**3-2 Assignment: Establishing a Baseline and Hypothesizing**

Kayla Sacks

Southern New Hampshire University

DAT-430 Leverage Data for Org Results

Professor Aubrey Condor

September 22, 2024

## Introduction

The goal of this analysis is to understand if the amount of training provided effects employee performance or job satisfaction. From the results of the analysis, we can learn if more training will have an affect on job satisfaction and employee performance. The three variables that will be used for the analysis are ‘training’ which is the independent variable and the two dependent variables, ‘JobSatisfaction’ and ‘PerformanceRating’.

## Analysis

The first step in creating the model is loading the libraries and data. It is also important to check that the data was read successfully by checking the dataset information and the first few rows as seen below where we can still see that there are 1000 rows and 35 columns as well as how many of each data type. Once the data was loaded, I checked the range for each variable.

A screenshot of a computer

Description automatically generated

A screen shot of a computer

Description automatically generated

Now we know the range of training (0-6), job satisfaction (1-5), and employee performance (1-5). Based on the variable ranges we can hypothesize that the more training that employees receive, the better their performance and the higher their job satisfaction. The next step I took was to find the correlation between the variables. Doing this allows me to see if there is a linear association between the variables. In the image below, we can see that the correlation between training and both dependent variables is above .9 which shows there is a strong positive correlation. With the strong correlation between the variables, I decided to create 2 regression models, one for each dependent variable, to show their relation to the independent variable of ‘training’. After creating the models, I found the root mean squared error (RMSE) and values to check the accuracy of the model before plotting.

A screenshot of a computer

Description automatically generated

*NOTE: For the models, I rounded the predicted values integers and changed any 0 values to 1 as the rating can only be provide as whole numbers from 1-5 for both job satisfaction and employee performance. This will make the predictions more informative as they not show values that are not possible to enter for the variables. This also means that the model will be a Non-Linear Regression model.*

A screenshot of a computer program

Description automatically generated

The lower RMSE the more accurate the model and the closer value is to 1, the better the fit between the model prediction and actual observation (Kumar, 2023). Finally, I plotted both models to visualize the data, which can be seen below.

A screenshot of a computer screen

Description automatically generated

## Conclusion

The hypothesis for this analysis after we found the range of each variable was that increased training will also increase job satisfaction rating and employment performance rating. Though the analysis, we found that job satisfaction has a .93 correlation and performance has a .90 correlation with training. We then found that the RMSE for both models were below .7 and the values were .93 and .86 respectively. From this information and based on the plots, which have overlapping data points, we can conclude that the more training that employees receive, the better their performance and job satisfaction ratings will be. It is also important to note that no scores of 3 or 4 appeared either of the dependent variables so there is a possibility of an error with the data set that should be taken into consideration but overall the conclusion remains that our hypothesis was correct and an increase in training will also improve job satisfaction or employee performance.

## References

Kumar, A. (2023, December 29). *Mean Squared Error or R-Squared - Which one to use?* Analytics Yogi. <https://vitalflux.com/mean-square-error-r-squared-which-one-to-use/#:~:text=Differences%3A%20Mean%20Square%20Error%20(MSE>