



# Statistical Methods for Business Analytics

## Project 2

### Analysis of Kroger Shopping Behavior

Josh Allen  
Sierra Funte  
Kevin Gardner  
Digvijay Sawant

## **Executive Summary**

This report examined sales trends for households and demographics over approximately two years. Specifically, it analyzed coupon campaigns and departmental sales changes. After thorough investigation, we have determined that the best course of action includes: targeting coupon campaigns towards the specific demographics most likely to redeem them and re-engaging customers in the grocery and drug departments.

Among households that spend less over time, engagement decreases by the largest amount in the grocery and drug departments. Over the 102-week period, we estimate these customers are spending about \$8 less per week in the grocery department and \$3 less in the drug department. The households that were spending more over time showed an increase spend in the same two departments of grocery and drugs.

In order to recover value from such customers, we recommend that efforts be made to attract customers to these departments. This could be accomplished through coupon campaigns targeted at budget shoppers or other groups that are likely to respond. The groups that are most likely to respond to coupons are customers between the ages of 45-54 and with an average salary of \$50-74K. The coupons should specifically be used for the products most likely to generate increased trips to the store. These products should be selected due to the level of popularity or growth among the groups most likely to utilize coupons in purchasing. Therefore, the recommended products to use in a targeted coupon campaign would include: beef, milk, cheese and soft drinks.

## Changes in Household Spending by Department

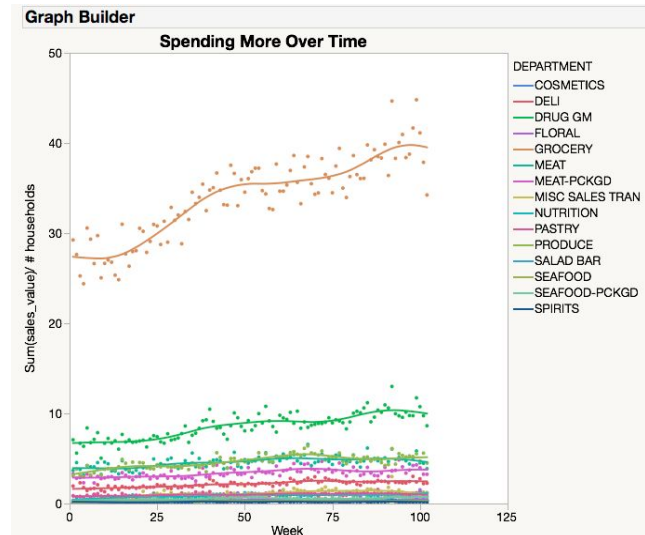
One of our primary questions concerns which categories households are becoming more or less engaged with over time. To answer this question we created two subsets of the journey transactions data and compared shopping behavior across two different sets of households: customers whose sales increased and customers whose sales decreased over the course of the two years.

After breaking these groups down, a significant trend in department sales emerged. For the customers whose sales increased over the two years, the increase in sales was mostly in the grocery and drug departments. Whereas for households whose sales decreased we observe the opposite trend. It appears that the grocery and drug departments take into account the majority of the increase or decrease in customer sales. Examining sub commodities within the grocery and drug departments would provide additional insight into factors driving these trends.

The following two charts compare weekly sales by department for the households whose overall sales decreased or increased over time. The regression estimates discussed in this section are presented in appendix A.



Among those that spent less over time, the average household reduced weekly spending in the grocery department from about \$36 in the first week to under \$28 the last week (down \$8) and reduced weekly spending in the drug department from about \$10 in the first week to about \$7 in the last week (down \$3).

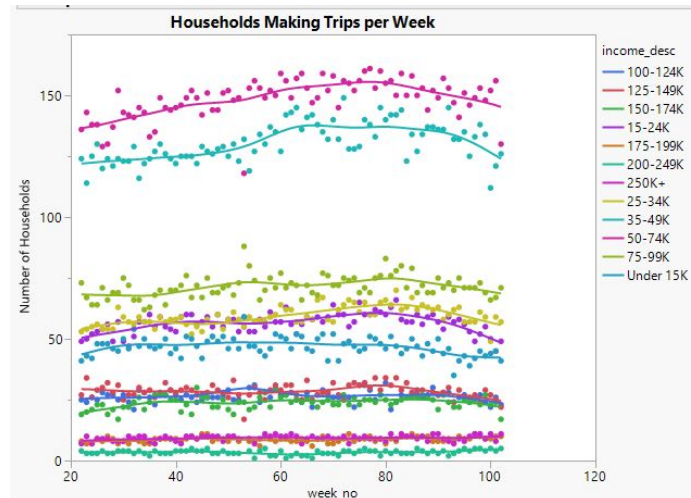


Among those that spent more over time, the average household increased spending in grocery from about \$27 in the first week to nearly \$41 in the last week (up \$14). Households increased spending in drugs from about \$6 in the first week to almost \$11 in the last week (up \$5).

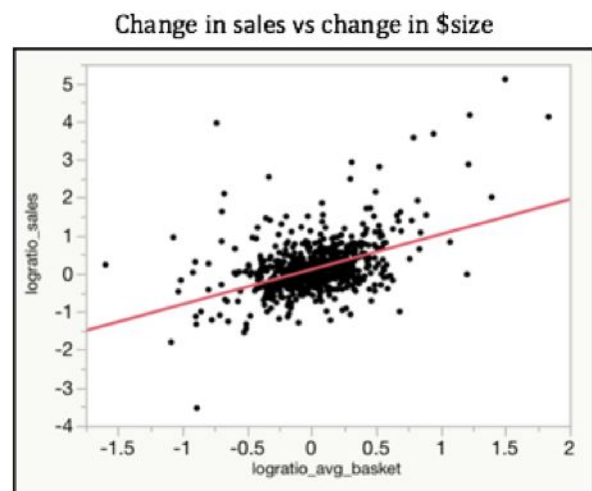
To better understand the groups most likely to increase or decrease spending over time, we decided to look at the demographics data. After careful analysis, the group which shows the most change in terms of sales is are the 11 households in the 175-199k income category.



While most other groups remained constant, the drop in sales for this income group was rather pronounced. Weekly sales per household in the group fell by around 33 percent. While this group is relatively small and the decrease is driven mostly by 1 or 2 households, there were also substantial decreases in other members of this income group. To understand this issue, we looked at the number of households making trips to Kroger on a weekly basis by income level. The results showed that the number of households shopping in any given week by income level remains constant. This is positive, as it means that those customers aren't leaving Kroger, simply that they are spending less at Kroger over time.



If the customers are spending less over time at Kroger, the question is raised as to whether this is due to fewer trips or if basket size is the primary driver of declining sales. To understand this question, log-log charts were developed to examine the percentage change in dollar sales based on the percentage change in the number of trips or in the basket size. The results showed that both basket size and number of trips have an impact on dollar sales, however there is a much stronger association with number of trips and dollar sales. Based on the generated model, there would be a predicted 0.976 percent increase in sales for every 1 percent increase in number of trips.



Response logratio_sales				
Whole Model				
Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	0.0295355	0.011537	2.56	0.0106*
logratio_trips	0.9766342	0.019114	51.09	<.0001*

Response logratio_sales				
Whole Model				
Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	0.113239	0.021042	5.38	<.0001*
logratio_basket	0.920108	0.065356	14.08	<.0001*

Following this analysis, it became clear that in order to make up for the declining sales in grocery and drug departments, it is critical to increase the number of trips made to the store. The most clear method to increase number of trips is to utilize coupon sales campaigns to drive customer engagement.

### Coupons and Campaigns

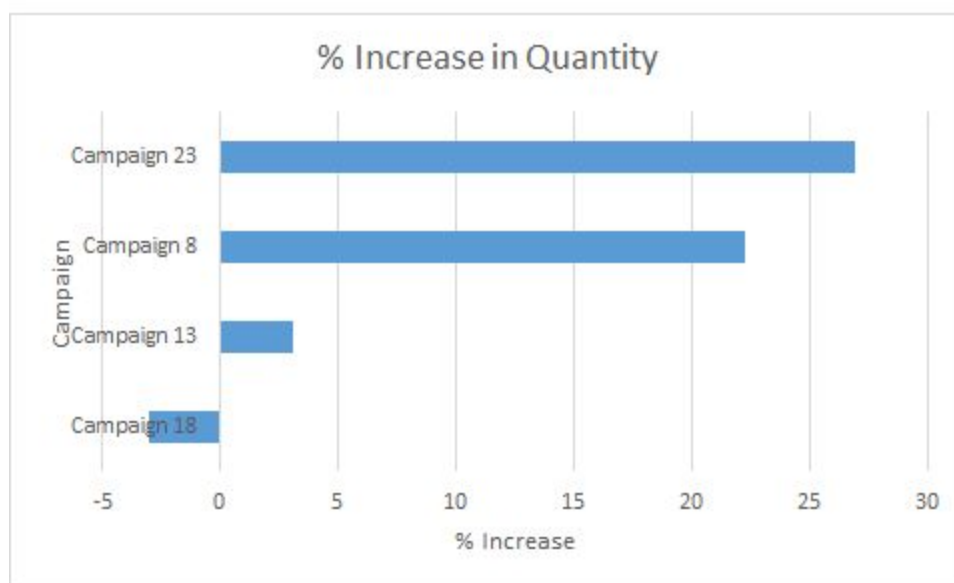
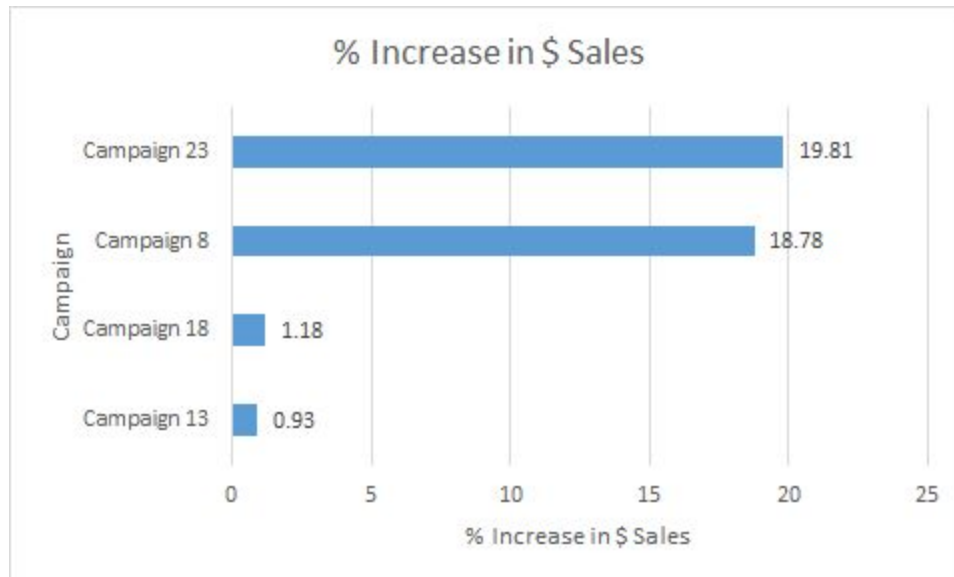
To better understand methods of customer engagement, it was necessary to research the effectiveness of coupon campaigns. In order to do so, a method must be developed to gauge success and compare campaigns. The most easily accessible methodology would be to compare coupon usage, sales of the products featured in the campaigns and total dollar sales. A major problem arises when attempting to use total dollar sales as a metric when comparing campaigns. Many campaigns overlap, such that any trends could possibly be the result of multiple campaigns and making it difficult to ascertain the impact of any singular sale. Therefore, it was decided to only use the sales of products featured in an individual campaign and the coupon usage for an individual campaign when examining campaign success.

The first step to obtain the necessary data in this analysis was to identify the number of households targeted in a campaign and the number of distinct households which used coupons (see SQL Query in Appendix B- Q1). This gives the coupon usage metric. To obtain the total product sales measure, JMP was utilized to join tables containing transaction\_data and

journey\_coupon\_redempt. This gave the campaigns, the products featured in the campaigns and the total sales for those products on a weekly basis. The first 22 weeks in the data set were excluded, as they did not contain the full number of households and thus could skew the data. Then campaign weeks were identified and the mean of product sales during those weeks were compared with control weeks to create the percent change in product sales compared to the norm.

Once the data was collected, it was possible to more closely examine the impact of the types of campaigns. There are three major coupon campaign types: Type A, Type B and Type C. Type A and Type B were chosen for further examination due to their widespread distribution and high coupon usage. Within the Type A subset, three significant campaigns were recognized: Campaign 8, Campaign 13 and Campaign 18. In each of these campaigns, over 1000 coupons were distributed to households. These campaigns were especially prominent when examining the rates of coupon usage. While most campaigns were receiving limited consumer attention, these campaigns were utilized by 34, 58 and 57 percent of targeted customer respectively. During Campaign 8, there was an increase in dollar sales of 18.8 percent. However, customer usage of coupons is not necessarily indicative of campaign success. In Campaign 18, despite the large percentage of customer usage, there was only a small increase in dollar sales and the quantity of products sold actually decreased over the course of the campaign. Campaign 13 faced similar challenges in increasing sales or quantity during the course of the campaign.

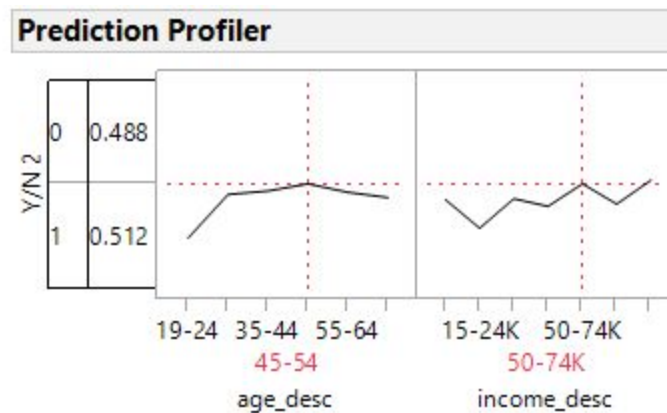




The fact that Campaign 8 had the lowest coupon usage rates and the highest increase in sales of the Type A campaigns we examined raises an interesting question. Is a campaign more successful due to the number of households included or the segments which are reached over the course of the campaign? Type B campaigns were useful aids in resolving this question. These campaigns did not permeate the households in the same way as the Type A campaigns.

In the most successful Type B campaign, Campaign 23, only 183 households were targeted. However, throughout the campaign there was a distinct trend in increasing dollar sales and quantity sold. In fact, Campaign 23 had the largest growth in sales of any coupon campaign at a nearly 20% increase. The success of this campaign demonstrates that the number of households targeted is not necessarily the best indicator for increasing sales. It is a better method to target coupons towards customer segments most likely to utilize them then to send out a massive campaign.

To understand the customer segments which most frequently participated in these campaigns, it was important to examine the odds of a certain demographic redeeming a campaign. The most prominent demographic groups were the 45-54 age and the 50-74k income brackets. These segments were highly predictive of coupon usage, with the two groups together carrying a 51 percent chance of utilizing a coupon during these campaigns. As Kroger moves forward, it is important to allow these groups and their buying trends to influence our coupon campaigns. (Refer Appendix B Table. 1).

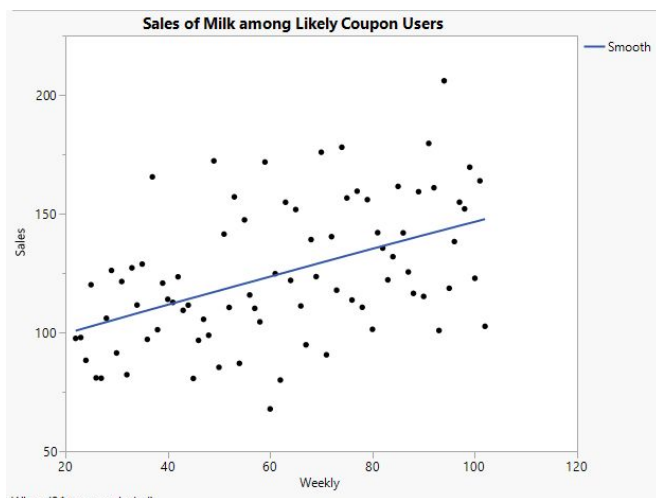


The best way to understand these demographics is to look at their sales trends. Among the segments identified as most likely to use coupons, the products with the highest weekly sales include beef, soft drinks, fluid milk products and cheese. Of these the largest

growth in dollar sales is seen in fluid milk products over the course of the two years examined.

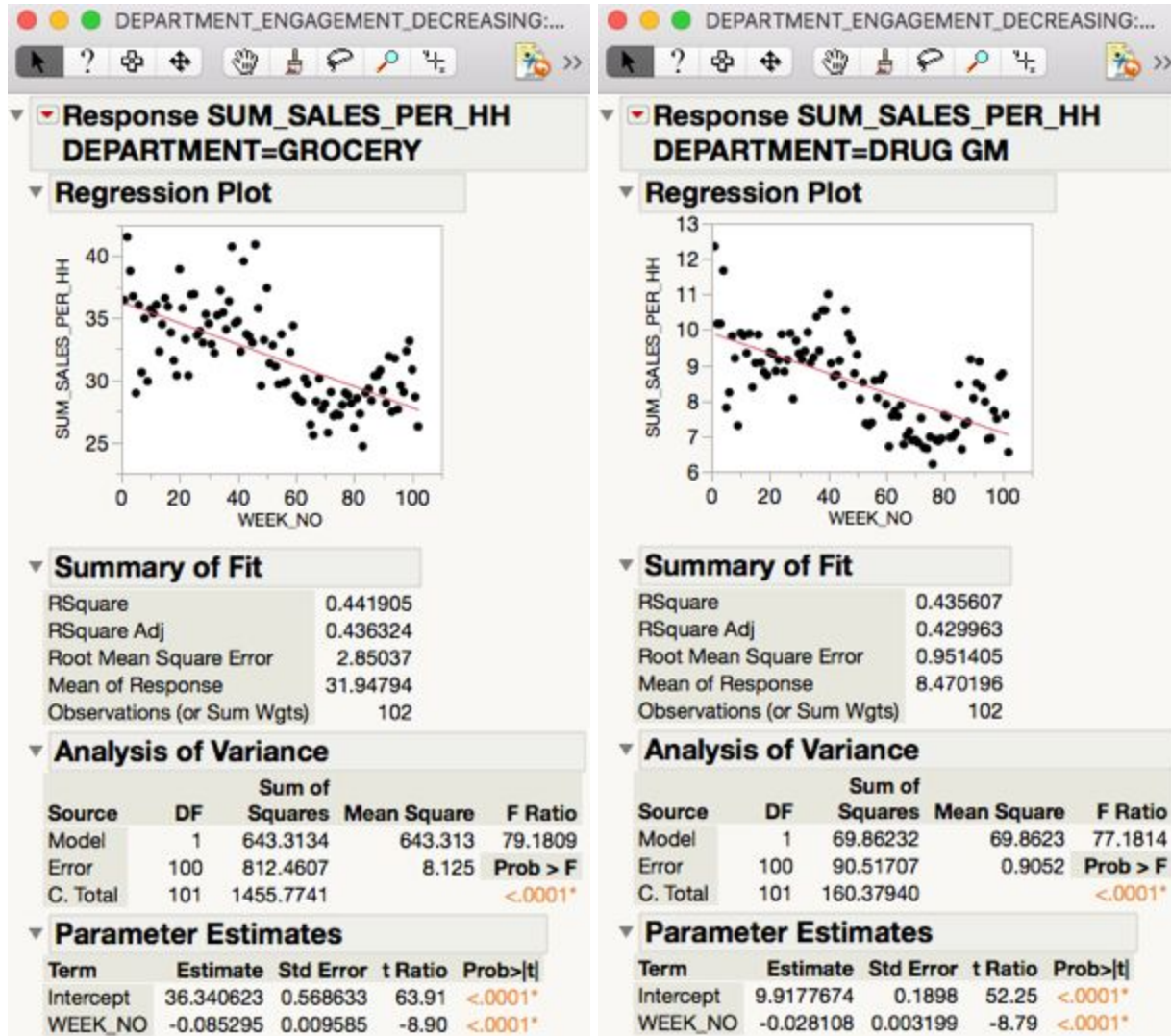
The predicted sales increased by nearly 50 percent from week 22 to week 102. This would indicate that milk should be prevalently featured in any coupon campaigns undertaken. Since it is already a product with large growth potential, it could drive coupon-oriented budget shoppers into the store. This will likely increase the frequency of trips, as customers will make special trips to the store if they have a coupon for a product which they were likely to buy anyways.

Considering the evidence presented earlier which suggests that number of trips to stores are more predictive of higher sales, it seems to be a wise decision to hold frequent coupon campaigns targeted at these demographics through featuring their most used and growing products.

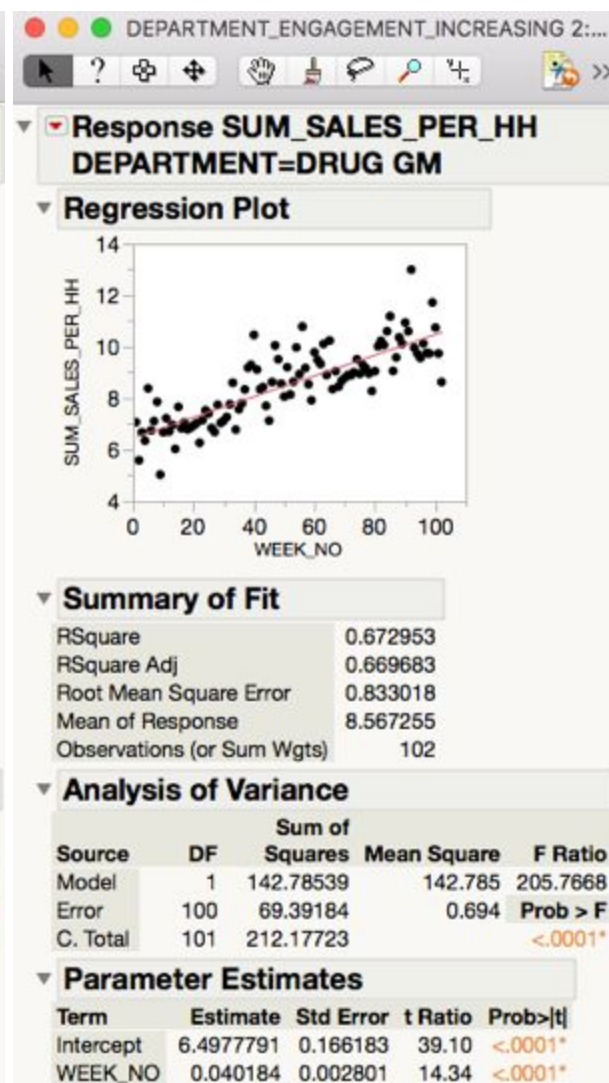
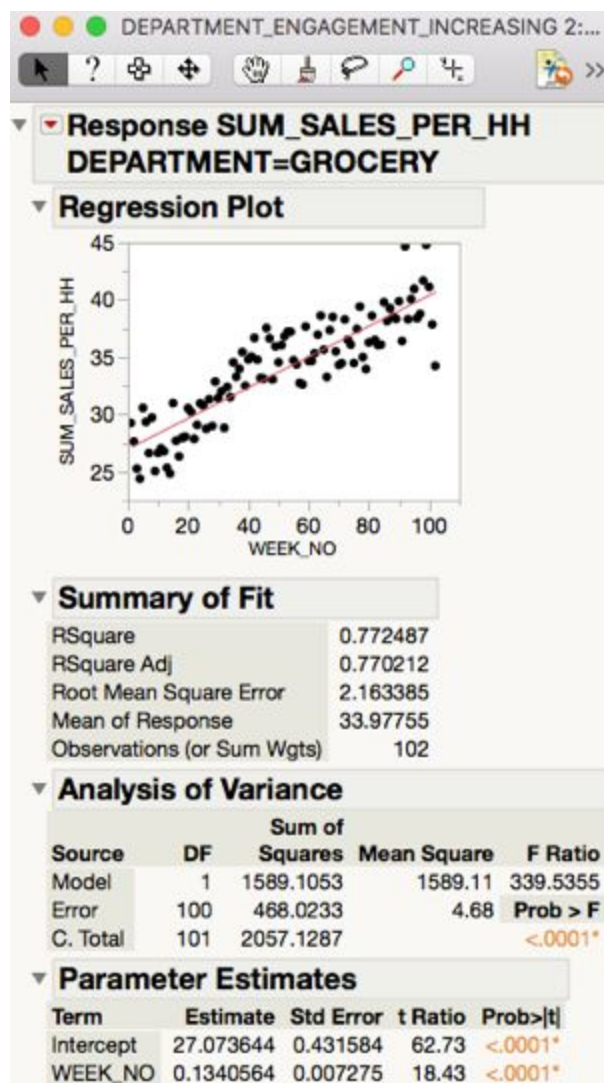


## Appendix A

Below are selected regression outputs for each of the two data subsets of our data.



From the decreasing engagement subset, the intercept and week are significant for Grocery and Drug GM. Interpreting the coefficients, we have that expected sum sales per household in grocery is about \$36.34 minus \$0.09 for every week. Expected sum sales per household in Drug GM is about \$9.92 minus \$0.03 each week.



For the increasing engagement subset the intercept and week are significant for Grocery and Drug GM. Interpreting the coefficients, the expected sum sales per household in grocery is about \$27.07 plus \$0.14 for every week. Expected sum sales per household in Drug GM is about \$6.50 minus \$0.04 for every week.

## Appendix B

### Q1.

```
select distinct campaign, count(distinct household_key) from  
journey_campaign_table group by campaign;
```

```
select distinct campaign, count(distinct household_key) from  
journey_coupon_redempt group by campaign;
```

**Table.1**  
**Lack Of Fit**

Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	30	21.72490	43.44979
Saturated	41	479.70720	<b>Prob&gt;ChiSq</b>
Fitted	11	501.43210	0.0535

### Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-0.5931518	0.1038753	32.61	<.0001*
age_desc[19-24]	-0.8217711	0.3236253	6.45	0.0111*
age_desc[25-34]	0.11439784	0.1774539	0.42	0.5191
age_desc[35-44]	0.18012654	0.1576162	1.31	0.2531
<b>age_desc[45-54]</b>	<b>0.31800896</b>	<b>0.1417731</b>	<b>5.03</b>	<b>0.0249*</b>
age_desc[55-64]	0.15518606	0.2401741	0.42	0.5182
income_desc[Under 15K]	0.01808744	0.247217	0.01	0.9417
income_desc[15-24K]	-0.5810561	0.2479644	5.49	0.0191*
income_desc[25-34K]	0.0342666	0.223606	0.02	0.8782
income_desc[35-49K]	-0.1159187	0.1620919	0.51	0.4745
<b>income_desc[50-74K]</b>	<b>0.32176148</b>	<b>0.1506297</b>	<b>4.56</b>	<b>0.0327*</b>
income_desc[75-99K]	-0.069717	0.204212	0.12	0.7328

For log odds of 1/0