

# Measuring Product-level Promotional Effectiveness using Multiple Linear Regression

Aaron Clark, Cory Weinmann, Donald Kirk

## ABSTRACT

Advertising weekly bargains in paper circulars, temporarily dropping regular prices, spooning out rewards club coupons, slashing prices in half during “super-sale” events, and marketing coupons to customer with unique purchasing habits are some of the many promotion strategies typically offered at your local retailer to lower the price of your favorite good. However, all promotions are not created equal from product-to-product. That is, customers that buy certain products respond differently to varying types of promotions at different times for different reasons, affecting the prediction of sales quantity and ultimately, the best type of promotion to run. To determine the most effective type of promotion for any given product, we built a multiple linear regression model which incorporated inputs such as promotional depth of discount, a store’s market location, seasonality of demand, advertisement effort, rewards member participation, the number of days since the last promotion, and how deep the previous discount was. Furthermore, we optimized the predictive results from the model to rank promotion types from the perspective of predicted sales quantity. This paper walks through the development and validation of a multiple linear regression model predicting quantity sold for an exemplar product using SAS® Enterprise Guide 6.1. A background in multiple linear regression is assumed.

## KEYWORDS

Retail, Big data, Multiple Linear Regression, PROC HPREG, Promotional Effectiveness

## INTRODUCTION

For any retailer, success is achieved by meeting the needs of their customers and devoting their pricing strategy to competing with other top retailers. Therefore, implementing the correct type of promotion at the correct time and marketed to the correct customer is imperative to increasing traffic, and thus storewide revenue. In addition to promotion type, the depth of a promotion also positively or negatively influences profit margin and thus needs to be evaluated and justified when considering a given promotion.

In order to investigate and rank the effectiveness of promotion types, we used SAS® Enterprise Guide 6.1 to build a multiple linear regression model for prediction of sales quantity for a specific product over the duration of the sale. The model took into account these other pertinent inputs:

- Geography: 6 different markets were tracked composing a varying number of stores in each market
- Seasonality of demand: A high level metric which looked at how well the product category has been performing, for all retailers within the footprint of this retailer
- The promotion depth, or price point’s distance from regular (everyday) pricing:  $(\text{Promo price} / \text{regular price})$
- Advertising efforts: was the product featured in circulars? Was it featured on the front, center, or back pages? Was the promotion unadvertised?
- Overall product category performance compares sales from this year to last year
- Rewards member participation: the percentage of rewards members who purchased the product out of all customers who made the purchase
- Days since the last promotion
- Depth of the previous promotion’s discount

The goal was to first predict, with accuracy, the amount of product sold within these above distinctions. Product X, who’s identify isn’t important for the scope of this paper, was the perfect exemplar to help us do so. Product X is promoted often, has high buy frequency and significant household penetration as a household staple. We trained the model using a subset of the data and tested how much more (or less) accurate the predictions were compared to the inventory-planning quantity forecast using the remaining data. This forecast, which we will now refer to as “The forecast” is sufficiently accurate as it helps the retailer’s staff prepare their stores’ inventory for the scheduled promotion. After the model development and validation phases, we created a table to rank which of the attempted promotion types were most effective at increasing the predicted sales quantity on average.

## METHODOLOGY

Product X has been a top seller among retailers' aisles for years. This retailer's stores have sold the majority of their product X at promotional (promo) pricing. For this retailer, 8.2% of customers have purchased this product X in the last year with a buying frequency of 1.8 on average. A rich sales history such as this yields a delightful opportunity for evaluating the effectiveness of the different promotions attempted over the recent years and suggest the best combination of pricing and advertising events to boost quantity sold.

The SAS®-created tool was designed to only use available data two years and six months prior to August 2014. A store's data was eligible for analysis if the store was open the entire duration of the promotion and didn't experience a significant event (i.e. remodel) during that time period, as those events would negatively influence sales quantities. Each observation represents conditions at the market level for a given promotion block. Stores were removed prior to the analysis because of a difference in marketing strategy that would cloud the results if included with the rest of the stores.

After incorporating all eligibility requirements, 87% of their stores were included in the analysis. Further geographic segmentation of the stores were needed to account for variability associated with store locations. For example, location can account for cultural differences in purchasing habits or the presence of a competitor with a large market share in one location but not in another. The breakdown of number of stores with eligible data for analysis from each region is displayed in table 1 below. We labeled each region with a 1 through 6 in order to maintain anonymity of the retailer's identity and regional pricing organization. From the table, regions 2 through 5 are fully intact, while region 1 lost only a handful of stores and region six lost a significant proportion of its stores due to not meeting the eligibility requirements. Due to inconsistencies between regions, a multiple linear regression model was built for *each* region to more accurately predict sales quantity within said region.

Region	Percent of Stores Included
1	72.5%
2	100%
3	100%
4	100%
5	100%
6	21.7%

**Table 1: Percent of stores used in analysis, by region**

Below, table 2 includes the response and explanatory variables we suspected would impact quantity sold and some necessary descriptions for each:

Variable Name	Label	Range of values		Description
		Min	Max	
RgBk_act_qt	Quantity Sold	See table 3 for region level statistics		The response variable. The actual quantity sold in each region for a given promotional block.
wk_pct_rwd	Rewards Member Participation	See table 3 for region level statistics		For those who purchased the product, we measured the proportion of customers who were rewards members out of all customers who made the purchase during promotion
wk_tyly_index	Category Performance Index	65.7%	190.8%	The product category's total sales performance during the week of the sale as compared to previous year's category performance during the same week. This accounted for increased sales that are due to more customers shopping stores for product X's product category.
wk_seas_Index	Seasonality Index	50.3%	180.3%	A weekly index which measures how sales of all of product X's product category has fluctuated over the last 2.5 years
BkDays	Block Days	1	27	Number of days the promotion block lasted
Variable Name	Label	Range of Values		Description
		Min	Max	
Bk_Depth	Depth	67%	97%	Promotion depth: promo price / regular baseline price

daysSinceLastPromo	Number of Days Since Last Promo	7	126	Days since last promotion block ended
prev_depth	Previous Depth	67%	97%	Prior Promotion Depth: promo price / regular baseline price
UAD	Unadvertised	0	1	A passive advertising effort means that the product was displayed on sale on the shelf edge only
frontpage	Front Page	0	1	Product was advertised on front page of weekly circular during the promo block
backpage	Back Page	0	1	Product was advertised on back page of weekly circular during the promo block
centerpage	Center Page	0	1	Product was advertised in any of the center pages of weekly circular during the promo block
circular	Circular	0	1	Product was advertised in the circular
insert	Insert	0	1	Product was advertised in a special insert which wraps around the circular
LongTermPromo	Long Term Promotion	0	1	Product was on a long term sale that lasts a minimum of 4 weeks and was signified with a different style tag on the shelf edge

**Table 2: Response and Explanatory variable possible values and descriptions**

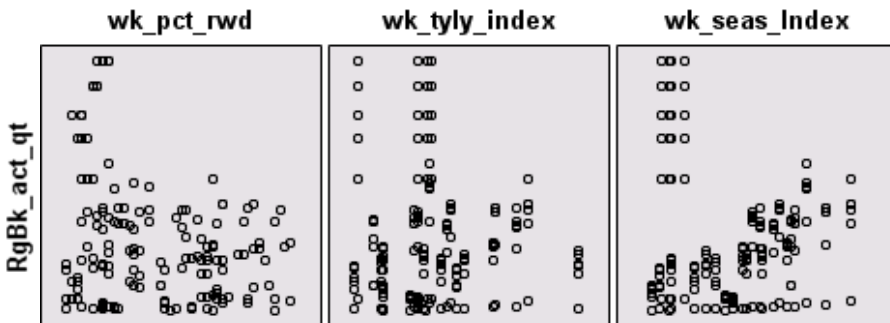
As some of the values vary by region, we may want to explore those possible values at the region-level. The variables in Table 3 include the min and max values recorded for the past 2.5 years of promo history for product X.

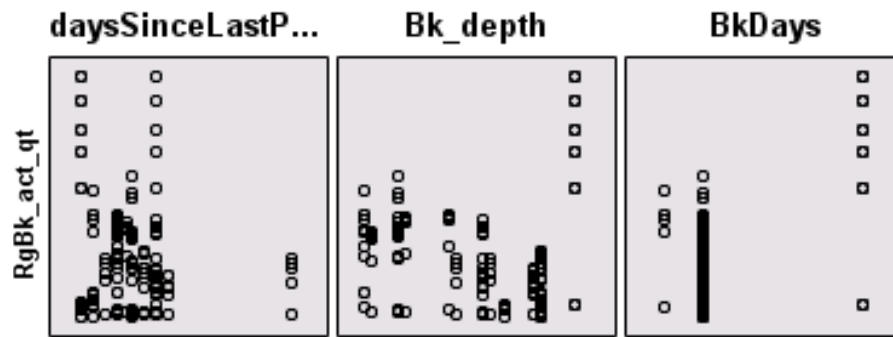
Variable	Statistic	Region					
		1	2	3	4	5	6
RgBk_act_qt	min	5,136	6,906	5,224	129	8,295	565
	max	83,100	122,150	140,925	156,525	107,720	16,138
wk_pct_rwd	min	9	8	11	13	8	8
	max	39	38	45	51	37	36

**Table 3. Regional minimum and maximum statistics for variables-of-interest**

In figure 1 (below) we used PROC SGSCATTER with a MATRIX statement to better understand the relationship between the quantitative explanatory variables and the response, quantity sold. Ideally, we'd like to observe a nice linear relationship that when the x-axis variable increases, then the quantity sold (RgBk\_act\_qt) increases or decreases at a constant rate. We see that a positive association exists between the seasonality index and quantity sold with one exception: at the very low values there seems to be a very high quantity sold in all six regions which doesn't follow the upward linear trend. In other words, some other source of variation allowed those promo blocks to break the mold and yield very high quantity sold. Looking at the scatter plot of days of promo (bk\_days) vs quantity sold, we can guess which promotion this was: the 27 day-long promotion. Most of the past promotions for product X at this retailer are a single week while a handful were one day in length. Another positively associated pattern exists in the category performance variables (wk\_tyly\_index), but slightly less distinct than that of the seasonality index.

Promotion block depth seems slightly negatively associated, meaning as the depth gets closer to 1 (indicating less savings at the register) then quantity sold during the given promotion decreases. Note that the other (daysSinceLastPromo and wk\_pct\_rwd) explanatory variables have a questionable relationship with quantity sold. We decided to include them in the model building process regardless.

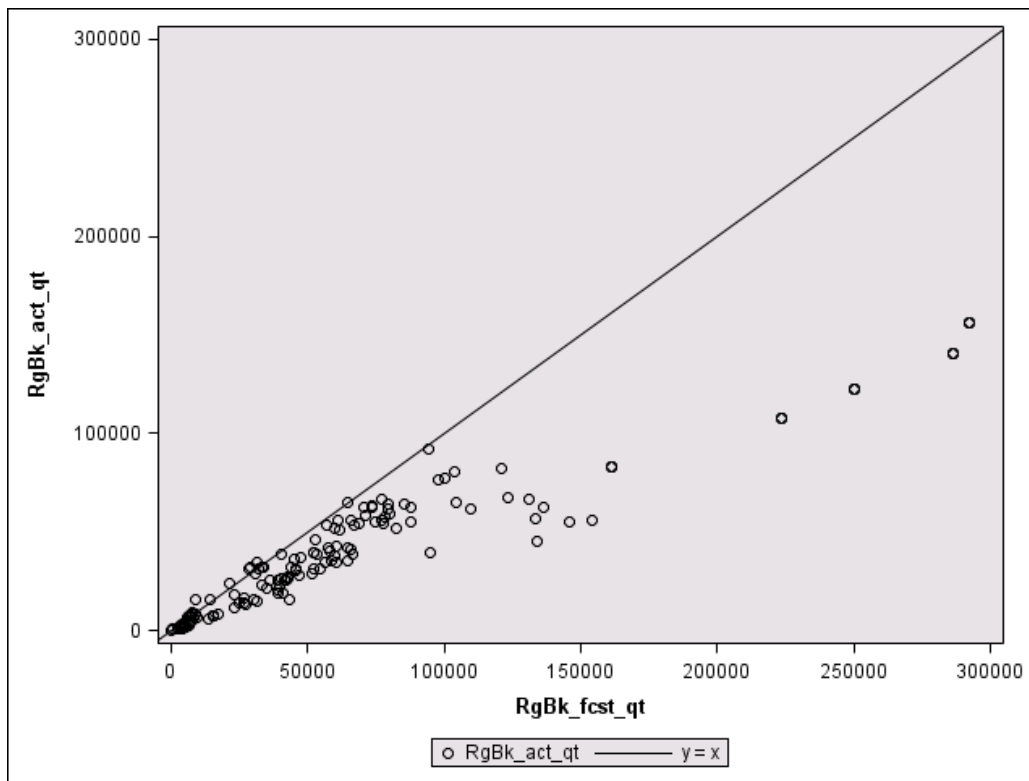




**Figure 1. Scatter plots of quantity sold with each quantitative explanatory variable**

As stated in the introduction, we used SAS® Enterprise Guide 6.1 to build six multiple linear regression models to predict region-based quantity sold for product X. Specifically, we used PROC HPREG with backward selection and a significance level of .10 for a covariate to stay in the model. To validate the model, we randomly sampled 19 (73%) of 26 promotion blocks recorded over the past 2.5 years to build the model. We did this 500 times using PROC SURVEYSELECT to create a distribution of statistics measuring the accuracy of the model's predictions in the test set (the remaining 27%) by comparing if we were closer to the actual quantity sold than the planning forecast. As figure 2 (below) shows very well, the planning forecast was generally accurate, with a tendency to overestimate inventory needed to supply stores during the sale. This makes sense, since product X is an item that does not perish, and buying more than needed only ensures the stores have ample supply for the promotion period, and extra to sell at regular pricing. Running out of the product would much more detrimental as an out-of-stock item advertised at a bargain price risks aggravating the customer during their shopping experience. Next, the line with equation  $y=x$  was plotted in figure 2 to show what an exactly correct planning forecast would look like, even though it's not likely. Therefore any data plotted above this line means if those regions followed the planning forecast, they would have ran out of inventory before the end of the sale. Note that this only happened a handful of times and in those cases, the planning forecast was only about 10% short of meeting the demand. However, in one case, the difference was drastic when product X was discounted to 33% off the regular price and region 6's planning forecast was 80% short of predicting the need for its stores. Conversely, most of this retailer's promotions were below the  $y=x$  line, meaning their stores were aptly in-stock.

We determined that if we are more accurate than this forecast for the majority of the time, then we are accurate enough to make inferences from the model. We should also note that in the case where the predictions are not closer to the actual quantity sold than the planning forecast doesn't necessarily mean that the prediction was not accurate, but rather the planning forecast may have been extremely accurate. As evidence, notice how some planning predictions fall very closely to the plotted  $y = x$  line.



**Figure 2. Product X: Planning Forecasted Quantity for vs. Actual Quantity Sold (Line:  $y=x$ )**

The next step was building a robust model so we could field requests regarding the optimal promotion type and depth, etc., to maximize sales quantity.

## Model Building and Validation

Using the HPREG procedure code below, we produced six models using backward selection. Note the use of the BY statement in PROC HPREG means we are also required to sort the data set by region prior to executing it. The ID statement allows us to see additional columns in the output data set we called "PredLinReg" where we requested to see both the resulting predicted and residual values. We've found PROC HPREG to be superior to its REG procedure predecessor for a number of reasons. First, there is no longer a need to manually create dummy variables for categorical covariates. Secondly, using "|" between variables allows for easy checking of all possible interaction variables' with the inclusion of "@ x" where x can be any integer greater than 0 and less than the number of variables on the right hand side of the "=" of the MODEL statement. Third, the HPREG procedure has a SELECTION statement where we've noticed more versatile functionality as opposed to a selection option on the MODEL statement in PROC REG. Regardless of which PROC you choose, here are many selection methods to choose from such as forward, backward, stepwise, forwardswap, lar, and lasso. Here we used backward, which by default selected covariates to stay in the model based off their significance level which must be less than .10.

```
proc sort data=regData;
  by region;
run;
proc hpreg data=regData;
  by region;
  id wk_pct_rwd wk_tyly_index wk_seas_index BkDays Bk_depth RgBk_act_qt
    RgBk_fcst_qt frontpage backpage centerpage circular LongTermPromo uad ;
  model RgBk_act_qt = wk_pct_rwd | wk_tyly_index | wk_seas_Index | BkDays |
    Bk_depth | frontpage | backpage | centerpage | circular | insert | uad |
    daysSinceLastPromo | prev_depth | LongTermPromo @1;
  selection method=backward;
  output out=WORK.PredLinReg predicted=pred_qt residual=resid_qt;
  TITLE "PROC HPREG: MLR Results using Backwards Selection";
run;
```

The results from PROC HPREG outputted the following models for each region:

Region	R <sup>2</sup>	Model
1	.972	$\widehat{Qty} = -57,25.8007 + 253.272(wk\_seas\_index) + 4,329.0168(BkDays) - 43,663(bk\_depth) + 41,147(insert) + 23,107(frontpage) + 20,362(centerpage) - 95.3552(daysSinceLastPromo)$
2	.971	$\widehat{Qty} = -397.2467 + 288.7629(wk\_seas\_index) + 6,177.8342(BkDays) - 64,388(bk\_depth) + 52,503(insert) + 34,232(frontpage) + 23,821(centerpage)$
3	.979	$\widehat{Qty} = -15,724 + 327.0229(wk\_seas\_index) + 7,255.281(BkDays) - 59,842(bk\_depth) + 57,736(insert) + 40,252(frontpage) + 33,455(centerpage) - 103.42(daysSinceLastPromo)$
4	.986	$\widehat{Qty} = -9,705.349 + 268.5463(wk\_seas\_index) + 7,734.4702(BkDays) - 61,369(bk\_depth) + 57,474(insert) + 35,237(frontpage) + 27,266(centerpage)$
5	.955	$\widehat{Qty} = 17,529 + 357.6028(wk\_seas\_index) + 5,754.2445(BkDays) - 87,838(bk\_depth) + 55,031(insert) + 37,887(frontpage) + 27,100(centerpage) - 143.5529(daysSinceLastPromo)$
6	.801	$\widehat{Qty} = 19,452 + 172.936(wk\_pct\_rwd) + 1,789.5162(BkDays) - 25,310(bk\_depth) + 6,787.1402(insert) - 25,666(LongTermPromo) - 12,503(prev\_depth)$

**Table 4. R-squared and multiple linear regression models from PROC HPREG using Backward Selection**

Overall, the amount of variability observed in quantity sold accounted for by the explanatory variables (R<sup>2</sup>) was rather high at greater than 95% for all regions with the exception of region 6 where we only had 21.7% of the stores included in the analysis. Notably, the covariates kept in the model for regions 1 through 5 were relatively the same with the exception of “daysSinceLastPromo” disappearing from the models for regions 2 and 4. Otherwise, 1 through 5 always included the product category seasonality index, the days and depth of the promo block, and three advertising efforts: a feature in the insert, front page, or center page of the circular. To interpret the coefficients of these covariates, we addressed those of model 1 only: With all covariates remaining constant, for every one unit increase in...

- the seasonality index leads to a 253.272 unit increase in predicted quantity sold, on average. This makes sense as the seasons change, demand for product X’s product category also changes. For example, during the colder months product X is typically used more often in the average household.
- the duration of the sale (in days) leads to a 4,329.0368 unit increase in predicted quantity sold, on average. This makes sense too, as the more opportunity you allow the customer to benefit from a promo, the more they will!
- the depth of promotion (as a percent of regular price) leads to a 43,663 unit decrease in predicted quantity sold, on average. Basically, there are customers who wait for the really deep sales to fill up their shelves and likely retain knowledge of what is a “good” price looks like for product X and don’t buy it until they see it. Also, there are customers who may not have knowledge of product X, but see the tag on the shelf and take advantage of deal by grabbing more than they normally would.

Next, the significant qualitative covariates regarding advertising effort: if all other covariates remain constant, when the product X was featured in the ...

- wrap-around insert of the circular, you can expect a 41,147 unit increase in quantity sold, on average.
- front page of the circular, you can expect a 23,107 unit increase in quantity sold, on average.
- center page of the circular, you can expect a 20,362 unit increase in quantity sold, on average.

It’s interesting to see that the wrap around insert has more visibility to the customer than the front or center pages of the circular. This may be due to the fact that the circular doesn’t have an insert every week, so the products featured there must be viewed as a “special” sale of some kind. There was not much difference in expected quantity sold from being featured on the front page vs. the center pages as both are quite visible to the customer, with the front page slightly more so than the center.

Lastly, for every day since the last promotion, we could expect a decrease in quantity sold of 95.3552 on average assuming the other covariates remain constant. This result doesn’t make as much sense as the others. Common sense would tell you that customers would buy more product X if a significant time period has passed since the last promotion, because those high-volume basket-loaders are just waiting to restock their shelves. Note that this variable was marginally significant in the models with a p-value near .10, hence why it didn’t remain in all the regions’ models. However, if it helped us predict quantity sold even slightly more accurately, we considered it useful.

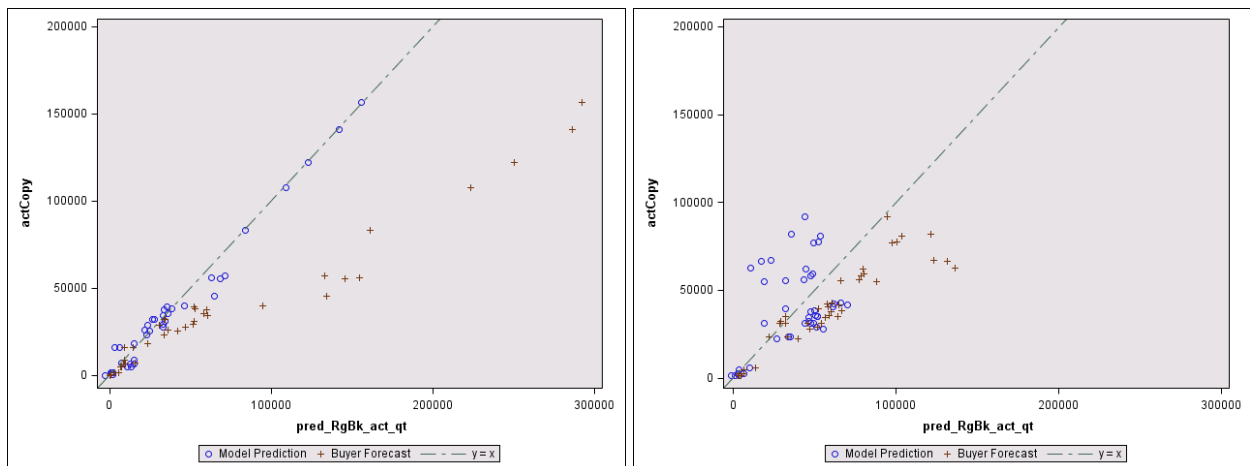
Region 6 had mostly the same, but somewhat different, influential covariates stay in the model: percent reward member participation, depth of the previous promo block, and the long-term promotion flag were introduced as significant predictors of quantity sold. Note that 80% of the variation in quantity sold was explained by these covariates, so we would expect region 6's model to predict slightly less accurate than 1 through 5. Interpreting these new additions within region 6:

- For every 1% increase in reward member participation during the promo block, then we can expect a 172.936 unit increase in quantity sold on average when all other covariates remain constant.
- With all other covariates remaining constant, we could expect an average increase of 25,666 units in quantity sold if the sale was a long term promotion.
- The interpretation for the previous promo block's depth is the same as the interpretation for the current promo block depth.

As mentioned in the methodology section, we performed a validation on the model's accuracy by testing each model built on 73% of the promotional data and tested on the remaining 27%. Using the using the PROC SURVEYSELECT code below, we were able to repeat this process 500 times, randomly selecting groupings for the training and test set every time.

```
PROC SURVEYSELECT data=unique_prBlks
  method=srs n=19 reps=500 outall out=random_PrBlks;
RUN;
```

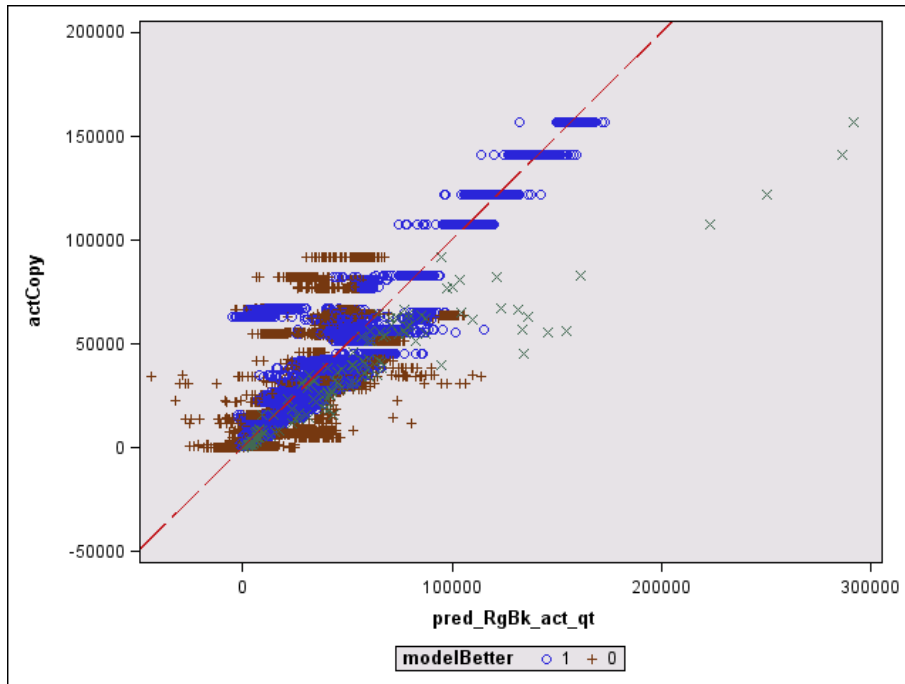
Walking through the above code, the input data set was called "unique\_prBlks" which contains a single column that uniquely identifies each promotion block. Being as there were 26 of them, we used a simple random sampling or "SRS" method to select 19 of the promotions and do it 500 times. Note that rep=500 creates a column called "Replicate" in the output data set with values 1 through 500 to keep track of this indicator. The "outall" option outputs both the 19 selected and 7 non-selected promotions within each replicate and utilizes a column called "Selected" to keep track of that indicator as well. Next, we used the SQL procedure to join the "Replicate" and "Selected" statuses in the output data set, called "random\_PrBlks" to the original "regData" data set from the PROC HPREG code seen previously. Next, we made a copy of the response, quantity sold, before setting it to missing when selected = 0. That way, when we run the PROC HPREG from above, we get a prediction for all observations, selected or not, while building the models off observations where the response variable existed. Note that the HPREG procedure's code must also be executed by replicate using the BY statement. From there we calculated some summary statistics and graphics to measure the model's accuracy using the observations where selected = 0 over the 500 replications.



**Figure 3. Comparing model predictions to planning's forecast among replicates (left: rep = 1; right: rep=4)**

As made visible by figure 3, there existed replicates where the collection of promotions used to build the model were extremely accurate at predicting quantity sold for the test set observations, as the predictions are very near the  $y = x$  line fitted in the plot area. However, in other replicates the model wasn't as accurate, as seen in the right-hand plot of figure 3. That's why it was much easier to simply calculate how often we are better than the planning-forecast. We should keep in mind that when the forecast was already extremely close to the actual sales quantity, that does not necessarily make the prediction "inaccurate".

Figure 4 below contains all of the 500 repetitions from the PROC HPREG output data set. The green “x’s” represent the planning forecast for each region while the red “+’s” and blue “o’s” represent instances when the predictions were worse than the forecast and better than the forecast respectively. Because there are so many repetitions, you can almost see the distribution of predictions that are possible, given the combination of covariates that exist in the test set. In general, the predictions were very accurate regardless of which 73% of the promotion blocks were selected to build the models, as evidenced by the concentration of blue “o’s” surrounding the  $y = x$  line imposed on the scatterplot.



**Figure 4. Plot of when regional models predicted better or worse than planning forecast, replicated 500 times**

Last, table 5 displays some summary statistics, by region, on how far off or close the model’s predictions were compared to the planning forecast. Negative values indicate the quantity the model was closer to the actual quantity sold compared to the planning forecast. For example, the model’s predictions for Region 3 were 29,896 units closer to the actual quantity sold than the planning forecast, on average. Conversely, a positive values indicate higher planning accuracy and zero means the prediction and planning forecast were the same distance away from the actual quantity sold. Predictions for models 2 through 5 were 6k to 10k more accurate than the planning forecast, according to the median. Model 6 was also more accurate, with a median of 275. Notice how far the means were skewed from center compared to their medians. This phenomena was due to extreme values on the bottom of the distribution. For that reason, we displayed the 1<sup>st</sup> and 99<sup>th</sup> percentile as opposed to the min and max to display a more representative account of the spread.

Region	Mean	1st Percentile	Median	99th Percentile
1	-16,066	-77,870	-6,347	21,109
2	-25,847	-127,880	-7,310	25,102
3	-29,896	-145,131	-9,664	31,872
4	-29,436	-135,288	-9,394	20,711
5	-21,820	-115,255	-7,635	43,356
6	-238	-7,985	-275	11,905

**Table 5. Summary Statistics on Difference between (predicted – act qty) and (forecast – actual qty)**

Finally, table 6 contains information on how frequently we are “beating” the planning forecast. Note that these statistics vary from run-to-run as the 500 samples taken vary, but not by more than 1% in either direction. All things considered, we are performing better than the planning forecast about 73.5% of the time for Regions 1 through 4. These proportions do not include an adjustment for when the planning forecast was extremely accurate. Because regions 5 and 6 are greater than 50% (66.59% and 53.26% respectively), we were satisfied with the model and its inferential abilities.



Region	Percent Model Better
1	72.02
2	72.58
3	74.07
4	75.31
5	66.59
6	53.26

**Table 6. Percent of 500 replications where model predictions were more accurate than the planning forecast, by region**

Having justified the inferential use of the models, we sought to learn from attempted promotions from the past.

## Results

The final component of this quantity-sold driven promotional effectiveness project was to optimize the model for each region to determine an order of best-to-worst quantity sold yields, on average. Unfortunately, we cannot control all of the covariates that were found significant in the model, such as the seasonality of demand in regions 1 through 5, or reward member participation in region 6. As a result, constant values were set for these inputs as we compare combinations of the remaining covariates.

Table 7 displays all 21 possible combinations of significant covariates that were found in the data and fed them through the model from region 1. The first seven columns represent inputs to the regression equation. Since we were comparing quantity sold for promotions of different duration (in days), it makes more sense to standardize the output from the model to expected quantity per week so we can compare “apples-to-apples”. Finally, a rank was assigned to the standardized expected quantity sold.

Multiple Linear Regression Model Inputs:									
Seasonality Index	Duration (Days)	Depth	Insert	Front page	Center page	Days Since Last Promo	Expected Quantity Sold	Expected Quantity Sold per Wk	Rank
100	1	0.67	1	0	0	14	34,488.29	206,929.72	1
100	6	0.73	0	1	0	28	34,138.62	34,138.62	2
100	6	0.67	0	0	1	28	34,013.40	34,013.40	3
100	6	0.72	0	1	0	35	33,907.76	33,907.76	4
100	6	0.68	0	0	1	35	32,909.28	32,909.28	5
100	6	0.72	0	0	1	28	31,830.25	31,830.25	6
100	6	0.72	0	0	1	33	31,359.19	31,359.19	7
100	6	0.79	0	0	1	28	28,773.84	28,773.84	8
100	6	0.85	0	0	1	21	26,821.54	26,821.54	9
100	6	0.84	0	0	1	42	25,255.72	25,255.72	10
100	6	0.84	0	0	1	49	24,588.23	24,588.23	11
100	6	0.92	0	0	1	14	24,432.62	24,432.62	12
100	6	0.84	0	0	1	56	23,920.74	23,920.74	13
100	6	0.92	0	0	1	28	23,097.65	23,097.65	14
100	6	0.92	0	0	1	35	22,430.16	22,430.16	15
100	6	0.92	0	0	1	42	21,762.68	21,762.68	16
100	6	0.91	0	0	1	49	21,531.82	21,531.82	17
100	27	0.97	0	0	0	49	89,459.39	19,879.86	18
100	6	0.8	0	0	1	126	18,992.40	18,992.40	19
100	6	0.87	0	0	0	7	6,921.26	6,921.26	20
100	6	0.92	0	0	0	7	4,738.11	4,738.11	21

**Table 7. Calculating, standardizing, and ranking quantity sold for various types of tried promotions**

Noticeably, the only one-day sale in the data produced an astronomical expectation for quantity sold per week because so much was sold in one single day. Extrapolating a one-day sale scenario over the duration of a week doesn't make sense in terms of accurately predicting the number items sold. However, it does show how tremendously effective the sale was considering it was so short, yet it could compete with weeklong promotions with the same depth of sale. The only difference between the one-day promo and the weeklong promo's was that the one-day promo advertised on the circular's insert-wrap around vs. one of the center pages. A second difference was that the one-day sale was 14 days after the previous sale vs. 28 days after the previous sale for the week long promo. Perhaps this says something important about duration of promotions: shoppers would get their product X on-sale regardless if we make the promo available for the whole week or only one day. According to that logic, it may be wiser to only offer product X on sale for shorter time periods assuming the longer the opportunity we provide for the

customers to purchase the product, the more likely customers who didn't consult the ad but had product X on their shopping list for the week, would luck out and get a great deal too. In other words, the one-day promotion was much more effective at driving traffic into the stores, where the week-long promotion may have driven traffic too, but likely sold product X to customers who weren't shopping for said product X bargains.

In general, the less deep the discount, the fewer items were purchased, as expected. However, there are some instances when less deep promotions yields higher expected quantity sold values. For example, promos with ranks 2 and 4 are worse deals than ranks 3 and 5, respectively, but the earlier promo's sold slightly more units than their more expensive counterpart when we'd expect them to sell less on average. The only other difference was their location in the ad circular. The more expensive product X was featured on the front page and sold more units vs. the cheaper product X was featured in the center pages. In other words, the product X got slightly more hype being on the front page, which stood out to customers and convinced them that "now is the time to buy" even though it wasn't necessarily the best deal. In general, ad-placement heightens the perceived savings of a given promotion and can be a useful tool to drive traffic into stores as well.

## **CONCLUSION**

In conclusion, insights into the effectiveness of the promotional pricing attempts are useful because they help us identify areas ripe for improvement that, after some tweaking, can drive the business' bottom line. Careful excavation of the factors that influence quantity sold during the promotional period using SAS® Enterprise Guide 6.1 proved to be worthwhile when predicting sales quantity. Specifically, we learned that product category seasonality of demand, the days and depth of the promotion, and a feature in the insert, front page, or center page of the circular are impactful measurements for product X. With this knowledge, the retailer's team may suggest tweaks to these promotion efforts to optimize a promotion. Next, evaluating what constitutes a "sufficiently accurate prediction" exposed that the planning forecasting for product x on promotion was usually over estimated, sometimes by a large margin. Furthermore, the process of ranking promotions relative to one another allowed us to produce results that help the retailer support its business areas and make well-informed promotion decisions. Utilizing the correct type of promotion at the correct time and marketed to the correct customer is imperative to increasing traffic, and thus storewide revenue. Specifically, one-day promotions are very effective at driving traffic into the stores and the product's ad placement can make or break the visibility of the customer's perceived savings. In conclusion, using SAS® Enterprise Guide 6.1 allowed us to develop a sophisticated data-driven tool to field future promotional effectiveness requests.

## **REFERENCES**

David L. Cassell. "Don't Be Loopy: Re-Sampling and Simulation the SAS® Way". Proceedings of the SAS Global Forum 2007. Web. <<http://www2.sas.com/proceedings/forum2007/183-2007.pdf>>

## **CONTACT INFORMATION**

Your comments and questions are valued and encouraged. Contact the author at:

Name: Aaron Clark  
Email: [clark.aaronchris@gmail.com](mailto:clark.aaronchris@gmail.com)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.