Joe Standerfer, Spriha Gupta, Jasmine Kaur, Daniel Lesser

Marketing Analytics 95-832

Assignment 3: Demand Analysis using Linear Regression

Throughout the 20th century, the incumbent marketing strategy in the retail grocery industry had been “Hi-Lo Pricing”. Grocers would keep prices high on average, while occasionally providing promotional discounts to drive traffic. However, the success of “Everyday Low Pricing” retailers in the past few years, such as Walmart, has cut into the profits of traditional grocers. Demand-based modeling provides an opportunity to combat these retailers and determine whether Hi-Lo Pricing or Everyday Low Pricing is the more effective strategy. By correlating price against quantity and several other predictive factors, there is an opportunity to set the optimal profit maximizing price for each product. In this case study we compared several 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.

The first regression model estimates the sales quantity of Tropicana orange juice using three factors: the log of the product’s sales price, the presence of in-store advertisements and shelf-tags, as well as the existence of out-of-store advertisements. Product demand appeared very elastic as sales quantity responded strongly to changes in price. The coefficient for the log of price was -2.02. This implies that unit sales will increase about 2% for each single percentage point decrease in price. The other two factors, in-store and out-of-store advertisements showed some effect on unit sales, but of a much lower magnitude than price. Their coefficients were 0.535 and 0.077, respectively. However, The accuracy of this model was not very robust. When tested against a hold-out test dataset, it achieved a mean-error of 0.53. Charts of the model predictions as well as summary statistics can be seen in the Figures 1 through 4.

A large portion of the prediction error can be traced back to a base assumption made by the first 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 to that store: surrounding demographics, closest competition, etc. To acknowledge that fact, a second model was used that estimates the distinct regression intercept and log price coefficient for each store. As expected, the prediction error for this “store model” is lower than those of the original “pooled model”, at 0.36. details on this model’s performance can be seen in Figures 5 through 7.

To improve the model further, we attempted to include the demographics of the stores – age, income, ethnicity, and education to our model. However, due to the fact we were already modeling at a store level, these variables were highly correlated to the ‘store’ attribute and hence, did not improve upon the model. It can be argued that ‘store’ already incorporates these demographic effects.

Our next avenue to improve the model, was to consider competition. By including the weekly prices of competing orange juice products for every store, the mean error fell to 0.34. Minute Maid, Florida Gold, and HH Orange Juice were a few of the products added for comparison. Figures 8 through 10 in the appendix provide details on this model’s performance. Adding the prices of other brands into the model considers competitive dynamic amongst the brands and the desire to gain market share, even if overall sales volume for the category remains stable or declines.

Another approach to considering the adequacy of the models is to display the residuals using a histogram. Figure 11 shows the histogram of the residuals of the Pooled model. The large deviations from actual values are expected as the model has a high mean error of 0.53. Figure 12 shows the residuals for the Store model. The deviations of the predicted values range from -0.5 to 1 and most of the deviations are in the range [-0.5, 0.5]. Given that this model includes store effects and has a mean error of about 0.36, we would expect the predictions to deviate less as compared to the Pooled model. Finally, Figure 13 shows the residuals of the third improved model. Most of the predicted values differ by 0.2 in absolute value from their actual values. Amongst the three models, the improved model has the least deviations from the actual quantities.

Figures 14 through 16 show the time series plots of the errors across weeks for store 2. The pooled model shows high variations in as compared to the store model. The improved model has errors which are more centered around 0. Both the histogram and time series plots align with the low mean error of the improved model and strengthen the case for the third model being superior. The implied optimal price for the log-linear demand model is given by:

where c = cost of the product. The optimal price for the pooled model across all stores is $4.35.

Figures 17 shows the computed optimal prices for each store in week #100 for the Store model and the improved model. The mean difference between the store model predictions and the actual values is -2.44. For the Improved model, the difference is -1.24. Both models imply Dominick’s should be charging more for this product. As Dominick’s Pricing Managers, we would implement the improved model for this case as it gives lower differences on average. However, additional constraints will be required in both models to avoid unreasonable prices, either negative or excessively high, which occurred in a few stores. This may be one limitation of using a simple regression model to predict prices.

**Appendix:**

**Model 1 Figures**

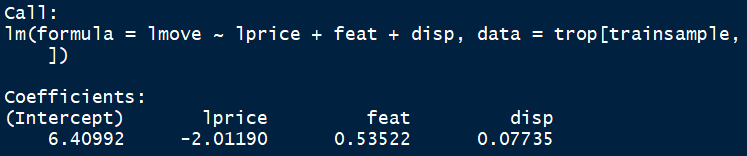


Figure 1: Pooled Model Regression Coefficients

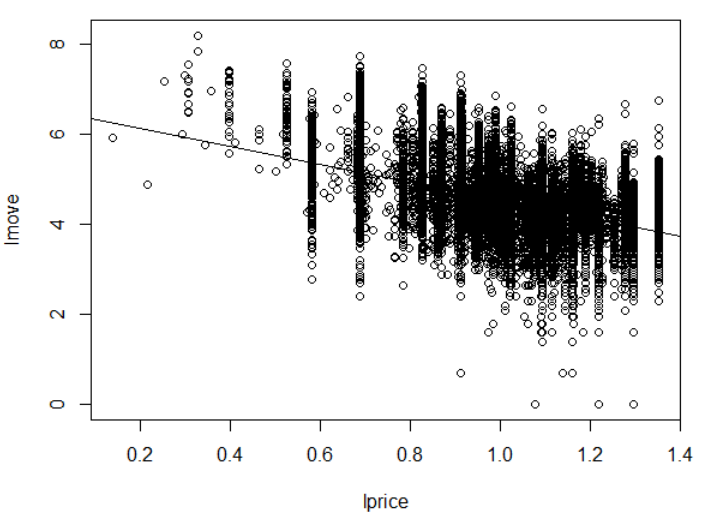


Figure 2: Pooled Model Predictions: All stores

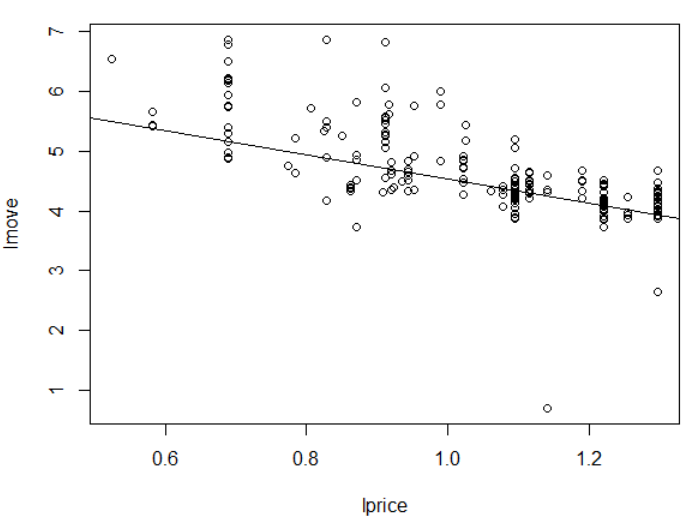


Figure 3: Pooled Model Predictions: Store 5

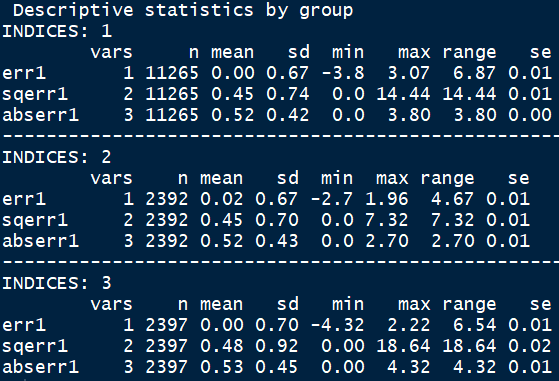


Figure 4: Pooled Model #1 Test Error

**Model 2 Figures**



Figure 5: Per-store Model #2 Regression Coefficients

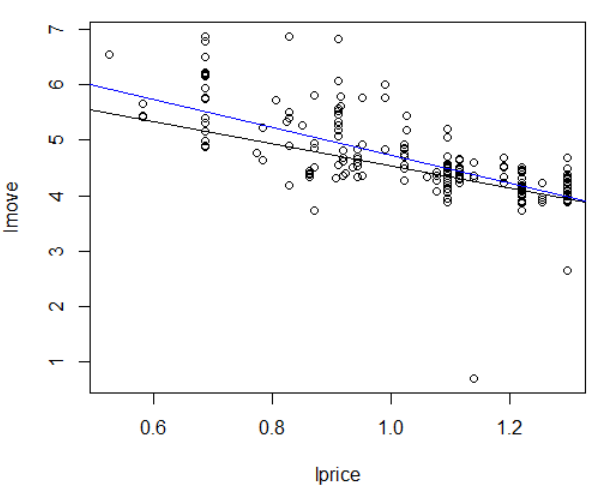


Figure 6: Per-Store Model #2 Predictions: Store 5 move predictions; price coefficent (black) vs all coefficients (blue)

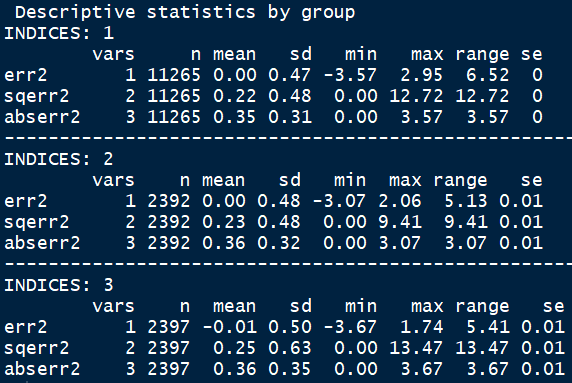


Figure 7: Per-Store Model #2 Test Error

**Model 3 Figures**



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

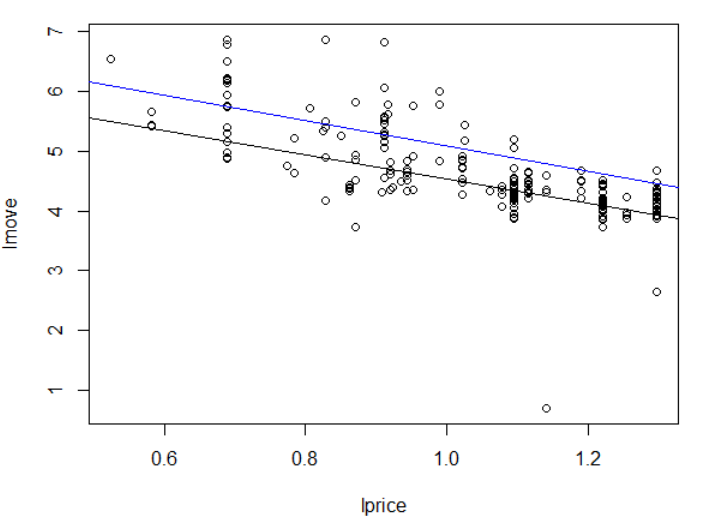


Figure 9: Model #3 Predictions - Store 5 move prediction; price coefficient (black) vs all coefficients (blue)

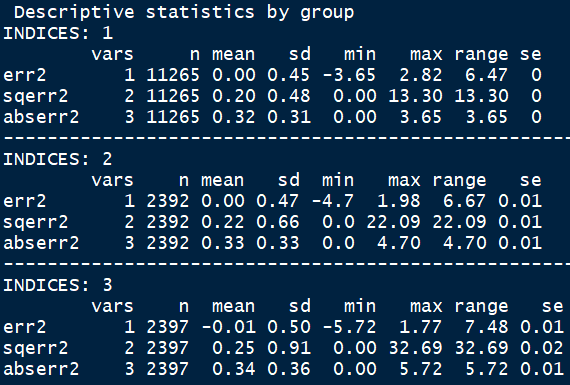


Figure 10: Improved Per-Store Model #3 Test Error

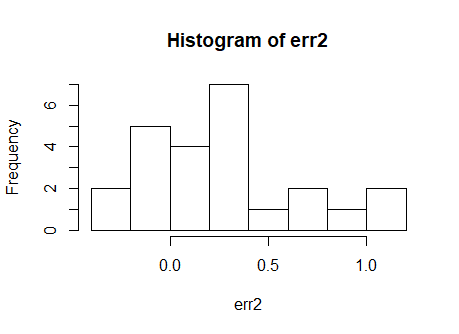


Figure 11: Histogram of errors of the Pooled model

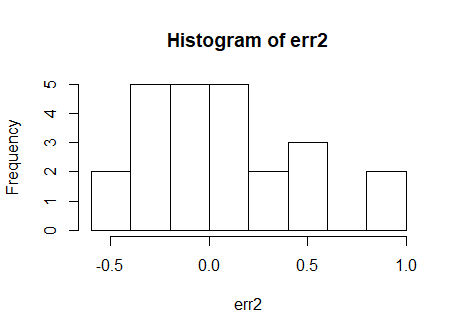


Figure 12: Histogram of errors of the Store model

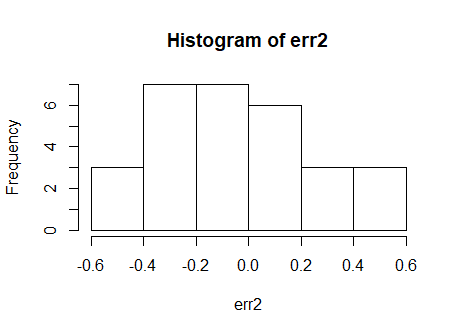


Figure 13: Histogram of errors of improved model

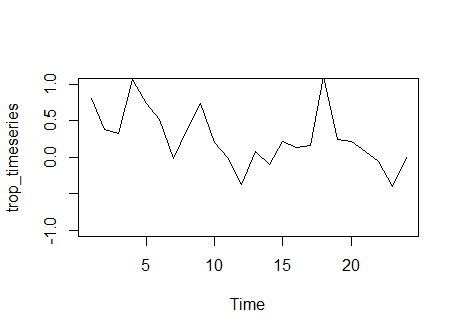


Figure 14: Pooled Model: Time Series plot of Store 2 price prediction errors

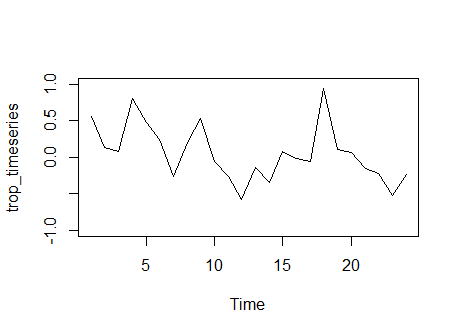


Figure 15: Store Model: Time series plot of store 2 price prediction errors

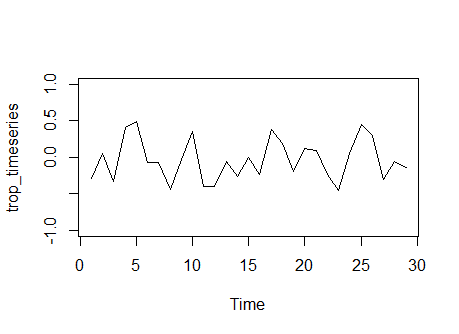


Figure 16: Improved Model: Time series plot of store 2 price prediction errors

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Store** | **Actual Price** | **Store Model Predicted Price** | **Store Model Difference** | **Improved Model Predicted Price** | **Improved Model Difference** |
| store5 | 2.19 | 4.18 | -1.99 | 4.00 | -1.81 |
| store8 | 2.19 | 3.66 | -1.47 | 3.96 | -1.77 |
| store9 | 2.19 | 4.27 | -2.08 | 6.15 | -3.96 |
| store12 | 2.19 | 3.37 | -1.18 | 4.06 | -1.87 |
| store14 | 2.19 | 5.59 | -3.40 | 11.62 | -9.43 |
| store18 | 2.19 | 3.78 | -1.59 | 6.43 | -4.24 |
| store21 | 2.19 | 3.55 | -1.36 | 3.52 | -1.33 |
| store28 | 2.19 | 4.52 | -2.33 | 8.69 | -6.50 |
| store32 | 2.19 | 7.01 | -4.82 | -35.05 | 37.24 |
| store33 | 2.19 | 5.15 | -2.96 | 9.17 | -6.98 |
| store40 | 2.19 | 3.32 | -1.13 | 3.48 | -1.29 |
| store44 | 2.19 | 3.71 | -1.52 | 5.01 | -2.82 |
| store45 | 2.19 | 6.30 | -4.11 | 8.02 | -5.83 |
| store47 | 2.19 | 3.98 | -1.79 | 5.22 | -3.03 |
| store48 | 2.19 | 5.04 | -2.85 | 19.55 | -17.36 |
| store49 | 2.19 | 4.44 | -2.25 | 5.95 | -3.76 |
| store50 | 2.19 | 4.12 | -1.93 | 5.18 | -2.99 |
| store51 | 2.19 | 4.10 | -1.91 | 9.19 | -7.00 |
| store52 | 2.19 | 4.41 | -2.22 | 7.41 | -5.22 |
| store53 | 2.19 | 4.18 | -1.99 | 6.01 | -3.82 |
| store54 | 2.19 | 4.37 | -2.18 | 4.53 | -2.34 |
| store56 | 2.19 | 21.03 | -18.84 | -4.99 | 7.18 |
| store59 | 2.19 | 3.42 | -1.23 | 3.87 | -1.68 |
| store62 | 2.19 | 8.73 | -6.54 | -46.64 | 48.83 |
| store64 | 2.19 | 3.75 | -1.56 | 4.63 | -2.44 |
| store67 | 2.19 | 4.50 | -2.31 | 6.72 | -4.53 |
| store68 | 2.19 | 4.25 | -2.06 | 4.71 | -2.52 |
| store70 | 2.19 | 3.73 | -1.54 | 4.82 | -2.63 |
| store71 | 2.19 | 3.61 | -1.42 | 4.38 | -2.19 |
| store72 | 2.19 | 5.18 | -2.99 | 6.32 | -4.13 |
| store73 | 2.19 | 4.69 | -2.50 | 5.36 | -3.17 |
| store74 | 2.19 | 4.28 | -2.09 | 4.97 | -2.78 |
| store75 | 2.19 | 4.25 | -2.06 | 5.65 | -3.46 |
| store76 | 2.19 | 3.24 | -1.05 | 4.06 | -1.87 |
| store77 | 2.19 | 4.00 | -1.81 | 7.54 | -5.35 |
| store78 | 2.19 | 3.47 | -1.28 | 3.62 | -1.43 |
| store80 | 2.19 | -17.70 | 19.89 | -8.91 | 11.10 |
| store81 | 2.19 | 4.09 | -1.90 | 5.05 | -2.86 |
| store83 | 2.19 | 3.35 | -1.16 | 4.17 | -1.98 |
| store84 | 2.19 | 4.37 | -2.18 | 6.28 | -4.09 |
| store86 | 2.19 | 4.01 | -1.82 | 4.00 | -1.81 |
| store88 | 2.19 | 21.92 | -19.73 | -16.33 | 18.52 |
| store89 | 2.19 | 3.62 | -1.43 | 4.00 | -1.81 |
| store90 | 2.19 | 3.23 | -1.04 | 3.45 | -1.26 |
| store91 | 2.19 | 3.87 | -1.68 | 4.26 | -2.07 |
| store92 | 2.19 | 5.42 | -3.23 | 6.35 | -4.16 |
| store93 | 2.19 | 4.83 | -2.64 | 6.76 | -4.57 |
| store94 | 2.19 | 15.59 | -13.40 | -91.90 | 94.09 |
| store95 | 2.19 | 3.35 | -1.16 | 4.19 | -2.00 |
| store97 | 2.19 | 3.81 | -1.62 | 3.89 | -1.70 |
| store98 | 2.36 | 3.78 | -1.42 | 4.83 | -2.47 |
| store100 | 2.38 | 4.34 | -1.96 | 6.81 | -4.43 |
| store101 | 2.39 | 7.93 | -5.54 | 11.86 | -9.47 |
| store102 | 2.51 | 4.57 | -2.06 | 4.09 | -1.58 |
| store103 | 2.43 | 3.86 | -1.43 | 4.34 | -1.91 |
| store104 | 2.19 | 3.63 | -1.44 | 3.93 | -1.74 |
| store105 | 2.33 | 3.25 | -0.92 | 3.85 | -1.52 |
| store106 | 2.19 | 3.86 | -1.67 | 4.32 | -2.13 |
| store107 | 2.19 | 4.63 | -2.44 | 4.90 | -2.71 |
| store109 | 2.19 | 4.53 | -2.34 | 8.47 | -6.28 |
| store110 | 2.19 | 5.34 | -3.15 | 19.17 | -16.98 |
| store111 | 2.19 | 3.19 | -1.00 | 3.39 | -1.20 |
| store112 | 2.37 | 10.24 | -7.87 | 16.69 | -14.32 |
| store113 | 2.19 | 4.29 | -2.10 | 5.28 | -3.09 |
| store114 | 2.39 | 3.56 | -1.17 | 3.96 | -1.57 |
| store115 | 2.39 | 3.72 | -1.33 | 4.04 | -1.65 |
| store116 | 2.19 | 4.98 | -2.79 | 18.74 | -16.55 |
| store117 | 2.19 | 4.47 | -2.28 | 5.12 | -2.93 |
| store118 | 2.19 | 4.22 | -2.03 | 4.87 | -2.68 |
| store119 | 2.19 | 6.94 | -4.75 | 15.37 | -13.18 |
| store121 | 2.35 | 3.67 | -1.32 | 4.85 | -2.50 |
| store122 | 2.31 | 7.63 | -5.32 | 13.89 | -11.58 |
| store123 | 2.19 | 3.57 | -1.38 | 3.55 | -1.36 |
| store124 | 2.19 | 4.01 | -1.82 | 4.43 | -2.24 |
| store126 | 2.34 | 3.19 | -0.85 | 3.42 | -1.08 |
| store128 | 2.38 | 3.84 | -1.46 | 4.93 | -2.55 |
| store129 | 2.42 | 4.08 | -1.66 | 4.60 | -2.18 |
| store130 | 2.19 | 3.10 | -0.91 | 3.50 | -1.31 |
| store131 | 2.42 | 4.53 | -2.11 | 7.04 | -4.62 |
| store132 | 2.39 | 3.74 | -1.35 | 3.77 | -1.38 |
| store133 | 0.00 | 2.97 | -2.97 | 4.11 | -4.11 |
| store134 | 2.37 | 3.68 | -1.31 | 4.87 | -2.50 |
| store136 | 0.00 | 3.86 | -3.86 | 13.54 | -13.54 |
| store137 | 2.19 | 3.47 | -1.28 | 4.00 | -1.81 |
| store139 | 2.19 | 2.99 | -0.80 | 3.17 | -0.98 |
| **Average** | 2.18 | 4.62 | -2.44 | 3.41 | -1.24 |

Figure 17: Computed Optimal prices for Store and improved models