Joe Standerfer, Spriha Gupta, Jasmine Kaur, Daniel Lesser

Marketing Analytics 95-832

Assignment 3: Demand Analysis using Linear Regression

In a familiar tale, the traditional marketing approaches of retailers have been upended in the age of computers and big data. For many decades, the dominant marketing strategy for these companies had been “Hi-Lo Pricing” - the use of periodic discounts to drive traffic while keeping prices high on average. However, the introduction of computers enabled a new, seemingly more effective, strategy to be born. Often referred to “Everyday Low Pricing”, large retailers such as Walmart began to use consumer data to set product prices. By correlating price and to sale quantity along with several other predictive factors, these retailers were able to set the optimal profit maximizing price for each product. In this case study we compared both approaches using orange juice sales data from a Chicago based grocery chain, Dominick’s Finer Foods. Specifically, we analyzed Dominick's sales of 64oz Tropicana Premium Orange juice and developed a series of regression models for predicting sales quantity given a price point. \*\*\*\*summary of conclusions\*\*\*\*

The first regression model we created, estimated a sales quantity for the Tropicana orange juice using 3 factors: the log of the product’s sales price, the presence of in-store advertisements/ shelf-tags, and the existence of out-of-store advertisements. What we found was that the product demand is very elastic, or in other words the sales quantity responds strongly to changes in price. The coefficient we predicted for the price was around -2.02. At that rate we can estimate the demand will increase 2% for each single percent decrease in price. The other two factors, in-store and out-of-store advertisements showed some effect on demand, but of a much lower magnitude than price. Their coefficients were 0.535 and 0.077 respectively. This first model, however, was far from perfect. When we tested the model estimates on a hold-out test data set it achieved a mean-error of around 0.52. Charts of the model predictions as well as summary statistics can be seen in the Figures %%%%%%%%%%%%- %%%% in the Appendix.

To combat the model error and develop a better sales quantity prediction, we created 2 more regression models. The first of these utilized a separate log price coefficient for each of the 88 Dominick’s grocery stores.

For each model compute the mean-error, root mean-squared error, and mean-absolute deviation separately for the estimation and hold-out

**Appendix:**

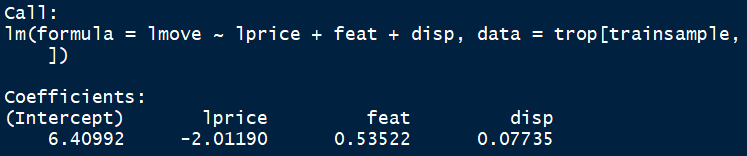


Figure 1: Pooled Model Regression Coefficients

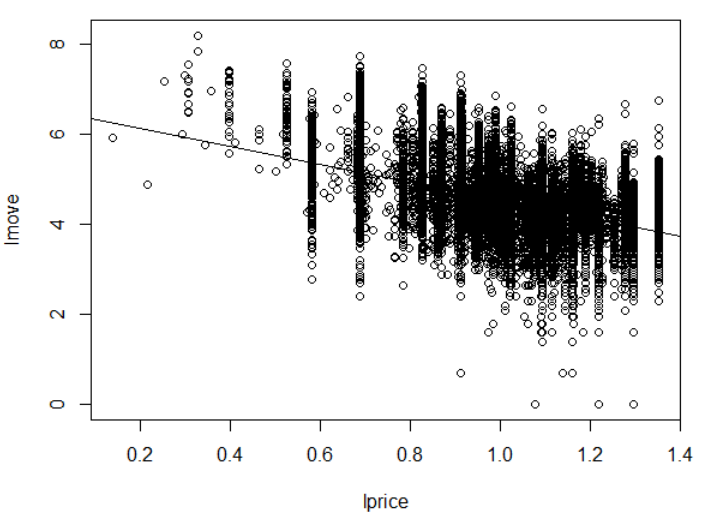


Figure 2: Pooled Model Predictions- All stores

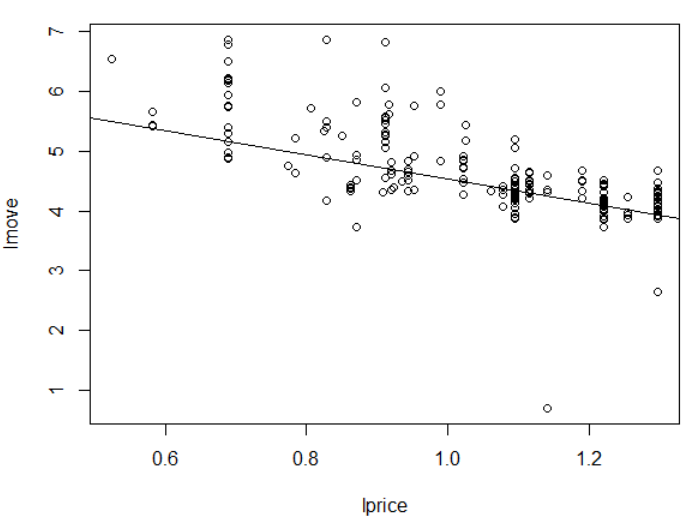


Figure 3: Pooled Model Predictions- Store 5

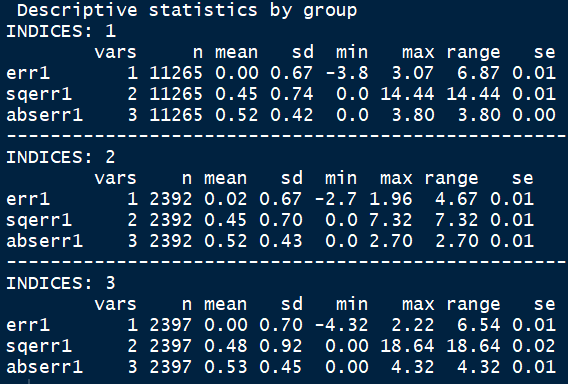


Figure 4: Pooled Model error