**Simple Linear Regression**

The **Number of Observations Read** and the **Number of Observations Used** are the same, which indicates that no missing values were detected for **Credit\_Limits** and **Months\_Inactive\_12\_mon**.

Pr > F is the *p*-value associated with the *F* value. The *F* value tests whether the slope of the predictor variable is equal to 0. The *p*-value is small (less than .05), so you have enough evidence at the .05 significance level to reject the null hypothesis. Thus, you can conclude that the simple linear regression model fits the data better than the baseline model. In other words, **Months\_Inactive\_12\_mon** explains a significant amount of variability of **Credit\_Limits**.

**Pr > |t|** is the *p*-value associated with the *t* statistic. It tests whether the parameter associated with each predictor in the model is different from 0. For this example, the slope for the predictor variable is statistically different from 0. Thus, you can conclude that the predictor variable explains a significant portion of variability in the response variable.

Because the estimate of bo=9061.33874 and b1=-183.40640, the estimated regression equation is given by: **Credit\_Limits** =9061.33874 -183.40640\* (**Months\_Inactive\_12\_mon**)

The model indicates that a one-unit greater value for **Months\_Inactive\_12\_mon** is associated with a 183.40640 lesser value for **Credit\_Limits**. However, ***extrapolation of the model beyond the range of your predictor variables is inappropriate***. You cannot assume that the relationship maintains in areas that were not sampled from.

**Table

Description automatically generated**

**Chart, box and whisker chart

Description automatically generated**

**Multiple Linear Regression Model**

A common correlation statistic used for continuous variables is the Pearson correlation coefficient, which is shown in this result. Pearson correlation coefficients are between -1 and 1 and are closer to either extreme if there is a high degree of linear association between the two variables. It is greater than 0 if there is a positive linear association and less than 0 if there is a negative linear association.

Correlations

|  |  |  |
| --- | --- | --- |
| Total\_Trans\_Ct and Total\_Trans\_Amt | 0.80719 | very strong |
| Total\_Revolving\_Balance and Avg\_Utilization\_Ratio | 0.62402 | strong |
| Months\_on\_book and Customer\_Age | 0.78891 | strong |
| Avg\_Open\_to\_buy and Avg\_Utilization\_Ratio | -0.53881 | moderate |

Graphical user interface, application

Description automatically generated

Graphical user interface, application, table

Description automatically generated

Graphical user interface, application, table, Excel

Description automatically generated

Now we build a multiple linear regression analysis of **Credit\_Limit** on all numeric variables, except **Total\_Trans\_Ct, Total\_Revolving\_Balance** and **Months\_on\_book** because they are strongly correlated with **Total\_Trans\_Amt**, **Avg\_Utilization\_Ratio** and **Customer\_Age**, respectively. (The choice of variables to be excluded due to high correlation is arbitrary here.)

Pr > F is small. Therefore, we reject the null hypothesis H0: b1=b2=0 and conclude that at least one bi10.

The R2 for this model is 0.9962 and the adjusted R2 is 0.9962, higher than the adjusted R2 of 0.0003for the **RunTime** only model. This suggests that this multiple linear regression model fits better than the **RunTime** only model. The p-value of individual predictors indicate that **Customer\_Age**, **Total\_Relationship\_count**, **Months\_inactive\_12\_mon**, **Avg\_Open\_to\_Buy**, **Total\_Amt\_Chng\_Q4\_Q1**, **Total\_Trans\_Amt**, **Avg\_Utilization\_Ratio** and **Naive\_Buyer\_Classifier\_Attririon** are statistically significant for predicting **Credit\_Limit** (with their p-values being 0.0399, 0.0371, 0.0936, <.0001, 0.0543, <.0001, <.0001 and <.0001, respectively.)

We can see that all Variance Inflation Factor (VIF) values are lower than 5, which means that multicollinearity is not a concern in the current model (mainly due to the fact that we have already taken three predictors out because of high pairwise correlation coefficients.)

Graphical user interface, text, application

Description automatically generated