domain + male$Domain2 +
    factor(male$GraduationYear) + factor(male$Specialization))

Residuals:
    Min      1Q  Median      3Q     Max
-4536.0 -1058.4   -44.6   947.7  6352.3

Coefficients:
                                                              Estimate
(Intercept)                                                 -2.901e+03
male$collegeGPA                                              7.839e+01
male$collegeGPA2                                            -1.154e-01
male$Quant2                                                 -1.595e-03
factor(male$CollegeTier)2                                   -5.892e+02
male$English                                                 3.082e+00
male$Logical                                                -6.626e+00
male$Logical2                                                7.563e-03
male$Quant                                                   4.499e+00
male$Domain                                                 -6.548e+02
male$Domain2                                                 1.038e+03
factor(male$GraduationYear)2009                              2.330e+03
factor(male$GraduationYear)2010                              2.255e+03
factor(male$GraduationYear)2011                              1.576e+03
factor(male$GraduationYear)2012                              8.673e+02
factor(male$GraduationYear)2013                              2.200e+02
factor(male$GraduationYear)2014                              1.452e+02
factor(male$GraduationYear)2015                              4.742e+02
factor(male$GraduationYear)2016                             -1.422e+02
factor(male$GraduationYear)2017                             -5.427e+02
factor(male$Specialization)automobile/automotive engineering          -1.017e+03
factor(male$Specialization)biotechnology                    -5.909e+02
factor(male$Specialization)civil engineering                 8.439e+02
factor(male$Specialization)computer engineering              1.700e+02
factor(male$Specialization)computer science & engineering   -1.314e+02
factor(male$Specialization)computer science and technology  -7.294e+01
factor(male$Specialization)electrical engineering           -4.475e+02
factor(male$Specialization)electronics & instrumentation eng -5.632e+01
factor(male$Specialization)electronics & telecommunications   5.525e+01
factor(male$Specialization)electronics and communication engineering   -1.635e+02
factor(male$Specialization)electronics and computer engineering       -2.565e+02
factor(male$Specialization)electronics and electrical engineering     -4.671e+02
```

```
factor(male$Specialization)electronics and instrumentation engineering -1.691e+02
factor(male$Specialization)electronics engineering                -4.026e+01
factor(male$Specialization)industrial & management engineering         2.959e+02
factor(male$Specialization)industrial & production engineering        -4.503e+02
factor(male$Specialization)information science engineering            3.409e+02
factor(male$Specialization)information technology                  2.086e+00
factor(male$Specialization)instrumentation and control engineering      8.684e+02
factor(male$Specialization)instrumentation engineering               5.424e+01
factor(male$Specialization)mechanical and automation               -1.094e+03
factor(male$Specialization)mechanical engineering                 -1.770e+02
factor(male$Specialization)telecommunication engineering             4.224e+02
                                                   Std. Error
(Intercept)                                         2.902e+03
male$collegeGPA                                      6.058e+01
male$collegeGPA2                                     4.213e-01
male$Quant2                                         1.559e-03
factor(male$CollegeTier)2                            1.279e+02
male$English                                        3.446e-01
male$Logical                                        3.320e+00
male$Logical2                                       3.328e-03
male$Quant                                          1.653e+00
male$Domain                                         5.162e+02
male$Domain2                                        4.531e+02
factor(male$GraduationYear)2009                      1.775e+03
factor(male$GraduationYear)2010                      1.592e+03
factor(male$GraduationYear)2011                      1.590e+03
factor(male$GraduationYear)2012                      1.589e+03
factor(male$GraduationYear)2013                      1.588e+03
factor(male$GraduationYear)2014                      1.588e+03
factor(male$GraduationYear)2015                      1.602e+03
factor(male$GraduationYear)2016                      1.830e+03
factor(male$GraduationYear)2017                      1.741e+03
factor(male$Specialization)automobile/automotive engineering          1.768e+03
factor(male$Specialization)biotechnology                         1.770e+03
factor(male$Specialization)civil engineering                     8.758e+02
factor(male$Specialization)computer engineering                   7.989e+02
factor(male$Specialization)computer science & engineering            7.958e+02
factor(male$Specialization)computer science and technology           1.377e+03
factor(male$Specialization)electrical engineering                 8.187e+02
factor(male$Specialization)electronics & instrumentation eng          8.928e+02
factor(male$Specialization)electronics & telecommunications          8.129e+02
factor(male$Specialization)electronics and communication engineering    7.949e+02
factor(male$Specialization)electronics and computer engineering        1.770e+03
factor(male$Specialization)electronics and electrical engineering       8.034e+02
factor(male$Specialization)electronics and instrumentation engineering  8.870e+02
factor(male$Specialization)electronics engineering                 9.922e+02
factor(male$Specialization)industrial & management engineering         1.770e+03
factor(male$Specialization)industrial & production engineering        1.375e+03
factor(male$Specialization)information science engineering            8.811e+02
```

factor(male$Specialization)information technology 7.969e+02
factor(male$Specialization)instrumentation and control engineering 9.929e+02
factor(male$Specialization)instrumentation engineering 1.371e+03
factor(male$Specialization)mechanical and automation 1.377e+03
factor(male$Specialization)mechanical engineering 8.022e+02
factor(male$Specialization)telecommunication engineering 1.766e+03

	t value
(Intercept)	-1.000
male$collegeGPA	1.294
male$collegeGPA2	-0.274
male$Quant2	-1.024
factor(male$CollegeTier)2	-4.606
male$English	8.942
male$Logical	-1.996
male$Logical2	2.272
male$Quant	2.722
male$Domain	-1.269
male$Domain2	2.292
factor(male$GraduationYear)2009	1.312
factor(male$GraduationYear)2010	1.416
factor(male$GraduationYear)2011	0.991
factor(male$GraduationYear)2012	0.546
factor(male$GraduationYear)2013	0.139
factor(male$GraduationYear)2014	0.091
factor(male$GraduationYear)2015	0.296
factor(male$GraduationYear)2016	-0.078
factor(male$GraduationYear)2017	-0.312
factor(male$Specialization)automobile/automotive engineering	-0.575
factor(male$Specialization)biotechnology	-0.334
factor(male$Specialization)civil engineering	0.964
factor(male$Specialization)computer engineering	0.213
factor(male$Specialization)computer science & engineering	-0.165
factor(male$Specialization)computer science and technology	-0.053
factor(male$Specialization)electrical engineering	-0.547
factor(male$Specialization)electronics & instrumentation eng	-0.063
factor(male$Specialization)electronics & telecommunications	0.068
factor(male$Specialization)electronics and communication engineering	-0.206
factor(male$Specialization)electronics and computer engineering	-0.145
factor(male$Specialization)electronics and electrical engineering	-0.581
factor(male$Specialization)electronics and instrumentation engineering	-0.191
factor(male$Specialization)electronics engineering	-0.041
factor(male$Specialization)industrial & management engineering	0.167
factor(male$Specialization)industrial & production engineering	-0.327
factor(male$Specialization)information science engineering	0.387
factor(male$Specialization)information technology	0.003
factor(male$Specialization)instrumentation and control engineering	0.875
factor(male$Specialization)instrumentation engineering	0.040
factor(male$Specialization)mechanical and automation	-0.795
factor(male$Specialization)mechanical engineering	-0.221

```
factor(male$Specialization)telecommunication engineering              0.239
                                                        Pr(>|t|)
(Intercept)                                                 0.31763
male$collegeGPA                                             0.19585
male$collegeGPA2                                            0.78413
male$Quant2                                                 0.30615
factor(male$CollegeTier)2                                   4.31e-06 ***
male$English                                                < 2e-16 ***
male$Logical                                                0.04609 *
male$Logical2                                               0.02314 *
male$Quant                                                  0.00653 **
male$Domain                                                 0.20471
male$Domain2                                                0.02200 *
factor(male$GraduationYear)2009                             0.18952
factor(male$GraduationYear)2010                             0.15681
factor(male$GraduationYear)2011                             0.32171
factor(male$GraduationYear)2012                             0.58515
factor(male$GraduationYear)2013                             0.88980
factor(male$GraduationYear)2014                             0.92717
factor(male$GraduationYear)2015                             0.76729
factor(male$GraduationYear)2016                             0.93809
factor(male$GraduationYear)2017                             0.75523
factor(male$Specialization)automobile/automotive engineering          0.56527
factor(male$Specialization)biotechnology                    0.73853
factor(male$Specialization)civil engineering                0.33536
factor(male$Specialization)computer engineering             0.83148
factor(male$Specialization)computer science & engineering             0.86883
factor(male$Specialization)computer science and technology            0.95777
factor(male$Specialization)electrical engineering           0.58471
factor(male$Specialization)electronics & instrumentation eng          0.94971
factor(male$Specialization)electronics & telecommunications           0.94581
factor(male$Specialization)electronics and communication engineering    0.83704
factor(male$Specialization)electronics and computer engineering       0.88479
factor(male$Specialization)electronics and electrical engineering     0.56102
factor(male$Specialization)electronics and instrumentation engineering  0.84882
factor(male$Specialization)electronics engineering          0.96764
factor(male$Specialization)industrial & management engineering        0.86730
factor(male$Specialization)industrial & production engineering        0.74334
factor(male$Specialization)information science engineering            0.69889
factor(male$Specialization)information technology           0.99791
factor(male$Specialization)instrumentation and control engineering    0.38190
factor(male$Specialization)instrumentation engineering      0.96844
factor(male$Specialization)mechanical and automation        0.42667
factor(male$Specialization)mechanical engineering           0.82542
factor(male$Specialization)telecommunication engineering    0.81103
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1579 on 2463 degrees of freedom
```

```
  (173 observations deleted due to missingness)
Multiple R-squared:  0.3427,  Adjusted R-squared:  0.3315
F-statistic: 30.58 on 42 and 2463 DF,  p-value: < 2.2e-16

> summary(lm9)

Call:
lm(formula = female$Salary ~ female$collegeGPA + female$collegeGPA2 +
    female$Quant2 + factor(female$CollegeTier) + female$English +
    female$Logical + female$Logical2 + female$Quant + female$Domain +
    female$Domain2 + factor(female$GraduationYear) + factor(female$Specialization))

Residuals:
    Min      1Q  Median      3Q     Max
-4724.8  -917.2    81.2   930.7  6579.0

Coefficients:
                                                                      Estimate
(Intercept)                                                          -5.956e+03
female$collegeGPA                                                     2.400e+02
female$collegeGPA2                                                   -1.351e+00
female$Quant2                                                         2.324e-03
factor(female$CollegeTier)2                                          -3.395e+02
female$English                                                        2.801e+00
female$Logical                                                       -1.011e+01
female$Logical2                                                       1.168e-02
female$Quant                                                          5.516e-01
female$Domain                                                        -1.879e+02
female$Domain2                                                        3.476e+02
factor(female$GraduationYear)2010                                     6.047e+02
factor(female$GraduationYear)2011                                     3.103e+01
factor(female$GraduationYear)2012                                    -7.587e+02
factor(female$GraduationYear)2013                                    -1.208e+03
factor(female$GraduationYear)2014                                    -1.316e+03
factor(female$GraduationYear)2015                                    -1.090e+03
factor(female$Specialization)civil engineering                        2.932e+02
factor(female$Specialization)computer application                    -2.776e+03
factor(female$Specialization)computer engineering                     9.371e+02
factor(female$Specialization)computer science & engineering           5.257e+02
factor(female$Specialization)electrical engineering                   3.116e+02
factor(female$Specialization)electronics & instrumentation eng        9.218e+01
factor(female$Specialization)electronics & telecommunications         7.552e+02
factor(female$Specialization)electronics and communication engineering    4.167e+02
factor(female$Specialization)electronics and electrical engineering       3.970e+02
factor(female$Specialization)electronics and instrumentation engineering  3.803e+02
factor(female$Specialization)electronics engineering                  1.236e+03
factor(female$Specialization)industrial & production engineering     2.077e+03
factor(female$Specialization)industrial engineering                   2.197e+03
factor(female$Specialization)information & communication technology  1.564e+03
```

factor(female$Specialization)information science engineering 5.491e+02
factor(female$Specialization)information technology 7.953e+02
factor(female$Specialization)instrumentation and control engineering 1.035e+03
factor(female$Specialization)mechanical engineering 1.077e+03

	Std. Error
(Intercept)	4.708e+03
female$collegeGPA	1.227e+02
female$collegeGPA2	8.309e-01
female$Quant2	3.114e-03
factor(female$CollegeTier)2	2.730e+02
female$English	6.065e-01
female$Logical	6.479e+00
female$Logical2	6.465e-03
female$Quant	3.079e+00
female$Domain	8.828e+02
female$Domain2	7.829e+02
factor(female$GraduationYear)2010	9.046e+02
factor(female$GraduationYear)2011	8.960e+02
factor(female$GraduationYear)2012	8.902e+02
factor(female$GraduationYear)2013	8.937e+02
factor(female$GraduationYear)2014	8.961e+02
factor(female$GraduationYear)2015	1.030e+03
factor(female$Specialization)civil engineering	1.082e+03
factor(female$Specialization)computer application	1.737e+03
factor(female$Specialization)computer engineering	8.616e+02
factor(female$Specialization)computer science & engineering	8.628e+02
factor(female$Specialization)electrical engineering	9.636e+02
factor(female$Specialization)electronics & instrumentation eng	1.002e+03
factor(female$Specialization)electronics & telecommunications	9.008e+02
factor(female$Specialization)electronics and communication engineering	8.618e+02
factor(female$Specialization)electronics and electrical engineering	8.951e+02
factor(female$Specialization)electronics and instrumentation engineering	1.212e+03
factor(female$Specialization)electronics engineering	1.708e+03
factor(female$Specialization)industrial & production engineering	1.726e+03
factor(female$Specialization)industrial engineering	1.705e+03
factor(female$Specialization)information & communication technology	1.358e+03
factor(female$Specialization)information science engineering	1.002e+03
factor(female$Specialization)information technology	8.610e+02
factor(female$Specialization)instrumentation and control engineering	1.217e+03
factor(female$Specialization)mechanical engineering	9.750e+02

	t value
(Intercept)	-1.265
female$collegeGPA	1.955
female$collegeGPA2	-1.626
female$Quant2	0.746
factor(female$CollegeTier)2	-1.244
female$English	4.619
female$Logical	-1.560
female$Logical2	1.807

```
female$Quant                                                    0.179
female$Domain                                                  -0.213
female$Domain2                                                  0.444
factor(female$GraduationYear)2010                               0.669
factor(female$GraduationYear)2011                               0.035
factor(female$GraduationYear)2012                              -0.852
factor(female$GraduationYear)2013                              -1.352
factor(female$GraduationYear)2014                              -1.468
factor(female$GraduationYear)2015                              -1.059
factor(female$Specialization)civil engineering                  0.271
factor(female$Specialization)computer application              -1.598
factor(female$Specialization)computer engineering               1.088
factor(female$Specialization)computer science & engineering     0.609
factor(female$Specialization)electrical engineering             0.323
factor(female$Specialization)electronics & instrumentation eng  0.092
factor(female$Specialization)electronics & telecommunications   0.838
factor(female$Specialization)electronics and communication engineering  0.484
factor(female$Specialization)electronics and electrical engineering     0.444
factor(female$Specialization)electronics and instrumentation engineering  0.314
factor(female$Specialization)electronics engineering            0.723
factor(female$Specialization)industrial & production engineering  1.204
factor(female$Specialization)industrial engineering             1.289
factor(female$Specialization)information & communication technology  1.152
factor(female$Specialization)information science engineering   0.548
factor(female$Specialization)information technology             0.924
factor(female$Specialization)instrumentation and control engineering  0.850
factor(female$Specialization)mechanical engineering             1.104
                                                     Pr(>|t|)
(Intercept)                                                     0.2062
female$collegeGPA                                               0.0509 .
female$collegeGPA2                                              0.1044
female$Quant2                                                   0.4557
factor(female$CollegeTier)2                                     0.2140
female$English                                                  4.51e-06 ***
female$Logical                                                  0.1191
female$Logical2                                                 0.0711 .
female$Quant                                                    0.8579
female$Domain                                                   0.8315
female$Domain2                                                  0.6572
factor(female$GraduationYear)2010                               0.5040
factor(female$GraduationYear)2011                               0.9724
factor(female$GraduationYear)2012                               0.3943
factor(female$GraduationYear)2013                               0.1768
factor(female$GraduationYear)2014                               0.1424
factor(female$GraduationYear)2015                               0.2899
factor(female$Specialization)civil engineering                  0.7865
factor(female$Specialization)computer application               0.1104
factor(female$Specialization)computer engineering               0.2771
factor(female$Specialization)computer science & engineering     0.5425
```

```
factor(female$Specialization)electrical engineering                    0.7465
factor(female$Specialization)electronics & instrumentation eng          0.9267
factor(female$Specialization)electronics & telecommunications           0.4021
factor(female$Specialization)electronics and communication engineering    0.6288
factor(female$Specialization)electronics and electrical engineering       0.6575
factor(female$Specialization)electronics and instrumentation engineering  0.7538
factor(female$Specialization)electronics engineering                    0.4698
factor(female$Specialization)industrial & production engineering         0.2291
factor(female$Specialization)industrial engineering                     0.1979
factor(female$Specialization)information & communication technology      0.2498
factor(female$Specialization)information science engineering            0.5837
factor(female$Specialization)information technology                     0.3559
factor(female$Specialization)instrumentation and control engineering      0.3956
factor(female$Specialization)mechanical engineering                     0.2698
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1470 on 778 degrees of freedom
  (47 observations deleted due to missingness)
Multiple R-squared:  0.302,    Adjusted R-squared:  0.2715
F-statistic: 9.901 on 34 and 778 DF,  p-value: < 2.2e-16
```