Project 1: Predicting Catalog Demand

1. What decisions needs to be made?

The main decision should be made is, either to send the catalog to the chosen 250 customers or not.

2. What data is needed to inform those decisions?

The historical data of revenues, purchased items, chance of buying for each customer.

Also, data about states, cities, segment of customers will be helpful in order to test more variables. I have used segment.

Step 2: Analysis, Modeling, and Validation

1. How and why did you select the predictor variables in your model? You must explain how your continuous predictor variables you've chosen have a linear relationship with the target variable. Please refer back to the "Multiple Linear Regression with Excel" lesson to help you explore your data and use scatterplots to search for linear relationships. You must include scatterplots in your answer.

I have selected average number products purchased because it is a continuous predictor variable and it has a linear relationship with the target variable.
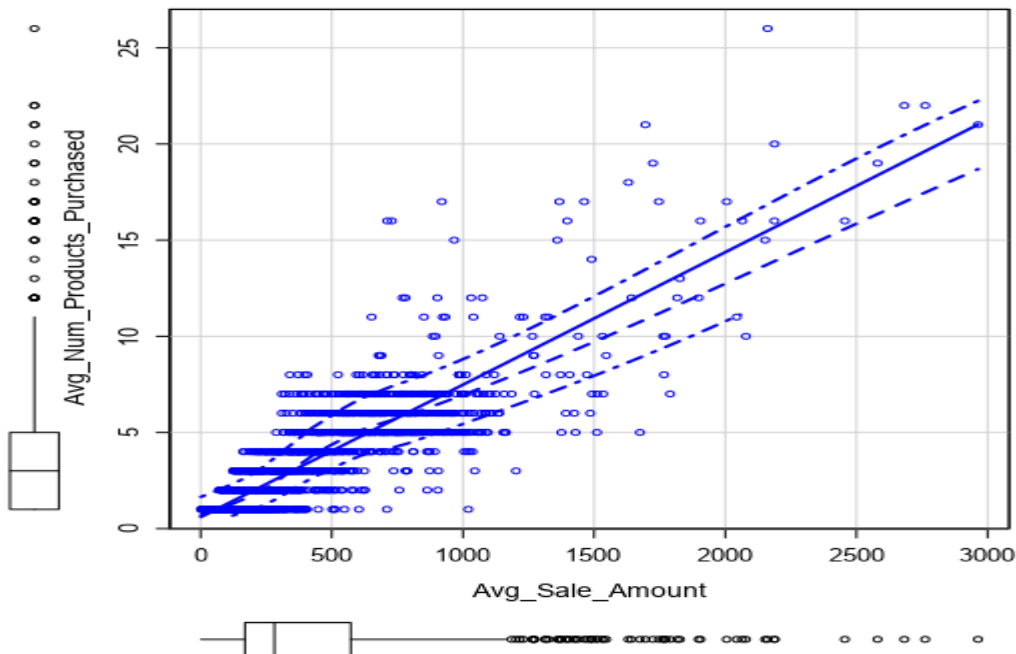
2. Explain why you believe your linear model is a good model. You must justify your reasoning using the statistical results that your regression model created. For each variable you selected, please justify how each variable is a good fit for your model by using the p-values and R-squared values that your model produced.

Average Number Products Purchased:
It has three starts which means less than 0.05 as p value.

Customer Segment with its all section (club, club and credit card and mailing list):
 It shows three starts means less than 0.05 as p value.

All of the above value have multiple r squared = 0.8369 and Adjusted R Squared = 0.8366

These values make my model strong and because of this I have chosen them.

3. What is the best linear regression equation based on the available data? Each coefficient should have no more than 2 digits after the decimal (ex: 1.28)

Y = 303.46 – 149.39 (club only) + 281.84 (club and credit card) – 245.42 (mailing list) + 66.98 (avg_num_products_purchased)

Note: the reason that I have separated the segments is because each one has its own p value and coefficient estimate.

Step 3: Presentation/Visualization

1. What is your recommendation? Should the company send the catalog to these 250 customers?

I recommend to send the catalog for the these 250 customers

2. How did you come up with your recommendation? (Please explain your process so reviewers can give you feedback on your process)

1 - I have chosen the linear regression model which is the suitable model for this situation
2 - Testing the given variables and chose the most suitable of them
3 - I set the target variable, which is the revenue, and then set the predictor variables, which is the average number of products purchased because of a linear relationship between them and the target variable.

4 - I chose the customer segment also as a predictor coefficient for the presence of strong statistical significance equivalent to the statistical significance of the first predictive coefficient

5 - I have appllied the model and generated the needed results and validate it again (I have validate it above as said in point 2 chosing the most suitable by looking for the statistical results)

6 - I calculated the predictable profit according to the revenues generated of the model

7 - According to the result I have provided my recommendation that support sending the catalog

3. What is the expected profit from the new catalog (assuming the catalog is sent to these 250 customers)?

According to the data provided and model used, the expected profit from the new catalog that will be send to 250 customers will be $21,987.4356