Case Study 2: GoodBelly

MSAN 601

Justin Midiri, Sheri Nguyen

October 14, 2016

1 Executive Summary

Through our analysis of GoodBellys promotional programs, particularly the Endcap Promotions and the in-store demonstrations, we found that overall the programs yielded an expected increase in units sold per store per week. We concluded that for Endcap Promotions, there was an expected 242% increase with the presence of a regional sales representative. For in-store demonstrations, we noticed a temporary increase in units sold and a reversion to previous sale numbers as the weeks progressed from the last demonstration. However, for a further and more in-depth conclusion, we would need the costs associated with the promotional programs to properly analyze ROI.

2 Introduction

The following is a case study to examine if the use of endcaps and demos boosted GoodBelly product sales. This study also explores whether the change in sales, if any, by these two marketing tactics are short-lived or long-lived.

3 Detailed Processes

3.1 Data and Testing Criteria

The dataset used is a compilation of 10 weeks worth of data from 126 Whole Foods stores and contains a total of 1386 observations over the weeks of May 4, 2010 to June 13, 2010. The variables in this case study are defined as follows:

Response Variable:

- **UnitsSold** : the number of units sold per store per week

Independent Variables:

- **Avg.RetailPrice**: the average retail price for GoodBelly products per store per week
- **SalesRep**: defined as 1 if the store had a regional sales rep (face-to-face contact) and 0 if the store had only the national sales rep (no face-to-face contact)
- **Endcap** : defined as 1 if a store participated in an endcap promotion
- **Demo**: defined as 1 if the store had a demo on the corresponding week
- **Demo1.3**: defined as 1 if the store had a demo 1-3 weeks ago
- **Demo4.5**: defined as 1 if the store had a demo at least 4-5 weeks ago
- **Natural Retailers**: the number of other natural retailers within 5 miles of each Whole Foods
- **Fitness Centers**: the number of fitness centers within 5 miles of each store

All tests in this study were compared to significance level of $\alpha = .05$. We declared all p-values less than α to be statistically significant. In order to test how well our model performs. To do this, we randomly sampled two thirds of our 1386 and assigned those rows to be our training set (data that we built our model on) and then we assigned the rest to be our testing set (data that we tested the performance of our model on).

3.2 Preliminary Examination

We began our investigation by looking at the relationship between Units Sold and Average Retail Price in our training set. To our surprise, we found that there wasnt a linear relationship between the two (figure 1).

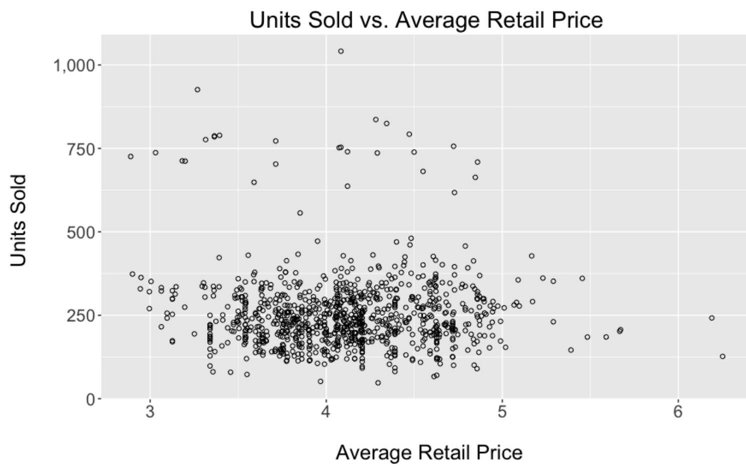


Figure 1: Average Retail Prices vs Units Sold

Based on the data, the units sold stayed relatively consistent regardless of the price level. We ran a model of $UnitsSold$ on $Av.RetailPrice$ to confirm our interpretation and got the following results with an Adjusted R-squared of -0.0003678:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	272.3267	26.5828	10.24	0.0000
Avg. Retail Price	-4.5059	6.4316	-0.70	0.4837

Table 1: Units Sold $\sim$ Avg. Retail Price

In addition to looking at their relationship, we also looked to see if the model is statistically insignificant because a transformation of the data is required. We performed normality tests on various potential transformations of our model and did not achieve success. From our preliminary examination, we established that for the scope of our study we did not need to include $Avg.RetailPrice$

3.3 Training the Model

We approached building our model with the backward elimination method and started with the full model in Appendix 4.1. Firstly, we noticed that the residuals are slightly left skewed (Appendix 4.10). Then, we eliminated both $Natural$ and $Fitness$ due to their insignificance in our model (Appendix 4.2, 4.3). Next, we checked if there were any terms that we were missing from the following model after our backward elimination:

$$\begin{aligned} UnitsSold = 180.1 + 65.8(SalesRep) + 333.4(Endcap) + \\ 108.7(Demo) + 80.7(Demo1.3) + 72.7(Demo4.5) \end{aligned} \tag{1}$$

After we ran the Ramsey RESET test with results suggesting that we have some missing second order terms, we checked for significant interaction terms first (Appendix 4.4). We then proceeded to add the proposed interaction terms, $SalesRep{*}Endcap$ and $SalesRep{*}Demo1.3$ (Appendix 4.5), and reran the RESET test to see if we were missing any other second-order variables. Our main issue with this model was that we were afraid that we were compromising interpretability power, so we delved deeper into the following equation:

$$\begin{aligned} UnitsSold = 187.2 + 53.9(SalesRep) + 3.1(Endcap) + 107.5(Demo) + 99.6(Demo1.3) + \\ 74.4(Demo4.5) + 457.3(SalesRep * Endcap) - 26.9(SalesRep * Demo1.3) \end{aligned} \tag{2}$$

Intuitively, we understood why $SalesRep * Endcap$ was important to add to the model– the presence of a sales representative would drive up the number of endcaps any given region has. However, we couldn't quite think through the use the second interaction term, $SalesRep * Demo1.3$. We ran the following models to compare our four model criteria (${R_a}^2$, MSE, BIC, Mallow's Cp):

1. UnitsSold ∼ SalesRep + Endcap + Demo + Demo1.3 + Demo4.5

2. UnitsSold ∼ SalesRep + Endcap + Demo + Demo1.3 + Demo4.5 + SalesRep*Endcap

3. UnitsSold ∼ SalesRep + Endcap + Demo + Demo1.3 + Demo4.5 + SalesRep*Demo1.3

We noticed that all of our criteria increased from model (1) to models (2) or (3). When we ran model (3), we noticed that the coefficient for $SalesRep * Demo1.3$ actually became insignificant. However, model (2) stayed significant and had the same ${R_a}^2$ as the model that included both of our interaction terms.

Once we established our best model (model (2) above) we ran tests for heterskedasticity and missing terms. With a p-value of 0.2117 for our Breusch-Pagan test and a p-value of 0.06377 for our RESET test, we reject that there is heteroskedasticity and missing terms in our final model. We also check for normality using QQPlots and see that our distribution of residuals is approximately normal (Appendix 4.9).

3.4 Testing the Model

After stating our final model, we moved forward to test how well the model does on our testing data set and compared our MSE of our training data to the MSPR in our test data (Appendix 4.6). Our MSE was 2552 and our MSPR was 2462.564. We felt that was a relatively small, approximately 3%, difference between the two. We then proceeded to see how well our model does relative to the model based on the whole dataset, and got an MSE of 2494 (Appendix 4.7), approximately 2% different than our training model.

3.5 Conclusion

We conclude that our model:

$$UnitsSold \sim SalesRep + Endcap + Demo + Demo1.3 + Demo4.5 + SalesRep * Endcap \quad (3)$$

does quite a good job predicting values of $UnitsSold$. The number of units sold per store per week was expected to be 188 units with no regional sales rep, no endcap display, and no demonstrations performed. From our model, we concluded that the presence of a sales rep, the stores involvement in the endcap promotion, and the presence of demonstrations significantly increases the expected number of units sold. When we analyzed units sold with respects to endcap promotions, we found that the largest increase in expected units sold was when there was a sales representative at a store running endcap promotions.

There was an expected 242% increase in expected sales at a given store. We found that the greatest ROI, an expected increase of 57% in sales, for product demonstrations came from when the demo was performed that same week. However, we noticed that sales started to revert to normal levels as time progressed from the first demo. Within a period of 5 weeks, the ROI changed from 57% to 39%. Therefore, we conclude that the endcap promotion in tandem with the presence of a regional sales representative can greatly boost expected sales while the implementation of demonstrations may only have a temporary effect.

3.5.1 Further Analysis

For further analysis, we can better conclude a proper ROI per program run by GoodBelly if given the costs of each program. We also believe that knowing the area population and the amount of people who visit a particular store can influence whether or not GoodBelly should invest in promotions at that store. For example, areas with a higher population, such as San Francisco, that shop at Whole Foods may be worth more marketing investment since the expected ROI would be greater than running programs in less populated areas such as Hartly Delaware.

4 Appendix

4.1

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	184.8849	5.5656	33.22	0.0000
Sales.Rep	66.4153	4.5516	14.59	0.0000
Endcap	332.9516	10.9832	30.31	0.0000
Demo	108.7698	9.0778	11.98	0.0000
Demo1.3	80.4615	5.9919	13.43	0.0000
Demo4.5	72.1291	7.9527	9.07	0.0000
Natural	-1.5853	2.2214	-0.71	0.4756
Fitness	-1.1456	1.3536	-0.85	0.3976

Residual standard error: 64.39 on 916 degrees of freedom
Multiple R-squared: 0.6875, Adjusted R-squared: 0.6851
F-statistic: 287.9 on 7 and 916 DF, p-value: < 2.2e-16

Table 2: Units Sold ∼ All independent variables excluding Avg. Retail Price

4.2

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	182.6341	4.5845	39.84	0.0000
Sales.Rep	66.0669	4.5242	14.60	0.0000
Endcap	332.4701	10.9595	30.34	0.0000
Demo	108.8529	9.0746	12.00	0.0000
Demo1.3	80.6555	5.9842	13.48	0.0000
Demo4.5	72.3620	7.9438	9.11	0.0000
Fitness	-1.0780	1.3499	-0.80	0.4248

Residual standard error: 64.37 on 917 degrees of freedom
Multiple R-squared: 0.6873, Adjusted R-squared: 0.6853
F-statistic: 336 on 6 and 917 DF, p-value: < 2.2e-16

Table 3: Units Sold ∼ All except Avg. Retail Price and Natural

4.3

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	180.0589	3.2578	55.27	0.0000
Sales.Rep	65.7664	4.5076	14.59	0.0000
Endcap	333.3614	10.9004	30.58	0.0000
Demo	108.6841	9.0704	11.98	0.0000
Demo1.3	80.7422	5.9820	13.50	0.0000
Demo4.5	72.6863	7.9319	9.16	0.0000

Residual standard error: 64.36 on 918 degrees of freedom
Multiple R-squared: 0.6871, Adjusted R-squared: 0.6854
F-statistic: 403.2 on 5 and 918 DF, p-value: $< 2.2e-16$

Table 4: Units Sold $\sim$ Sales Rep + Endcap + Demo + Demo1.3 + Demo4.5

4.4

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	186.9631	2.6971	69.32	0.0000
Sales.Rep	54.7863	3.9447	13.89	0.0000
Endcap	3.9071	16.5333	0.24	0.8132
Demo	109.6820	17.2746	6.35	0.0000
Demo1.3	100.1849	11.3609	8.82	0.0000
Demo4.5	79.4930	12.8863	6.17	0.0000
Sales.Rep:Endcap	455.5297	20.0328	22.74	0.0000
Sales.Rep:Demo	-11.2080	19.3857	-0.58	0.5633
Sales.Rep:Demo1.3	-27.5080	12.6698	-2.17	0.0302
Sales.Rep:Demo4.5	-3.4805	15.2101	-0.23	0.8191
Endcap:Demo	-0.9358	25.6958	-0.04	0.9710
Endcap:Demo1.3	-6.2162	19.7906	-0.31	0.7535
Endcap:Demo4.5	93.8073	68.8641	1.36	0.1735
Demo:Demo1.3	27.7198	18.4576	1.50	0.1335
Demo:Demo4.5	6.5675	37.9186	0.17	0.8625
Demo1.3:Demo4.5	-28.0921	18.0374	-1.56	0.1197

Table 5: Testing for two-way interaction term signifcance

4.5

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	187.2422	2.6294	71.21	0.0000
Sales.Rep	53.8779	3.7206	14.48	0.0000
Endcap	3.1811	16.1648	0.20	0.8440
Demo	107.5403	7.1059	15.13	0.0000
Demo1.3	99.6444	11.3109	8.81	0.0000
Demo4.5	74.4020	6.2137	11.97	0.0000
Sales.Rep:Endcap	457.2511	19.0005	24.07	0.0000
Sales.Rep:Demo1.3	-26.8709	12.4262	-2.16	0.0308

Residual standard error: 50.41 on 916 degrees of freedom
Multiple R-squared: 0.8085, Adjusted R-squared: 0.807
F-statistic: 552.3 on 7 and 916 DF, p-value: < 2.2e-16

Table 6: Model including significant interaction terms

4.6

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Sales.Rep	1	2385357.18	2385357.18	934.88	0.0000
Endcap	1	4332729.22	4332729.22	1698.10	0.0000
Demo	1	558296.85	558296.85	218.81	0.0000
Demo1.3	1	726190.53	726190.53	284.61	0.0000
Demo4.5	1	347837.81	347837.81	136.33	0.0000
Sales.Rep:Endcap	1	1462751.06	1462751.06	573.29	0.0000
Residuals	917	2339735.38	2551.51		

Table 7: Anova II for Training Model

4.7

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Sales.Rep	1	3446903.62	3446903.62	1382.09	0.0000
Endcap	1	5557089.45	5557089.45	2228.21	0.0000
Demo	1	856789.90	856789.90	343.54	0.0000
Demo1.3	1	988019.29	988019.29	396.16	0.0000
Demo4.5	1	409542.90	409542.90	164.21	0.0000
Sales.Rep:Endcap	1	2366660.44	2366660.44	948.95	0.0000
Residuals	1379	3439189.26	2493.97		

Table 8: Anova II for Whole Dataset

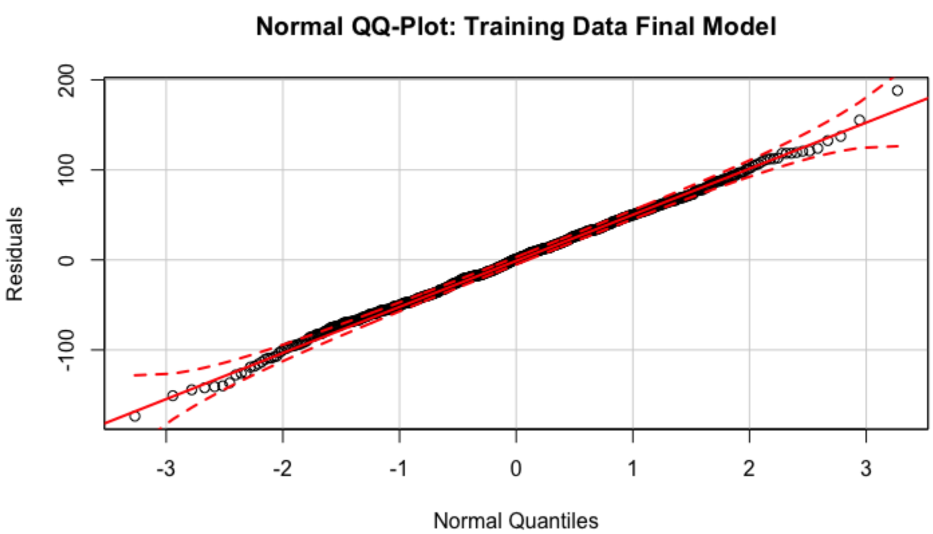
4.8

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	188.3900	2.5804	73.01	0.0000
Sales.Rep	51.6805	3.5864	14.41	0.0000
Endcap	4.2849	16.1891	0.26	0.7913
Demo	107.2839	7.1191	15.07	0.0000
Demo1.3	77.3846	4.6971	16.48	0.0000
Demo4.5	74.2331	6.2257	11.92	0.0000
Sales.Rep:Endcap	455.3720	19.0187	23.94	0.0000

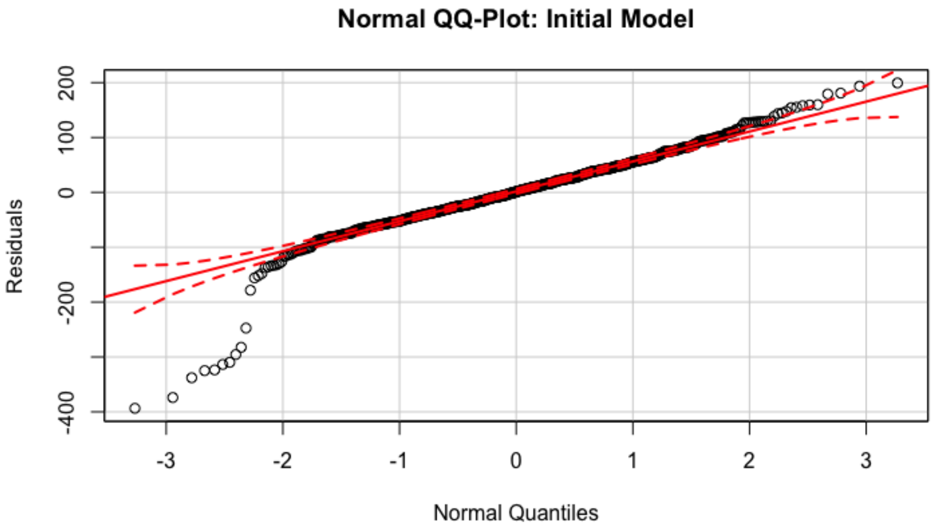
Residual standard error: 50.51 on 917 degrees of freedom
Multiple R-squared: 0.8075, Adjusted R-squared: 0.8062
F-statistic: 641 on 6 and 917 DF, p-value: < 2.2e-16

Table 9: Final Model

4.9



4.10


Normal QQ-Plot: Initial Model
Residuals
200
100
0
-200
-400
-3
-2
-1
0
1
2
3
Normal Quantiles