QANT 530 Quiz 1

Phan Nguyen - 770034391

07/27/2022

This dataset represents stores of a company, where each row is a seperate stores. There are 150 rows with four variables For this exercise, the response variable is AverageMonthlyOrders. Below is a sample of the dataset.

Region is categorical and assumably nominal. Number of employees, inventory value and average monthly orders are continuous.

Table 1: Dataset example

```
##   Region NumberOfEmployees InventoryValue AverageMonthlyOrders
## 1      1              4.00            690                  277
## 2      1              2.07            786                  254
## 3      1              4.71            765                  244
## 4      1              5.32            767                  259
## 5      1              5.69            729                  238
## 6      1              6.83            688                  245
```

Scatterplots between response and explanatory variables

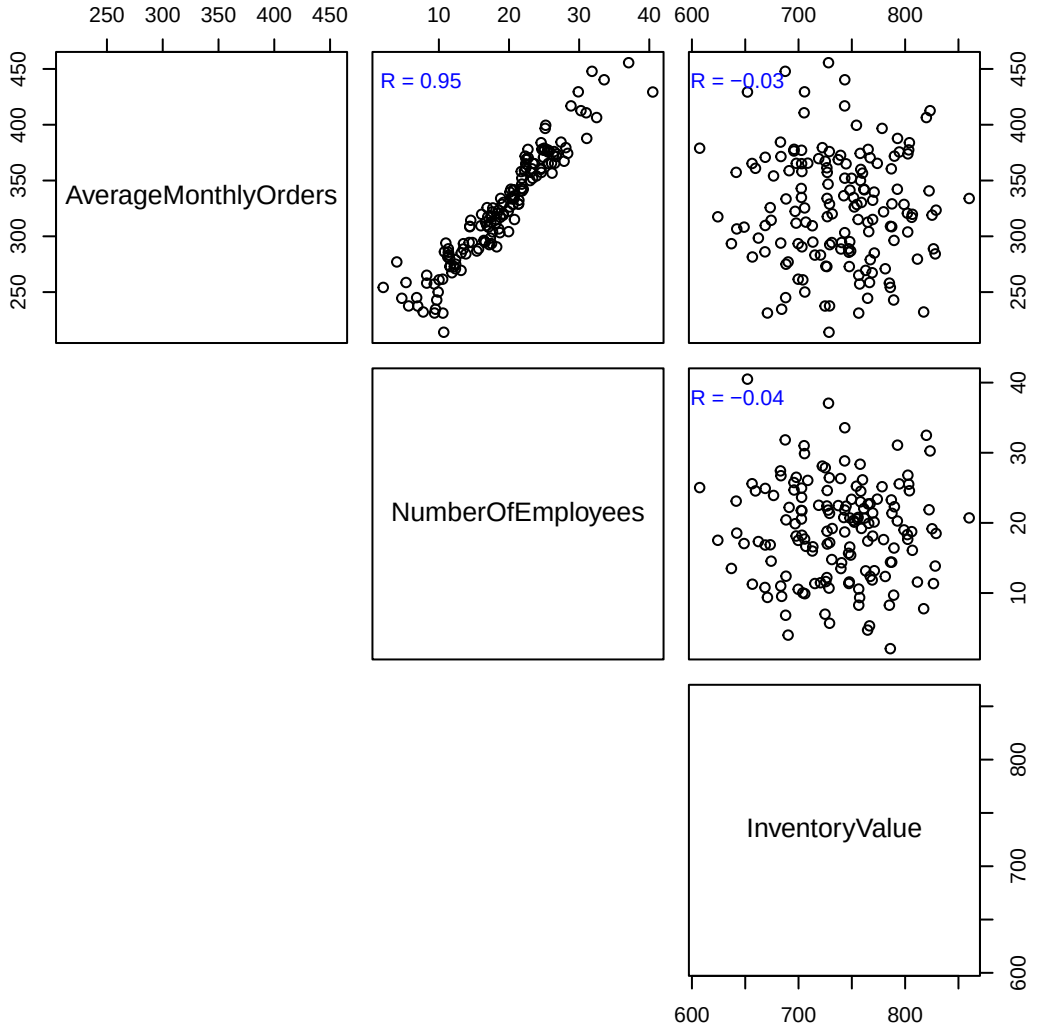

Figure 1: Scatterplot Matrix

Since Region is a nominal variables, we will exclude it out of our analysis. Figure 1 show a scatteplot matrix and Pearson correlation coefficient for each pairs of variables.

We can see that Average Monthly Order and Number of Employees have a positive linear relationship. The correlation coefficient is 0.95, which is almost 1, suggest a very strong correlation.

On the other hand, Average Monthly Orders and Inventory has weak correlation. The data points are scattered, and they have a near 0 correlation coefficient.

Correlation Matrix between explanatory variables

Table 2: Correlation Matrix

```
##                     NumberOfEmployees InventoryValue
## NumberOfEmployees               1.000         -0.043
## InventoryValue                 -0.043          1.000
```

There is weak correlation between Number of Employees and Inventory Values. The correlation coefficient is -0.43, according to Table 2. This means that the two variables are not correlated, and we can safely conclude that there is no multicollinearity problem.

Regression Coefficients

We ran multiple regression with AverageMonthlyOrders as our response variable, and NumberOfEmployees and InventoryValue as our explanatory variables.

Table 3: Unstandardized coefficient table

```
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          190.698     19.746   9.657    0.000
## NumberOfEmployees      6.797      0.184  36.950    0.000
## InventoryValue         0.008      0.026   0.311    0.756
```

Table 3 shows the unstandardized regression coefficients from the multiple regression model we ran. 190.698 indicates the average monthly orders for a store with no employee and no inventory value. For each additional employee and no change in inventory value, the average monthly orders increases around 6.797 orders. For every one unit of inventory value increased with no adjustment in the labor force, the average monthly orders increases around 0.0081.

These coefficients show us the rate of change of average monthly orders based on different factors. However, they don't tell us which factor have a bigger effect on the order numbers. Since they are in different units, a bigger number doesn't equal to greater significance.

Therefore, we are going to find the standardized regression coefficients.

Table 4: Standardized regression coefficient

```
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)            0.000      0.026   0.000    1.000
## NumberOfEmployees      0.951      0.026  36.950    0.000
## InventoryValue         0.008      0.026   0.311    0.756
```

The standardized regression coefficients are shown in table 4. Number of Employees has the largest number, meaning it has the highest effect on average monthly orders. 0.951 is the increase in standard deviation in average monthly orders if there is an increase of one standard devation in number of employees. With this logic, we can see that inventory value doesn't affect average monthly orders by much. Every one standard deviation increase only change the average monthly orders by 0.008 standard deviation. This also corresponds to what we found earlier from the correlation matrix, where inventory value and average monthly orders have low correlation coefficient.