A8 Retail Analytics

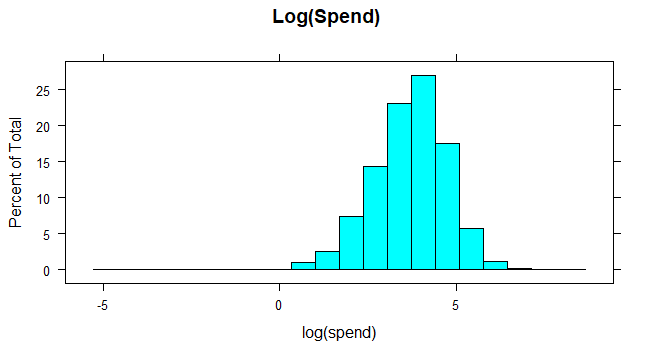
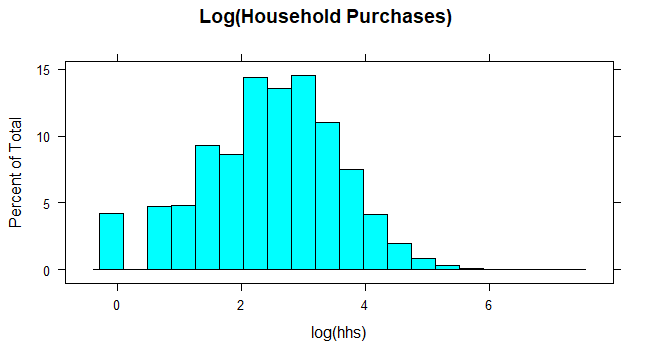
**Action** **Plan**:

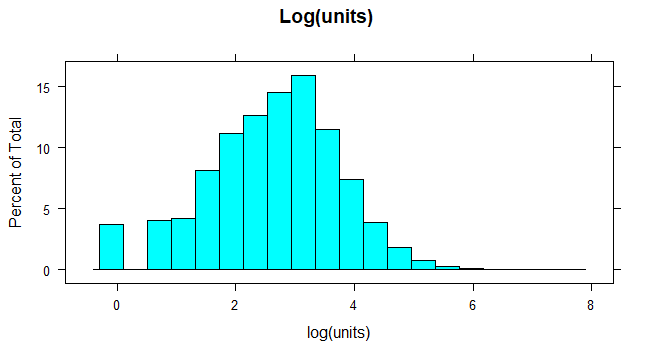
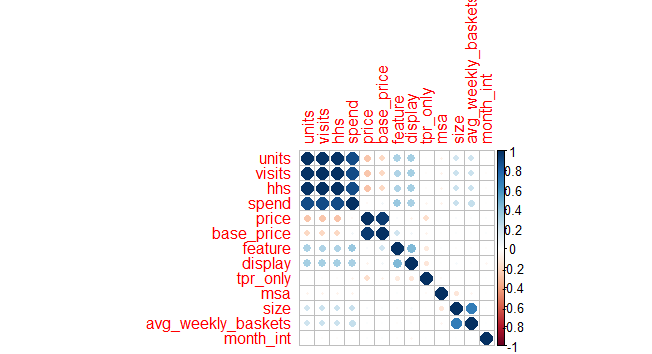
1. Explore Excel file and use filter feature to spot any useful patterns and understand variables
2. Read the data and merge into a main data frame using the
   1. Transactions table contains store\_id, and UPC, which are the primary keys for the stores and products tables, respectively, and will be used as keys to merge.
3. Describe data frame
   1. Descriptive stats and structure
   2. Test data quality
      1. Explore NA’s, duplicates, nonsensical observations(i.e. price= $0)
4. Fix all data types and engineer additional features if necessary
   1. Extract month and year from date
   2. Price difference variable computed from base price vs actual price
5. Visualize the data
   1. Histograms for all dependent variables
   2. Correlations between variables
   3. Optional: Boxplots of DV’s by product category and segment
6. Build models for sales(spend) DV
   1. Build a simple model for question 1
   2. Build an interaction model(segment/category for each promotion for question 2
   3. Build a model with an interaction effect between price and products to answer questions 3 and 4
   4. Evaluate Assumptions
7. Build 3 model for unit sales DV
   1. Build a simple model for question 1
   2. Build an interaction model(segment/category for each promotion for question 2
   3. Evaluate assumptions for the best model
8. Build 3 model for household DV
   1. Build a simple model for question 1
   2. Build an interaction model(segment/category for each promotion) for question 2
   3. Evaluate assumptions for the best model
9. Coefficient Interpretations and Recommendations

|  |  |  |
| --- | --- | --- |
| **Predictor** | **Effect** | **Rationale** |
| *Dependent Variables: Spend, Unit\_Sales, HHS* | | |
| Price | - | The price will have similar effect on all DV’s as more expensive items will sell less in units, dollar sales, and number of purchasing households |
| Tpr\_only | +/none | Temporary price reduction could boost sales slightly because it is not advertised, and customers may not know the reduction is in place and take advantage of it. |
| Feature | + | This should have a big effect on increasing all DV’s as high visibility combined with a possible discount is a well established sales boosting strategy |
| Display | + | This should have a big effect on increasing all DV’s as in-store circular combined with a possible discount will grab customers’ attention and boost sales. |
| Category | +/- | Some product categories sell differently than others |
| Segment | +/- | Different store types will have different clientele with different buying patterns |
| Upc | +/- | Control for product-level variance |
| Store\_num | +/- | Control for store-level variance |
| Month | +/- | Control for monthly time effect on DV’s |
| Year | +/- | Control for year over year effect on DV’s |
| Price\_diff | + | Larger discounts will lead to more sales |

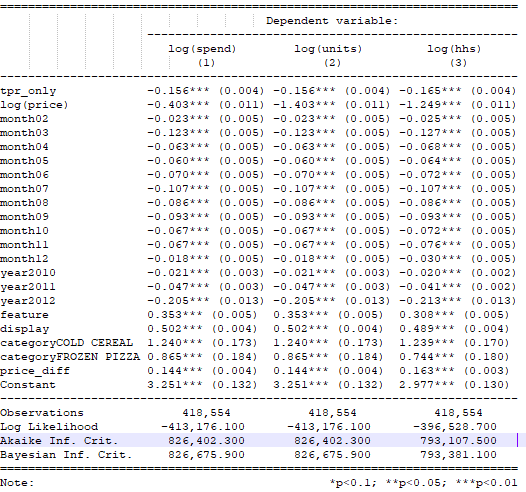
Visualization Summary:

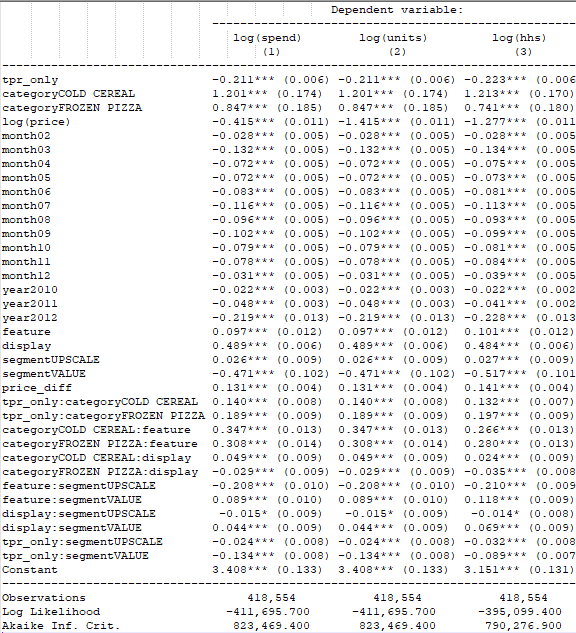
Distributions were heavily skewed for all three independent variables but looked more normal after a log transformation. Households and units are classic Poisson data as they are count, but I will use log(spend) in an OLS model for sales since it is continuous and very stabilized after the transformation. Correlations further support the predicted effect of most independent variables chosen for models. Boxplot visualizations further showed support the need to control for segments and categories as these highly affect the sales pattern.



The correct model for units and hhs should be glmer model with a poisson base because they are count data and will most likely violate the normality assumption of OLS. Due to limit in computing power and after multiple failed attempts, I computed the closest appropriate model I could, which is lmer using a log transformation on units and hhs to stabilize the variance. Assumptions are most likely violated, but our estimates should still yield valuable results.





1. What is the effect of promotions, displays, or being featured in the circular on product sales (spend), unit sales, and number of household purchasers? (3 points)
   1. These marginal effects were derived from models with no interaction effects
   2. Sales
      1. Dollar **sales** for products with a **temporary price reduction** were **15.6% lower** on average than products that did not have a reduction.
      2. Dollar **sales** for **featured** products were **35.3% higher** on average than non-featured products
      3. Dollar **sales** for products **displayed in the in-store circular** were **50.2% higher** on average than products not displayed on
   3. Units
      1. The marginal effects are identical for unit sales.
   4. HHS
      1. **Purchasing households** for products with a **temporary price reduction** were **16.5% lower** on average than products that did not have a reduction.
      2. **Purchasing households** for **featured** products were **30.8% higher** on average than non-featured products
      3. **Purchasing households** for products **displayed in the in-store circular** were **48.9% higher** on average than products not displayed on
2. How do the above effects vary by product categories (cold cereals, frozen pizza, bag snacks) and store segments (mainstream, upscale, value)? (2 points)
   1. Sales and Units yielded identical coefficients for all interaction terms as expected while households showed slight deviances from the first two models. Overall, all variables showed the same patterns with respect to product category and store segments as far as marginal effects of promotions. The assumptions below were based on dollar sales.
   2. Feature\*Category
      1. The feature promotion worked best for cold cereal products with dollar sales 34.7% higher than bag snacks and 4% higher than pizza products on average
      2. The dollar sales for feature promotion for Pizza products was 30.8% higher than Snacks on average
      3. The feature promotion was least successful on bag snack items
   3. Display\*Category
      1. The display promotion worked best for cold cereal products with dollar sales 7.8% higher than Pizza and 4.9% higher than bag snacks on average
      2. The dollar sales for the display promotion on Bag Snack products were 2.9% higher on average than Pizza products
      3. The display promotion was least successful on pizza products
   4. Tpr\_only\*Category
      1. The temporary price reduction promotion worked best for pizza products with dollar sales 4.9% higher than Cold cereal products and 18.9% higher than bag snack products
      2. The dollar sales for the temporary price reduction promotion on Cold cereal products were 14% higher than bag snack products
      3. The temporary price reduction promotion was least successful on Pizza products
   5. Feature\*Segment
      1. The feature promotion worked best for value stores with dollar sales 8.9% higher than Mainstream stores and 30% higher than upscale stores
      2. The dollar sales for the feature promotion for value stores was 20.8% higher on average than upscale stores
      3. The feature promotion was least successful in upscale stores
   6. Display\*Segment
      1. The display promotion worked best for value stores with dollar sales 4.4% higher than Mainstream stores and 5.9% higher than upscale stores
      2. The dollar sales for the display promotion for mainstream stores was 1.8% higher on average than upscale stores
      3. The display promotion was least successful in upscale stores
   7. Tpr\_only\*Segment
      1. The temporary price reduction promotion worked best for mainstream stores with dollar sales 2.4% higher than upscale stores on average and 13.4% higher than value stores
      2. The dollar sales for products with a temporary price reduction in Upscale stores was 11% higher than value stores
      3. The temporary price reduction promotion was least successful in value stores
3. What are the five most price elastic and five least price elastic products? Price elasticity is the change in sales for unit change in product price?
   1. If we model log(price) interaction effect with the product, the coefficients for the interaction term for each product will essentially be a measure of its elasticity. After ignoring the sign, since elasticity is always a negative value, the largest values represent the most elastic products while the smallest coefficient values represent the most inelastic products. Output was not included due to huge number of rows.



|  |  |  |  |
| --- | --- | --- | --- |
| Most Elastic Products | | | |
| UPC | Product Name | Category | Elasticity |
| 3800039118 | KELLOGS FROOT LOOPS | Cold cereal | 4.0 |
| 7192100336 | DIGIORNO THREE MEAT  PIZZA | Frozen Pizza | 3.9 |
| 7218063979 | Tony’s FRSC PEPPERONI PIZZA | Frozen Pizza | 3.8 |
| 2066200532 | NWMN OWN SUPREME PIZZA | Frozen Pizza | 3.68 |
| 7218063052 | FRSC BRCK OVN ITL PEP PZ | Frozen Pizza | 3.6 |

|  |  |  |  |
| --- | --- | --- | --- |
| Least Elastic Products | | | |
| UPC | Product Name | Category | Elasticity |
| 1111009507 | PL TWIST PRETZELS | Bag Snacks | 0.029 |
| 1111009497 | PL PRETZEL STICKS | Bag Snacks | 0.61 |
| 1111085319 | PL HONEY NUT TOASTD OATS | Cold cereal | 0.99 |
| 7027316404 | SHURGD PRETZEL STICKS | Bag Snacks | 1.26 |
| 7027316204 | SHURGD MINI PRETZELS | Bag Snacks | 1.28 |

1. As the retailer, which products would you lower the price to maximize (a) product sales and (b) unit sales, and why?
   1. The strategy will be the same for sales and unit sales since sales is simply a function of units\*price. The most elastic products are ones that react the strongest to price changes. With this in mind, we would simply lower the price for the most elastic products since that change will generate more units sold than the same price change for inelastic products. Lowering prices for inelastic products will mean that the business is unnecessarily losing money as the demand might be the same or only slightly higher regardless of price.
   2. Recommendations:
      1. Rank of best to worst products to offer promotions on (based on elasticity):
         1. Meat pizza products,
         2. Cheese pizza products
         3. Cereal producs
         4. Non-pretzel bag snacks
         5. Pretzel bag snacks
2. Assumptions Evaluation for all models:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Assumption | Test | Sales | Unit Sales | HHS |
| Multicollinearity | Vif Test | Satisfied | Satisfied | Satisfied |
| Homoscedasticity | Resid vs Fitted Plot | Satisfied | Violated | Violated |
| Normality | QQ plot and KS test | Violated | Violated | Violated |
| Linearity | Resid vs Fitted Plot | Violated | Violated | Violated |

Assumptions were violated as hypothesized since I build lmer instead of glmer models. This is most likely since we have count data for two dependent variables and highly skewed sales data for the third variable.