

Final Research Paper

Dao (Cherry) & Khanh

Introduction:

In 2012, spending per employee remained stable (Miller, 2013). The Solow Paradox theory states that in the era of technology, the benefits of technologies have not yet been materialized at scale. Despite the high investment in technological innovation, many companies have not seen a direct and immediate impact on productivity growth. This brings into question the role of training on productivity growth, especially when capability gaps are one of the top challenges of corporates, controlling for other factors.

Researchers find different results on the impacts of training on performance. One research concludes that training will improve productivity if the problems are caused by insufficient skills and knowledge (Imran, 2013). This finding is supported by another paper that states a positive correlation between the impact of training and the performance of employees at universities in Jordan (Al-Mzary, et al., 2015). The authors also suggest more suitable training programs for employees and equal opportunities for the employees to attend the training courses. The third study also agrees that investment in employee training can create better employee performance. The researchers explore the impacts of training investment on productivity in 14 Canadian industries from 1999 to 2005 and conclude that in 12 out of 14 industries, training increases productivity. Employers provide employee training because they see the threat of new technology, which requires an investment in training to maintain the firm's current labor productivity. It also challenges our idea of the Solow Paradox that firms do not care about and invest enough in employee training to boost productivity with the rise of technological innovation.

This paper plans to explore the correlation between training and employee performance, controlling for other factors such as educational background, gender, and level of job involvement. Based on literature reviews, our paper hypothesizes that more investment in employee training will increase productivity as it equips the employees with the necessary skills and knowledge.

Materials and Methods:

This research looks at the HR Analytics Case Study dataset from Kaggle, which revolves around a fictitious company with a high turnover rate. The data sources contain several datasets, but we will mainly focus on the three datasets called the employee_survey and draw relevant explanatory variables from the two datasets called general_data and management_survey. Our data will contain explanatory variables from the employee_survey and general_data and the dependent variable from the manager_survey. The main dependent variables are Monthly Income and Performance Rating. The research paper focuses on employees who are currently still working at companies. Proxies for relevant factors that affect the performance are drawn from those three datasets.

The merged dataset requires some data cleaning and removal of variable(s) that we consider are irrelevant, such as whether the employees are over 18 or not, stock options level, standard hours, etc. Regarding Gender, empty cells are replaced with Female genders, arguing that this will help to balance out our dataset with the Male population of 2646 and Female population of only 1752. With variables of Job Involvement, Environment Satisfaction, Job Satisfaction, Work-Life Balance and Monthly Income, empty cells are replaced with the previous values that are filled with values. This paper assumes that a person is affected by the coworkers around them or their workplace, which reflects their satisfaction with their work and lifestyle. Other variables such as Training Times Last Year, Age, Years At Company, Education, Performance Rating with empty values are replaced with the mean of the column rounded to the nearest integer for easy usability in modeling. By filling out empty cells with substitute values, we do not remove any rows that contain missing data.

After the cleaning process, we create the indicator variable for the explanatory categorical variables. Our analysis includes 1) performing multiple linear regression with Monthly Income as a response variable, 2)

checking the conditions, 3) performing logistics regression with Performance Rating as the dependent variable, and 4) conducting ANOVA test to choose the best model.

In the first part, we use Monthly Income as an indicator for the response variable productivity with the main chosen variable as Training Times Last Year employees received. This paper hypothesizes that the more training employees receive, the better they perform at their work, and so the higher income. However, the model with Monthly Income as a dependent variable violates the conditions for linear regression. Therefore, in the second part, we choose to use Performance Rating as the response variable for logistic regression. Performance Rating is a quantitative variable that only contains two values “3” and “4,” but as we would like to this variable as our response variable, we transform it into a binary variable and introduce a new dataset (Appendix). This method allows us to conduct a logistics regression model to quantify the relationship between our outcome and predictor variables.

Result:

Model 1: Examining the relationship between our first quantitative response variable Monthly Income with individual explanatory variables such as Gender, Marital Status, Training, and Age, we notice that the distribution of data points results in many outliers (Appendix). This gives us a hint that Monthly Income might not be a good outcome variable. More variables included in our model would result in even more outliers. Therefore, we perform a multiple regression between the response variable Monthly Income and only several explanatory variables Training Received, Age, Gender and Marital Status. The model shows that Gender and Marital Status are not significant with p-value >0.05 . This suggests that we should include Training and Age in our next model and remove the other two variables.

Model 1:

```
summary(modelTrainMonInvsAgeMarriedMale)

##
## Call:
## lm(formula = MonthlyIncome ~ TrainingTimesLastYear + Age + Male +
##     Married, data = TrainvsProdv2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -59706 -34796 -15643  17687 139023
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    66918.20    3545.61  18.874 < 2e-16 ***
## TrainingTimesLastYear 1810.76     549.36   3.296 0.000988 ***
## Age             -230.61      77.81  -2.964 0.003055 **
## Male              784.09    1446.42   0.542 0.587785
## Married          2373.46    1425.68   1.665 0.096024 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 46970 on 4405 degrees of freedom
## Multiple R-squared:  0.005037,    Adjusted R-squared:  0.004133
## F-statistic: 5.575 on 4 and 4405 DF,  p-value: 0.0001789
```

Model 2: Having the incentive to include Training and Age in our next model, we conduct a linear regression analysis to quantify the correlation between those two variables and Monthly Income. The summary statistics reveal that both of those two variables are significant with p.values <0.05 . The coefficient on Training is positive and highly significant, stating that an increase in one time of training during last year would lead to an increase of 1787.07 dollars of Monthly Income, holding Age constant. We also notice a negative coefficient

of 221.49 dollars on Age, which indicates that an increase in one-year-old can lead to a reduction of 221.49 dollars of Monthly Income when we control for Training. We are 95% confident that for every additional training a random employee received, the monthly income of that employee increases between 710.6071 and 2863.52791 dollars. In addition, we can expect 95% that the monthly income ranges from -373.4069 to -69.56626 dollars for every one year increase in one employee's age.

Model 2:

```
summary(modelTrainMonInvsAgeTrain)

##
## Call:
## lm(formula = MonthlyIncome ~ TrainingTimesLastYear + Age, data = TrainvsProdv2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -58787 -34821 -15876  18166 139619
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    68204.87    3360.03   20.299 < 2e-16 ***
## TrainingTimesLastYear 1787.07     549.07    3.255  0.00114 **
## Age           -221.49       77.49   -2.858  0.00428 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 46980 on 4407 degrees of freedom
## Multiple R-squared:  0.004357,    Adjusted R-squared:  0.003905
## F-statistic: 9.643 on 2 and 4407 DF,  p-value: 6.629e-05
```

```
confint(modelTrainMonInvsAgeTrain)

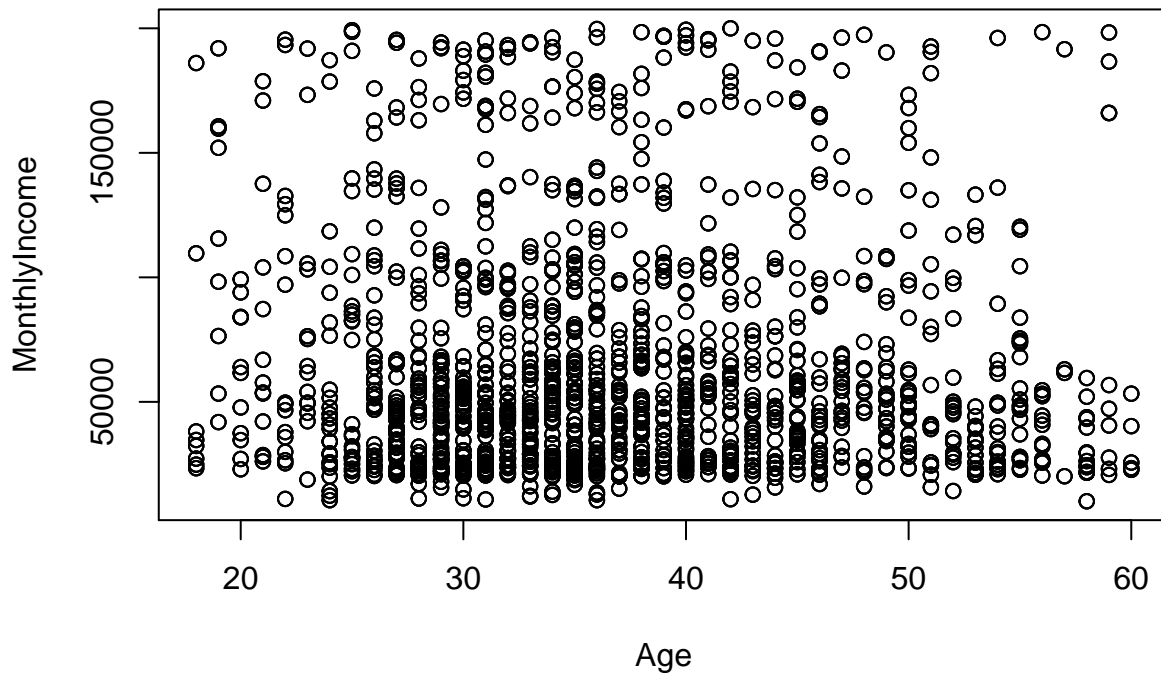
##              2.5 %      97.5 %
## (Intercept)  61617.5226 74792.21295
## TrainingTimesLastYear  710.6071 2863.52791
## Age          -373.4069  -69.56626
```

To have a closer look at the relationships between the response variable with each individual predictor, we use graphs to illustrate those relationships visually (Figure 1):

Figure (1):

```
# Plot of Monthly Income by Age
plot(MonthlyIncome ~ Age, data = TrainvsProdv2, main = "Scatterplot of Monthly Income by Age")
```

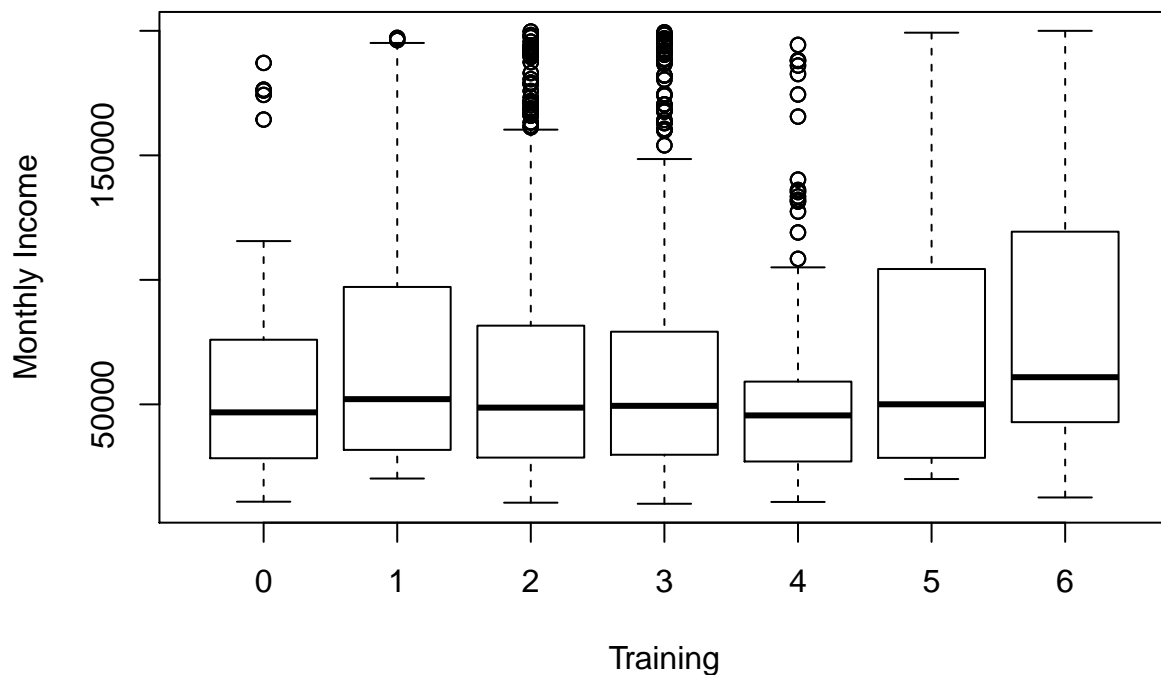
Scatterplot of Monthly Income by Age



```
# Boxplot of Monthly Income by Training
```

```
boxplot(MonthlyIncome ~ TrainingTimesLastYear, main = "Boxplot of Monthly Income by Gender", data = Tra
```

Boxplot of Monthly Income by Gender



The graphs above from Figure (1) demonstrates that there are many outliers at the tail of each box and far from the mean of the variables we are considering. Similarly, the data points in the plot between Monthly Income and Age seems messy. These data points do not indicate clearly what is the trend or what sort of relationship

between these variables. This gives us the incentive to check the conditions of linear regression from our model 2.

Figure (2):

```
plot(modelTrainMonInvsAgeTrain)
```

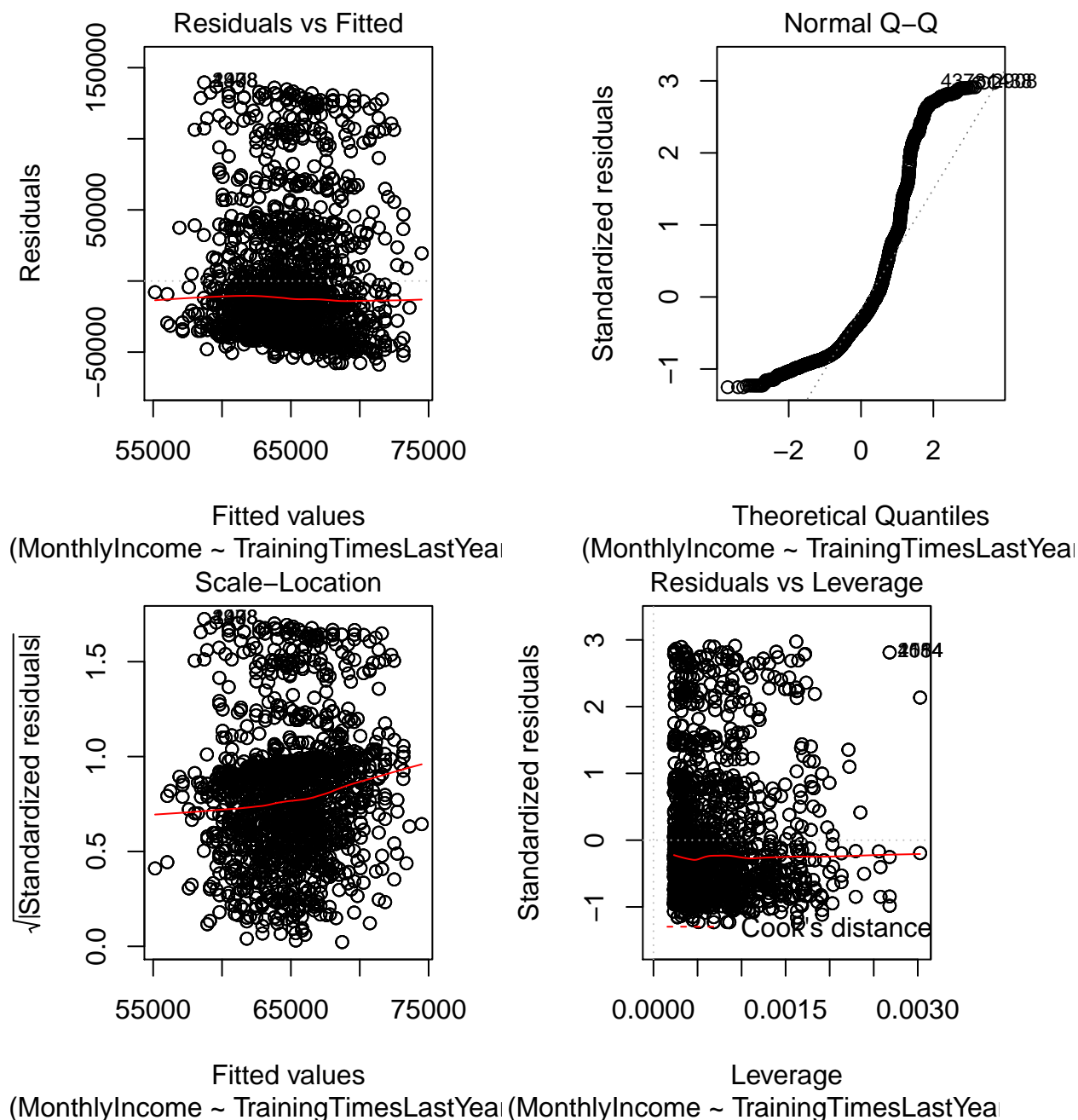


Figure (2) above shows that our model 2 does not satisfy the condition of normal distribution. Here, we observe an upper tail in one end. This result, combined with the existence of multiple outliers observed in data exploration provides us with an incentive to change our response variable to Performance Rating. Here, we hypothesize that more training provided will lead to better performance and thus higher performance rating.

Model 3: In this second section, we run the multiple logistic regression model in predicting the Performance

Rating based on Income, Age, Gender, Education, Job Involvement, Marital Status, Years At Company, Environment Satisfaction, Job Satisfaction, Work-Life Balance, and Training Times Last Year. The summary statistics illustrate very small coefficients in this model. Moreover, Monthly Income, Age, Job Involvement, Marital Status, Years At Company, Training Times Last Year (our main predictor), Environment Satisfaction, and Work-Life Balance are insignificant with p-value >0.05.

Model 3:

```
summary(PerfRatevsAllwoDepart)
```

```
##
## Call:
## glm(formula = PerformanceRatingv2 ~ MonthlyIncome + Male + Education +
##      Age + JobInvolvement + Married + YearsAtCompany + TrainingTimesLastYear +
##      EnvironmentSatisfaction + JobSatisfaction + WorkLifeBalance,
##      family = binomial, data = TrainvsProd3)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7621  -0.6088  -0.5570  -0.4939   2.2696
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.526e+00  3.775e-01  -4.044 5.26e-05 ***
## MonthlyIncome    8.924e-07  8.765e-07   1.018  0.30861
## Male           2.824e-01  8.807e-02   3.206  0.00134 **
## Education      -9.320e-02  4.064e-02  -2.293  0.02182 *
## Age            -7.828e-03  4.913e-03  -1.593  0.11108
## JobInvolvement   3.030e-02  5.941e-02   0.510  0.61012
## Married         5.290e-02  8.452e-02   0.626  0.53141
## YearsAtCompany  -1.194e-03  7.415e-03  -0.161  0.87205
## TrainingTimesLastYear -4.782e-02  3.291e-02  -1.453  0.14619
## EnvironmentSatisfaction 4.002e-02  3.858e-02   1.037  0.29953
## JobSatisfaction  1.060e-01  3.879e-02   2.731  0.00631 **
## WorkLifeBalance  -8.337e-02  5.861e-02  -1.422  0.15491
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 3785.1  on 4409  degrees of freedom
## Residual deviance: 3750.5  on 4398  degrees of freedom
## AIC: 3774.5
##
## Number of Fisher Scoring iterations: 4
```

Therefore, we perform a drop-in-deviance test to have the best final model. The full model includes all the predictors and the reduced model removes those that prove to be insignificant in our logistic regression analysis above. The Analysis of Deviance table shows a p-value of 0.3052, which fails to reject the null hypothesis. Therefore, we cannot argue that including those insignificant variables in our final model will give us a better model than the reduced model, and thus we decide to include only predictors Gender, Education, and Job Satisfaction in our final model - Model 4.

Model 4:

```
summary(PerfRatevsGenEduJobSatis)
```

```
##
## Call:
## glm(formula = PerformanceRatingv2 ~ Male + Education + JobSatisfaction,
##      family = binomial, data = TrainvsProd3)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6982  -0.6114  -0.5576  -0.5076   2.1411
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.91977    0.17369 -11.053  < 2e-16 ***
## Male           0.29393    0.08773   3.350 0.000807 ***
## Education     -0.09348    0.04050  -2.308 0.020995 *
## JobSatisfaction 0.10798    0.03874   2.787 0.005318 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 3785.1  on 4409  degrees of freedom
## Residual deviance: 3760.0  on 4406  degrees of freedom
## AIC: 3768
##
## Number of Fisher Scoring iterations: 4
```

```
anova(PerfRatevsGenEduJobSatis, PerfRatevsAllwoDepart, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: PerformanceRatingv2 ~ Male + Education + JobSatisfaction
## Model 2: PerformanceRatingv2 ~ MonthlyIncome + Male + Education + Age +
##      JobInvolvement + Married + YearsAtCompany + TrainingTimesLastYear +
##      EnvironmentSatisfaction + JobSatisfaction + WorkLifeBalance
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         4406      3760.0
## 2         4398      3750.5  8   9.4572  0.3052
```

Model 4: Our final model includes Performance Rating as a response binary variable and Gender, Education, and Job Satisfaction as explanatory variables. This model describes a relationship between log(Performance Rating) with Gender, Education, and Job Satisfaction. By taking the exponential terms of the reduced multiple logistic model, we have the following interpretations. If the gender of an employee is a male, then we can expect there is a 1.34169 multiplicative change in the odds of the performance rating, after adjusting for other variables. Each degree increase in an employee's education level (on a scale from 1 to 5) is associated with an 8.924% increase in the odds of having performance rating of 4 holding other variables constant. Finally, for every increased level of job satisfaction of an employee, the odds that this employee's performance rating is 4 changes by a factor of 1.11402, adjusting for other factors.

Model 4 (exponentiated):

```
exp(PerfRatevsGenEduJobSatis$coefficients)
```

```
##      (Intercept)           Male      Education JobSatisfaction
##      0.1466414       1.3416931      0.9107607      1.1140218
```

Discussion:

Our final model does not attest to our hypothesis that training will improve employee performance. Instead, Gender, Education, and Job Satisfaction are correlated with the odds of achieving the highest rating of 4 for performance. We find it interesting that Job Satisfaction plays an important role here. Employees with a high level of job satisfaction will make good business sense and perform higher productivity as well as career enhancement (Shmailam, 2016). Therefore, it is important for employers to understand the needs of their employees and . This can be done through assigning them the work that they are most passionate about, rewarding them when necessary, acknowledging their contribution, and providing assistance to employees when needed. Better employee performance can lead to more employee engagement and organizational success (Shmailam, 2016).

Due to the limited time period, we are not able to conduct linear and logistics regressions on another dataset that has a more relevant dependent variable to our hypothesis. Even though we do think that performance rating is a good response variable that is correlated with training, the fact that we only have values of “3” and “4” for the rating can create bias in our analysis. That is also why we choose to use monthly income in the beginning instead of performance rating, but the income that employees received monthly turns out to be not a good indicator for an employee’s performance. Although the performance rating might have many biased due to other confounding factors and is not an excellent indicator in how training affects on an employee’s productivity, it does make sense when it comes to explaining why there is a difference in performance among employees based on other variables in the analysis. We believe that our final model could have been different if we do not have any missing values from the dataset.

We are aware that there might be some potential confounding variables in our research, such as the department that the employee is working for. Since we have three different departments, we also think work allocation can vary across departments, with different departments having different deadlines, budgets, and busy seasons. Another potential confounding variable is the total number of companies the employee has worked for. This variable can help us understand how to adjust training programs appropriately for maximum effectiveness. The last one potential confounding factor is the level of relationship satisfaction. The interaction between co-workers seems to be a valuable indicator for one’s productivity in an organization. The close-knit community in one department can reflect or resonate the effectiveness of training programs.

Furthermore, the dataset we collect might come from one specific company, but our analysis might be generalised to a broader perspective on the impacts of training programs for boosting human resources outcome in order to bring benefits for their company. We suggest that researchers should look at the impact of gender on employee performance. Do men perform better because they are actually good at the job or because it is the gender bias that shapes the perception of the employers on their employees’ achievement? Can women do better at work if they are responsible for doing all the household chores and taking care of the children? Those are some interesting things that future researchers can explore to help improve the overall performance of an organization. Overall, our research provides helpful insights into what can potentially determine the performance of employees so that the employers can construct their welfare programs for their workers.

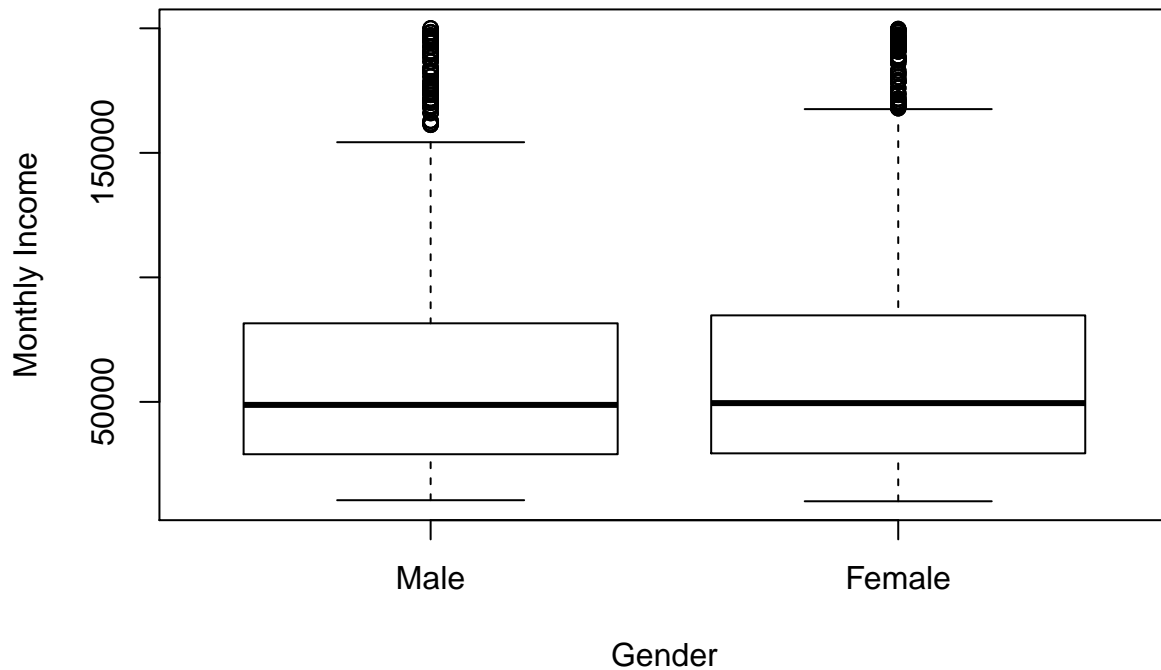
Appendix:

```
TrainvsProd <- read.csv("~/Stats 272b F19/Project/Dao (Cherry) & Khanh/Completed dataset.csv")
```

Exploratory Data Analysis

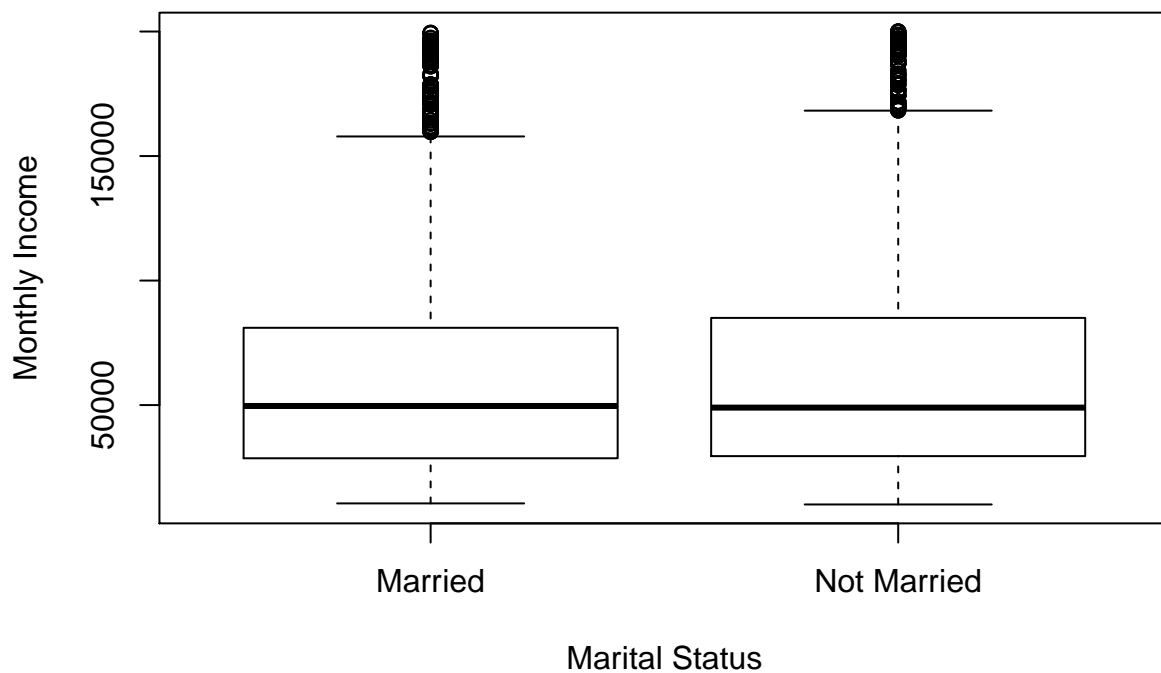
```
par(mfrow = c(1,1))  
# Boxplot of Monthly Income by Gender  
boxplot(MonthlyIncome ~ Male, main = "Boxplot of Monthly Income by Gender", data = TrainvsProdv2, names
```


Boxplot of Monthly Income by Gender



```
# Boxplot of Monthly Income by Marital Status  
boxplot(MonthlyIncome ~ Married, main = "Boxplot of Monthly Income by Marital Status", data = TrainvsPr
```

Boxplot of Monthly Income by Marital Status



```
par(mfrow = c(1,1))
```

We see that average monthly income is pretty much similar for both genders. We also do not see any difference

in monthly income between 2 different stages of marital status.

Model 1:

```
modelTrainMonInvsAgeMarriedMale <- lm(MonthlyIncome ~ TrainingTimesLastYear + Age + Male + Married, data = TrainvsProdv2)
summary(modelTrainMonInvsAgeMarriedMale)

##
## Call:
## lm(formula = MonthlyIncome ~ TrainingTimesLastYear + Age + Male + Married, data = TrainvsProdv2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -59706 -34796 -15643  17687 139023
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    66918.20    3545.61  18.874 < 2e-16 ***
## TrainingTimesLastYear  1810.76     549.36   3.296 0.000988 ***
## Age             -230.61      77.81  -2.964 0.003055 **
## Male              784.09     1446.42   0.542 0.587785
## Married          2373.46     1425.68   1.665 0.096024 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 46970 on 4405 degrees of freedom
## Multiple R-squared:  0.005037, Adjusted R-squared:  0.004133
## F-statistic: 5.575 on 4 and 4405 DF, p-value: 0.0001789
```

Model 2:

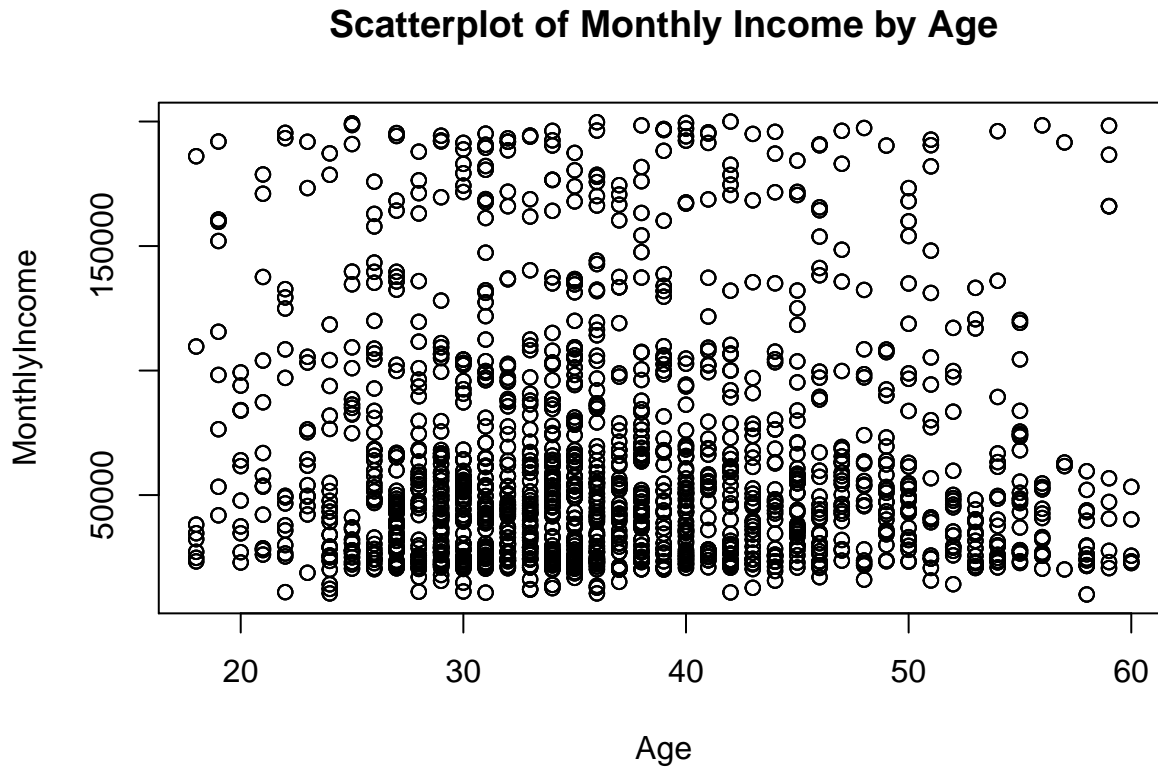
```
modelTrainMonInvsAgeTrain <- lm(MonthlyIncome ~ TrainingTimesLastYear + Age, data = TrainvsProdv2)
summary(modelTrainMonInvsAgeTrain)

##
## Call:
## lm(formula = MonthlyIncome ~ TrainingTimesLastYear + Age, data = TrainvsProdv2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -58787 -34821 -15876  18166 139619
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    68204.87    3360.03  20.299 < 2e-16 ***
## TrainingTimesLastYear  1787.07     549.07   3.255 0.00114 **
## Age             -221.49      77.49  -2.858 0.00428 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 46980 on 4407 degrees of freedom
## Multiple R-squared:  0.004357, Adjusted R-squared:  0.003905
## F-statistic: 9.643 on 2 and 4407 DF, p-value: 6.629e-05
confint(modelTrainMonInvsAgeTrain)
```

```
##                2.5 %      97.5 %
## (Intercept)    61617.5226 74792.21295
## TrainingTimesLastYear 710.6071 2863.52791
## Age            -373.4069  -69.56626
```

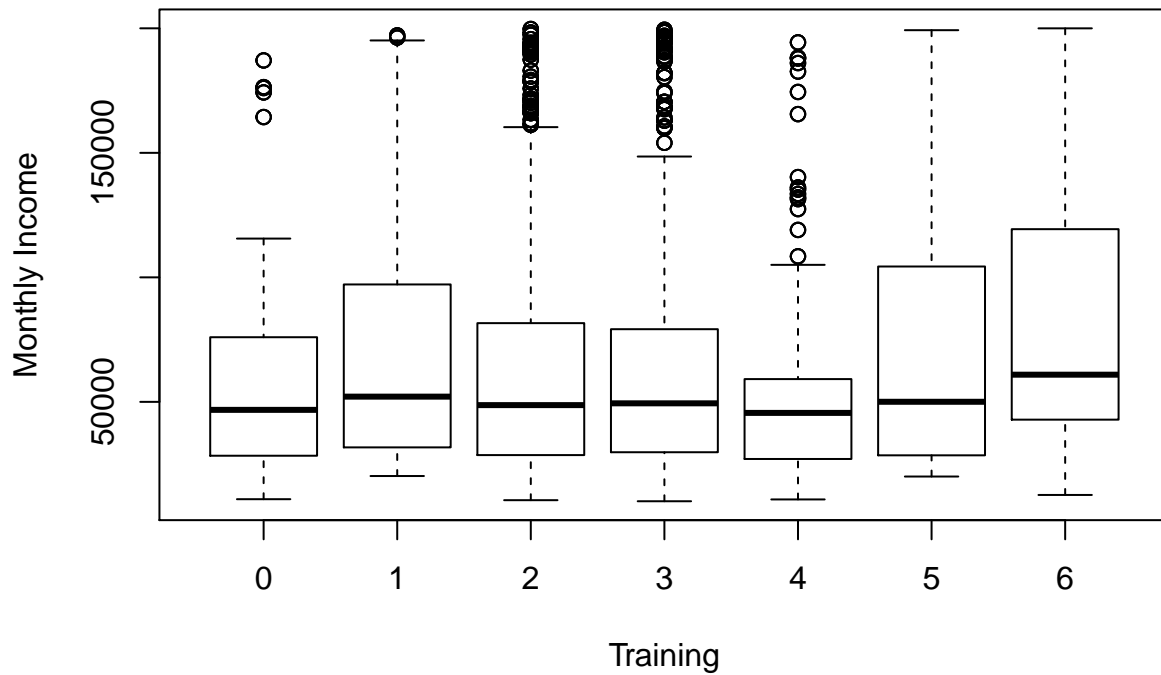
Figure (1): We can also examine the relationship between Monthly Income and Age through a scatterplot:

```
# Plot of Monthly Income by Age
plot(MonthlyIncome ~ Age, data = TrainvsProdv2, main = "Scatterplot of Monthly Income by Age")
```



```
# Boxplot of Monthly Income by Training
boxplot(MonthlyIncome ~ TrainingTimesLastYear, main = "Boxplot of Monthly Income by Gender", data = Tra
```

Boxplot of Monthly Income by Gender

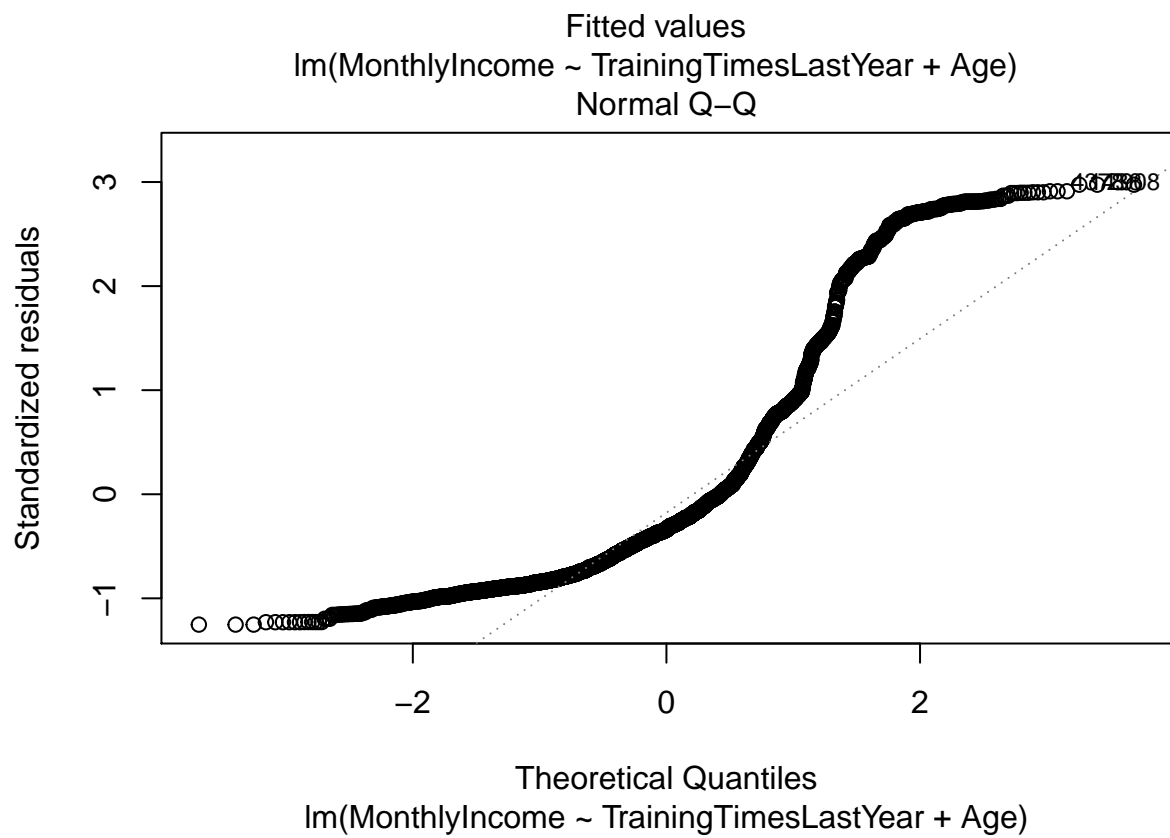
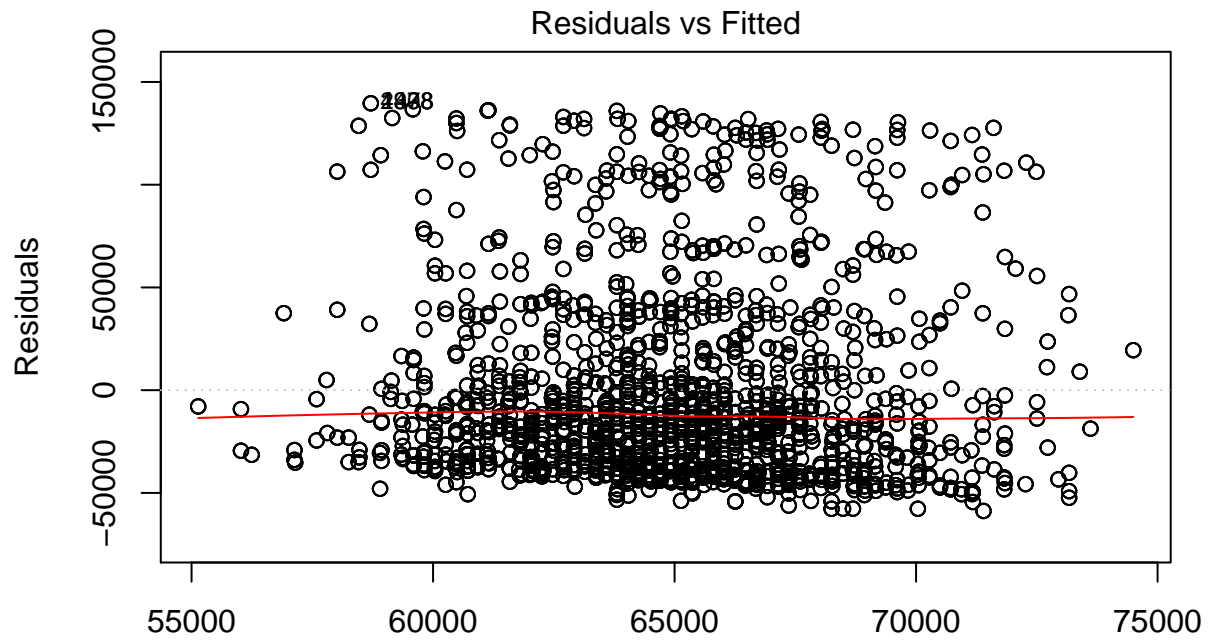


Re-
garding training, we can see the difference in monthly income between those who have the least training and the most training. For the one with the least training, we do have some outliers, but they are few. Majority have a lot of training, with higher range of income distribution (up to 100,000). This implies a correlation between income and training.

The graphs above from Figure (1) demonstrates that there are many outliers at the tail of each box and far from the mean of the variables we are considering. Similarly, the data points in the plot between Monthly Income and Age seems messy. These data points do not indicate clearly what is the trend or what sort of relationship between these variables.

Figure (2):

```
plot(modelTrainMonInvsAgeTrain)
```



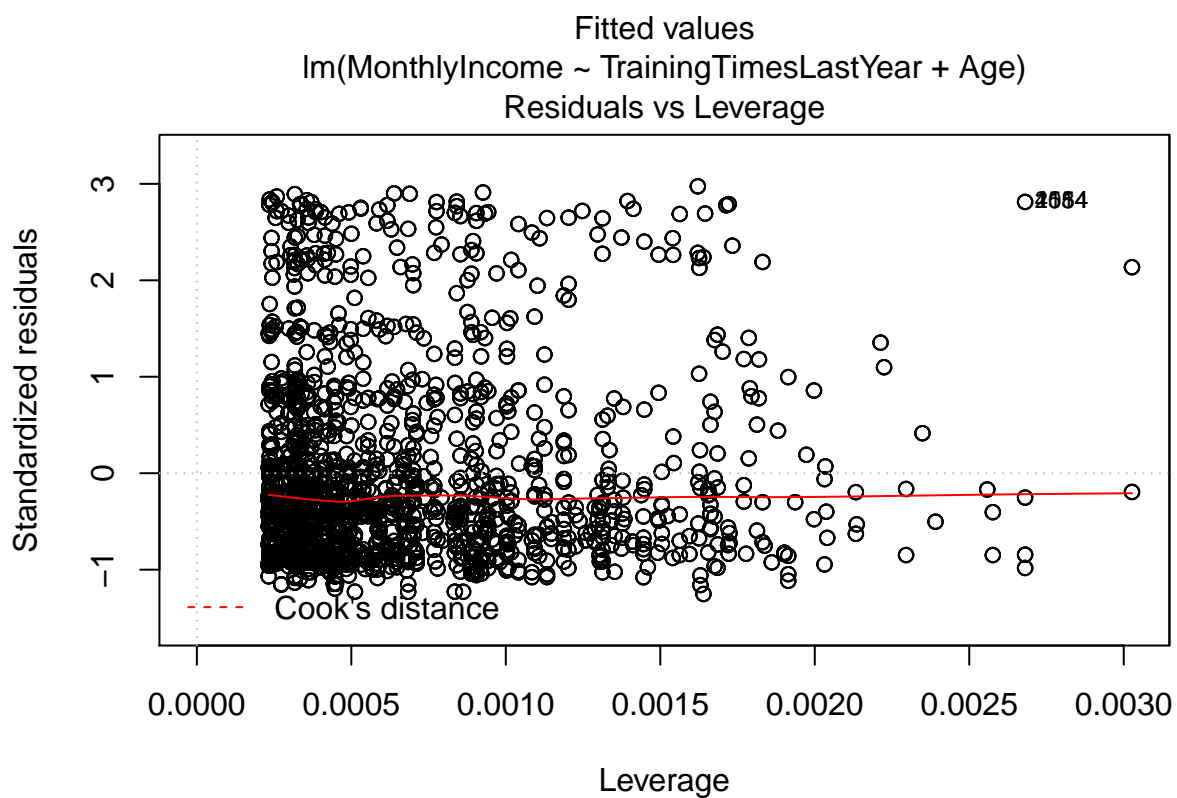
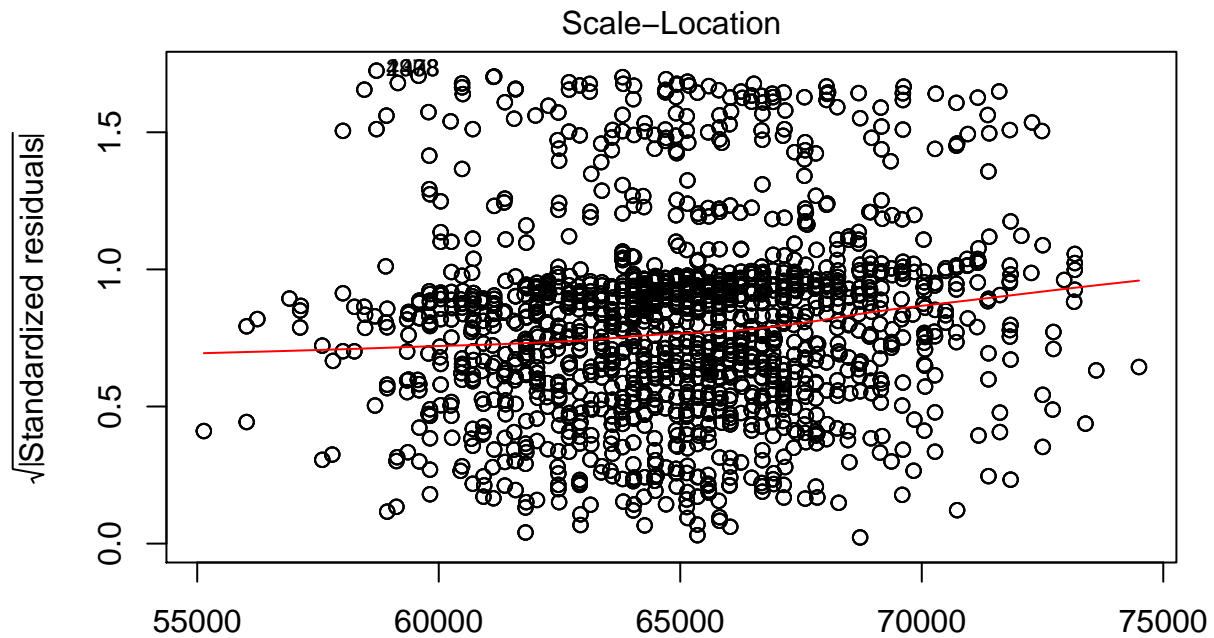


Fig-

ure (2) shows that our model 2 does not satisfy the condition of normal distribution. Here, we observe an upper tail in one end.

```
# We create the indicator variable for the quantitative variable 'PerformanceRating' and introduce a new variable
TrainvsProdv3 <- TrainvsProdv2 %>%
  mutate(PerformanceRatingv2 = ifelse(TrainvsProdv2$PerformanceRating == 4, 1, 0)) %>% # Let '4' be 1
  select("Departmentv2", "Age", "Education", "EducationFieldv2", "Male", "JobInvolvement", "PerformanceRatingv2")
```

Model 3:

```
PerfRatevsAllwoDepart <- glm(PerformanceRatingv2 ~ MonthlyIncome + Male + Education + Age + JobInvolvement +  
summary(PerfRatevsAllwoDepart)
```

```
##  
## Call:  
## glm(formula = PerformanceRatingv2 ~ MonthlyIncome + Male + Education +  
##     Age + JobInvolvement + Married + YearsAtCompany + TrainingTimesLastYear +  
##     EnvironmentSatisfaction + JobSatisfaction + WorkLifeBalance,  
##     family = binomial, data = TrainvsProdv3)  
##  
## Deviance Residuals:  
##      Min       1Q   Median       3Q      Max  
## -0.7621  -0.6088  -0.5570  -0.4939   2.2696  
##  
## Coefficients:  
##              Estimate Std. Error z value Pr(>|z|)  
## (Intercept)    -1.526e+00  3.775e-01  -4.044  5.26e-05 ***  
## MonthlyIncome     8.924e-07  8.765e-07   1.018  0.30861  
## Male             2.824e-01  8.807e-02   3.206  0.00134 **  
## Education        -9.320e-02  4.064e-02  -2.293  0.02182 *  
## Age              -7.828e-03  4.913e-03  -1.593  0.11108  
## JobInvolvement    3.030e-02  5.941e-02   0.510  0.61012  
## Married           5.290e-02  8.452e-02   0.626  0.53141  
## YearsAtCompany   -1.194e-03  7.415e-03  -0.161  0.87205  
## TrainingTimesLastYear -4.782e-02  3.291e-02  -1.453  0.14619  
## EnvironmentSatisfaction 4.002e-02  3.858e-02   1.037  0.29953  
## JobSatisfaction    1.060e-01  3.879e-02   2.731  0.00631 **  
## WorkLifeBalance   -8.337e-02  5.861e-02  -1.422  0.15491  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## (Dispersion parameter for binomial family taken to be 1)  
##  
##      Null deviance: 3785.1  on 4409  degrees of freedom  
## Residual deviance: 3750.5  on 4398  degrees of freedom  
## AIC: 3774.5  
##  
## Number of Fisher Scoring iterations: 4
```

Only Gender, Job Satisfaction, and Education are significant.

Model 4:

```
PerfRatevsGenEduJobSatis = glm(PerformanceRatingv2 ~ Male + Education + JobSatisfaction, family = binomial,  
summary(PerfRatevsGenEduJobSatis)
```

```
##  
## Call:  
## glm(formula = PerformanceRatingv2 ~ Male + Education + JobSatisfaction,  
##     family = binomial, data = TrainvsProdv3)  
##  
## Deviance Residuals:  
##      Min       1Q   Median       3Q      Max  
## -0.6982  -0.6114  -0.5576  -0.5076   2.1411
```

```
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.91977    0.17369 -11.053  < 2e-16 ***
## Male          0.29393    0.08773   3.350 0.000807 ***
## Education    -0.09348    0.04050  -2.308 0.020995 *
## JobSatisfaction 0.10798    0.03874   2.787 0.005318 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 3785.1  on 4409  degrees of freedom
## Residual deviance: 3760.0  on 4406  degrees of freedom
## AIC: 3768
##
## Number of Fisher Scoring iterations: 4
```

```
exp(PerfRatevsGenEduJobSatis$coefficients)
```

```
##           (Intercept)           Male           Education JobSatisfaction
##           0.1466414         1.3416931         0.9107607         1.1140218
```

Drop in Deviance:

```
anova(PerfRatevsGenEduJobSatis, PerfRatevsAllwoDepart, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: PerformanceRatingv2 ~ Male + Education + JobSatisfaction
## Model 2: PerformanceRatingv2 ~ MonthlyIncome + Male + Education + Age +
##           JobInvolvement + Married + YearsAtCompany + TrainingTimesLastYear +
##           EnvironmentSatisfaction + JobSatisfaction + WorkLifeBalance
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         4406      3760.0
## 2         4398      3750.5  8    9.4572  0.3052
```

P-value >0.05 -> We cannot reject the null hypothesis. Model 4 is a better model to predict performance.

Reference:

- Al-Mzary, M, et al., (2015). Training and its Impact on the Performance of Employees at Jordanian Universities from the Perspective of Employees: The Case of Yarmouk University . Journal of Education and Practice , 6(32). Retrieved from <https://files.eric.ed.gov/fulltext/EJ1083504.pdf>
- Imran2, A., & Imran, A. (2013). The Effect of Training on Employee Performance . European Journal of Business and Management , 5(4), 137–147. Retrieved from <https://pdfs.semanticscholar.org/354c/2c8c60f37f5e25f63f557b3573ec366197ae.pdf>
- Jennifer, P., Brian, C., & Steven, F. (2013). Return on investment for workplace training: the Canadian experience, 17(1). Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1111/ijtd.12002>
- Smailan, Abdulwahab. (2016). The Relationship Between Job Satisfaction, Job Performance, and Employee Engagement: An Exploratory Study. Retrieved from <https://pdfs.semanticscholar.org/9916/a165d256e4622b7e494389cde7edf1767c3e.pdf>.