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 as “Everyday Low Pricing”, large retailers such as Walmart began to use consumer data to set product prices. By correlating price to sale quantity, along with several other predictive factors, these retailers were able to set a 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 for a given 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 log 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 estimates on a hold-out dataset 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 1 through 4 in the Appendix.

A large portion of the prediction error can be traced back to a base assumption made by the model. By utilizing a single log price coefficient, the model assumes that all of Dominick’s stores have similar demand curves. However, this is likely not the case. Price sensitivity for a store’s product can be affected by a variety of factors unique that store: surrounding demographics, closest competition, etc. To acknowledge that fact and thus lower the prediction error, we developed a second model that estimates distinct regression intercepts and log price coefficients for each store. As expected, the resulting prediction error for this “store model” is much lower than that of the original “pooled model”, at around 0.36. Figures 5 through 7 in the appendix provide details on this model’s performance.

To further improve the regression model, we decided to consider yet another significant factor in product sales – competition. When a commodity is priced higher than its peers, we can expect that consumers will defer to the cheaper option. This tradeoff is abundant within grocery stores, where comparable products are displayed side-by-side along with clearly labeled prices. In order to account for the in-store competition, our team collected the price of several products similar to the 64oz Tropicana Premium Orange juice. These were Minute Maid OJ 64 OZ (upc# 2500002606), Florida Gold Valenci OJ 64 OZ (upc# 1110000142), HH OJ 64 OZ (upc# 3828154001), and Tree Fresh OJ 64 OZ (upc# 7271850001). With those prices in hand, we computed the percent difference between the Tropicana juice’s price and the mean price of its competition. When this feature was added into the regression model, the prediction error was reduced to 0.33. A quick comparison between figure 6 and 8 illustrates the impact of our new feature. In Figure 6 you can see that the predictions made using just price coefficients (black) is very similar that those made with all the coefficients (blue). The gap is much wider in Figure 8, where the blue line includes our “competition” coefficient.

**Appendix:**

**Model 1 Figures**

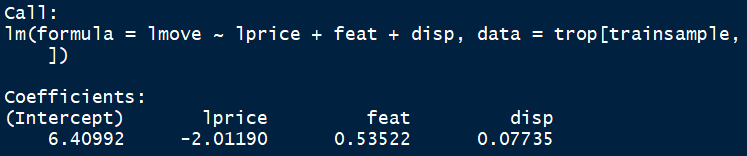


Figure 1: Pooled Model Regression Coefficients

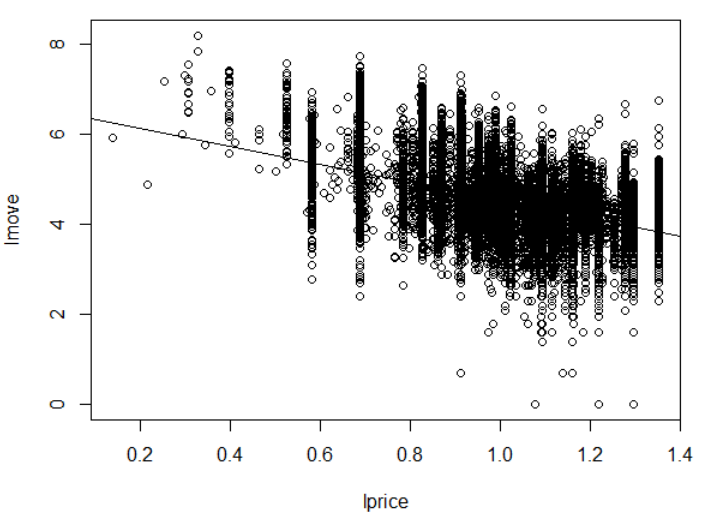


Figure 2: Pooled Model Predictions: All stores

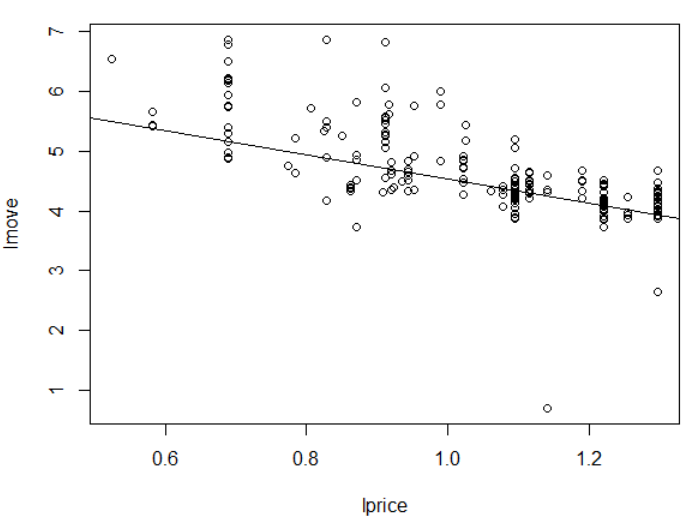


Figure 3: Pooled Model Predictions: Store 5

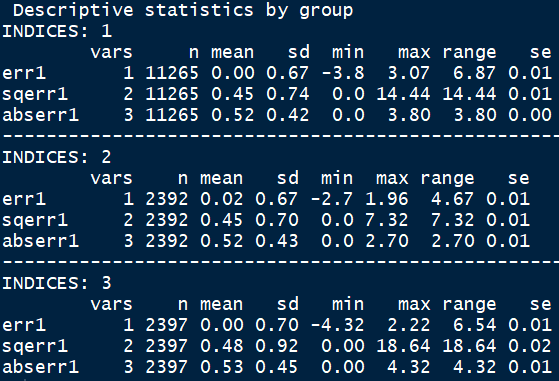


Figure 4: Pooled Model #1 Test Error

**Model 2 Figures**



Figure 5: Per-store Model #2 Regression Coefficients

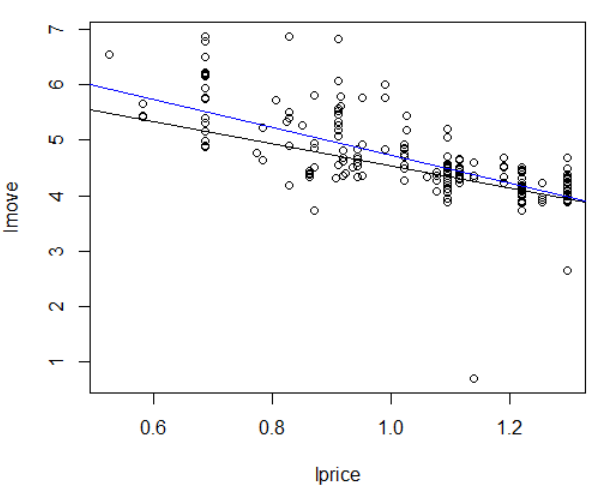


Figure 6: Per-Store Model #2 Predictions: Store 5 move predictions; price coefficent (black) vs all coefficients (blue)

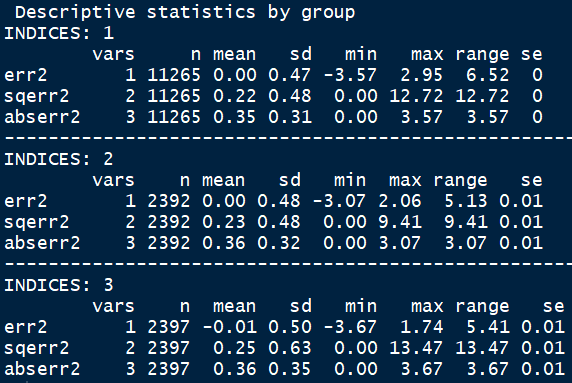


Figure 7: Per-Store Model #2 Test Error

**Model 3 Figures**



Figure 8: Improved Per-store Model #3 Regression Coefficients

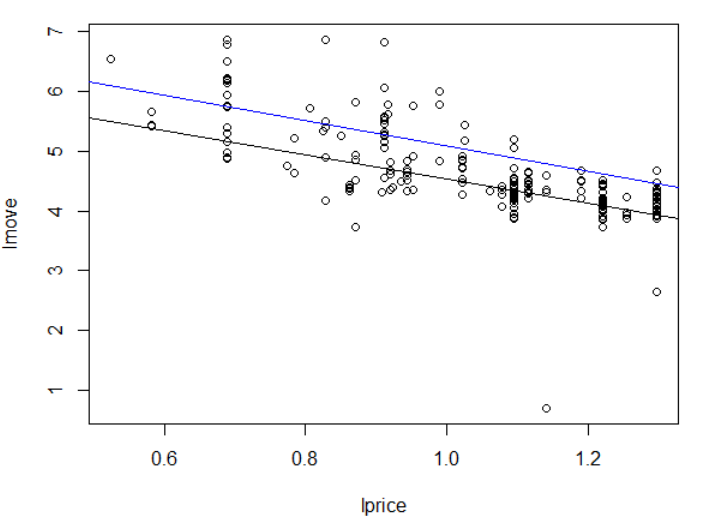


Figure 9: Model #3 Predictions - Store 5 move prediction; price coefficient (black) vs all coefficients (blue)

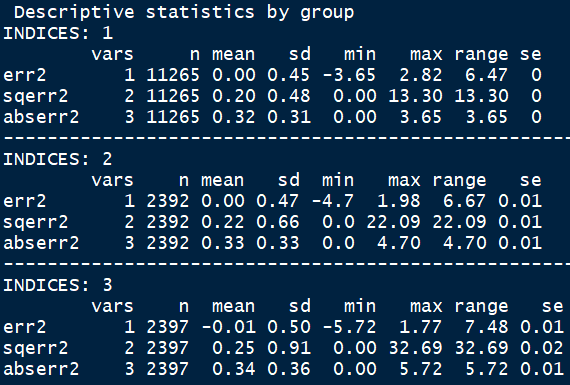


Figure 10: Improved Per-Store Model #3 Test Error