CJ Kieler

David Lee Miller II

Tony Zhang

MBA 8040 Data Driven Decisions

August 1st, 2019

Professor Sub

Freight Charges

**Introduction:**

The data for our regression analysis is taken from the company where CJ Kieler is currently employed as a logistics and supply chain manager. In this position, CJ works with various freight companies to ship products throughout the United States - easily exceeding one million dollars worth of freight a year. Unfortunately, CJ began to notice that one of their main shipping vendors, SAIA, has frequently been charging more for their shipping service after the products were shipped than originally quoted.

CJ has since contacted SAIA to resolve the issue, but SAIA denied his claims and refused to accept responsibility. On the other hand, however, this problem does not happen with every shipment that CJ employs SAIA to transport, and he did not want to discontinue business with the vendor because they are generally the cheaper option outside of these negative random occurrences. Therefore, it is our goal to utilize this analysis to attempt to isolate what has the highest correlation with the change in quoted shipping costs and invoiced shipping costs in order to see where his company can still utilize SAIA's cheaper rates, but avoid using them to ship products associated with the excess charges.

Taking this into consideration, our independent variables in this analysis will be the total weight, price per pound, volume, density, height, the assigned freight class, the value of the product(s) being shipped, and the quoted price of shipping. Our dependent variable is the invoice price following delivery. The "total weight" is broken up per pallet and is measured in pounds both before and after shipping. The "volume" of shipments is important because they will charge excessive length fees once the length gets past a certain point. "Density" can cause additional fees because the vendors will charge if we have very large pallets with little to no density, believing that they are losing money because of what we are shipping and try to recoup by charging this fee. Assigning a "freight class" gives the vendor an idea of what is being shipped from CJ's facility and allows them to know if they will be shipping hazardous materials or any other circumstances needing to be considered while handling the shipment.

All of this information is gathered using a software called Freight View. With this program, CJ's company can send out freight quotes for pricing and then audit the invoiced amount that comes back after delivery. Analytics can be obtained on a weekly, monthly, or yearly basis and compared against the quoted price and the data is sent to an excel sheet where it is continually monitored for changes in charges and where we can improve with our shipments.

**Analysis:**

For this analysis, the free software R was used to run a regression analysis, while Excel was used for descriptive statistics, correlations, and scatterplots of the dependent and independent variables. The reason R was used for regression analysis is because of the way it handles categorical variables. While performing linear regression in R, R automatically creates all the dummy variables for each categorical variable in the regression output. After the raw data from the Excel spreadsheet was imported into R, all of the variables had a default numeric data type. Our only categorical variable, "Freight Class", has 10 different categories. This means we would have a total of 9 dummy variables and, in the context of regression analysis, categorical variables must be specified as factors in the R environment.

A factor is a data structure in R that consists of different "levels," with each level representing a different category of the variable. The reference level for our "Freight Class" variable is "300." This means that in the regression equation, the coefficients of all the 9 dummy variables would be multiplied by 0 if we are predicting the price of a shipment with a Freight Class of "300." This resulted in the full (initial) model consisting of all of the independent variables from the beginning to be written as follows:

$$\hat{Y}\text{(Predicted Invoice Amount)} = 158.149 + 1.145\times\text{(Quoted Price)} + 0.0194\times\text{(Total Weight)} - 79.61\times\text{(Price per Pound)} - 0.299\times\text{(Volume)} - 3.953\times\text{(Density)} - 0.503\times\text{(Height)} + 0.00251\times\text{(Product Cost)} - 31.098\times\text{(Freight Class 60)} - 57.549\times\text{(Freight Class 65)} - 61.003\times\text{(Freight Class 70)} - 79.263\times\text{(Freight Class 85)} - 85.213\times\text{(Freight Class 92.5)} - 74.863\times\text{(Freight Class 100)} - 85.582\times\text{(Freight Class 125)} - 67.093\times\text{(Freight Class 175)} - 64.440\times\text{(Freight Class 250)}.$$

After the initial model was given by R, we began to drop variables with the single highest p-value in the regression output and ran a new model without the insignificant variables – one at a time. This process was repeated a total of 4 times before we obtained our final model. First, the level of "60" in the Freight Class variable was dropped because it had the highest p-value of 0.434. This resulted in 5 observations being deleted as there are 5 observations in the dataset with a level of "60" for Freight Class. After this revised model was run, we then dropped the "Product Cost" variable with a p-value of 0.520. Next, Height was dropped with a p-value of 0.185. Finally, "Density" was dropped with a p-value of 0.175 and we ended with the following final model:

$$Y\text{(Predicted Invoice Amount)} = 77.470 + 1.098\times\text{(Quoted Price)} + 0.0187\times\text{(Total Weight)} - 50.094\times\text{(Price per Pound)} - 0.213\times\text{(Volume)} - 87.816\times\text{(Freight Class 65)} - 67.099\times\text{(Freight Class 70)} - 73.374\times\text{(Freight Class 85)} - 72.660\times\text{(Freight Class 92.5)} - 53.119\times\text{(Freight Class 100)} - 59.628\times\text{(Freight Class 125)} - 38.962\times\text{(Freight Class 175)} - 40.673\times\text{(Freight Class 250)}.$$

This leaves the final model with an F-value of 628.9 and a corresponding p-value/significance of less than 2.2E-16, meaning that the final model is significantly different than a model with no independent variables and only the intercept term. The R-squared value is 0.9703, meaning that 97.03% of the variance in the invoiced amount is explained by these independent variables. Next, the coefficients of each independent variable were interpreted. All of the following interpretations are only true when all other independent variables are held constant: the invoiced amount increases by $1.098 for each 1-dollar increase in the "Quoted Price"; the invoiced amount increases by $0.0187 for each 1-pound increase in the "Total Weight"; the invoiced amount decreases by $50.094 for each 1-dollar increase in the "Price per Pound"; and the invoiced amount decreases by $0.213 for each 1-unit increase in "Volume". However, the interpretation of the categorical variable "Freight Class" is a little different, with all of the following interpretations being compared to the baseline category of "300": shipments in the Freight Class category "65" decrease the invoiced amount by $87.816 on average compared to shipments in the Freight Class category "300," shipments in the Freight Class category "70" decrease the invoiced amount by $67.099 on average, shipments in the Freight Class category "85" decrease the invoiced amount by $73.374 on average, shipments in the Freight Class category "92.5" decrease the invoiced amount by $72.660 on average, shipments in the Freight Class category "100" decrease the invoiced amount by $53.119 on average, shipments in the Freight Class category "125" decrease the invoiced amount by $59.628 on average, shipments in the Freight Class category "175" decrease the invoiced amount by $38.962 on average, and lastly shipments in the Freight Class category "250" decrease the invoiced amount by $40.673 on average.

**Conclusion:**

Ultimately, our analysis found that most variables of the variables associated with CJ's shipping process caused a negative impact to the predicted invoiced amount, thus reducing the estimated invoice cost. However, as previously stated, this analysis was performed with categorical variables with "Freight Class 300" being the baseline with all other freight categories also further reducing the invoice price. Therefore, we conclude that our data suggests that "Freight Class 300" shipments are more likely to be the main variable in shipping that causes SAIA to charge higher than their quoted price.