The dummy coded variables that I considered are from the department:

Management, RandD and Sales

Table

Description automatically generated

Table

Description automatically generated

***Multicollinearity***-The pearson correlation between the different variables determines the multicollinearity status. If the coefficient value is greater than 0.8 then the multicollinearity assumption is violated. The above table consists of the coefficient values ranging from -0.145 to +0.459 Therefore the values are less than 0.8 hence multicollinearity assumption is satisfied.

Additionally, to strengthen the multicollinearity status I had generated a coefficients table which consists of collinearity statistics in which tolerance and VIF are generated. If the tolerance value is less than 0.4 and if the variance integration factor is less than 10. The above tolerance values are less than 0.4 and the variance integration values are less than 10. Hence it satisfies the multicollinearity assumption.

**Figure-1**

Chart, histogram

Description automatically generated**Figure-2**Chart, line chart

Description automatically generated

The above figure-1 shows the normally distributed curve with satisfaction level plotted in which the satisfaction level frequency is plotted on Y-axis, and regression standardized residual is plotted on the Y-axis. In the above graph (Figure-1) the frequency distribution mostly follows the normal curve. Hence, the satisfaction level is mostly normally distributed.

Figure-2 shows the normal probability plot in which the dependent variable i.e., satisfaction level is plotted. The satisfaction level mostly follows the diagonal line as shown in figure 2. Hence, it satisfies the ***normality***.

**Figure -3**

Chart, scatter chart

Description automatically generated

The graph above shows the concentration of the points plotted on the satisfaction level is mostly linear across the graph and the points are mostly confined to a certain region that is linear and thus satisfies ***linearity***.

In figure-3 most of the points are held between the -3 to +3 while some of the points on the X-axis has distributed above the +3 and -3 points. To prove homoscedasticity the points, must be most of the points should be restricted to the region of -3 to +3 area. In the above graph we can see that most of the points have occupied the region between -3 to +3 area hence it satisfies ***homoscedasticity***.

Table

Description automatically generated

The promotion in the last 5 years variable explains the employees who got promoted in the organization in the last 5 years. If yes, it is denoted by 1 if not is denoted by zero. The satisfaction level explains the level of the employees' satisfaction with their work and their experience in the organization. The department variables give the info of the different branch names in the organization. The time spend in the company gives the amount of time that the employees spent in the company. Average monthly working hours give the number of hours the employee spent in the organization monthly.

The first model constitutes the dependent variable satisfaction level, and the independent variables work accident, promotion in last five years, dummy variables of sales, management, and RandD.

The second model I considered is the enhancement of the first model it constitutes the same variables as the first additionally it contains time spent in the company and average monthly hours.

The hierarchical multiple regression is tested for the two models that I considered above. The r-squared value for the first model is 0.004 which constitutes only 0.04 percent of the predicted model while the r squared value obtained for the second model is 0.031 which constitutes 0.31% of the predicted model. By adding predictor variables in the second model the percentage increased about 0.28% from the first predicted model.

The ANOVA test is performed between the two models to exact the statistical significance status between the two models. By performing the ANOVA test the result turned out that the two models are statistically significant to each other with the degrees of freedom of 5 for the first model and 8 for the second model.

Adding more predictor variables to model 1 is increasing the accuracy of the model to some extent but very far from the proposed model or deemed model. I recommend director of human resources include more number of predictor values since model-2 with more predictor values than model-1 is showing only 0.31% of the predicted model, it’s extremely low to recommend any structural changes in the management. Further, I recommend considering new models with different combinations of predicted variables that could potentially match the predicted model.