8-2008

Analyzing the Effects of Coupons and Promotion in the Grocery Retail Sector

Milena Hanna Chotard
University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a thesis written by Milena Hanna Chotard entitled "Analyzing the Effects of Coupons and Promotion in the Grocery Retail Sector." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Statistics.

Dr. Russell Zaretzki, Major Professor

We have read this thesis and recommend its acceptance:

Dr. Mary Leitnaker, Dr. Kenneth Gilbert

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a thesis written by Milena Hanna Chotard entitled "Analyzing the Effects of Coupons and Promotion in the Grocery Retail Sector." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Statistics.

Dr. Russell Zaretzki, Major Professor

I have read this thesis and
recommend its acceptance:

Dr. Mary Leitnaker

Dr. Kenneth Gilbert

Accepted for the Council:

Carolyn R. Hodges, Vice Provost and
Dean of the Graduate School

(Original signatures are on file with official student records.)

Analyzing the Effects of Coupons and Promotion in the Grocery Retail Sector

A Thesis Presented for
the Master of Science
Degree
The University of Tennessee, Knoxville

Milena Hanna Chotard
August 2008

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my major advisor, Dr. Russell Zaretzki, for giving me the inspiration to write this Thesis and providing the doorway to the data and analysis. Without his support and constant positive attitude this endeavor would have been most difficult and unattainable.

Secondly, much appreciation and thanks goes to Dr. Mary Leitnaker for being the most flexible and supportive advisor. Thank you so much for your time and exploration of this analysis.

I would also like to thank the University Libraries for the use of their collections.

Furthermore, a big thanks to dunnhumbyUSA for the use of their data. Their generous contributions to academia are very much appreciated and treasured – it is my hope that this Thesis ignites even more interest for future students to study and analyze this data.

Last, but not least, I would like to thank my family and friends for their support. My husband deserves the most credit as he provided the initial inspiration for me to go back to school and never wavered in his constant encouragement. Thank you for your positivism and enduring strength. A big thank you to my daughter Lauren for providing hugs and kisses throughout the writing of this Thesis. And a huge thank you to my parents and parents in law in their interest and questions that propelled me to think outside of “the box”.

ABSTRACT

Modern scanner technology is pervasive throughout the retailing sector of the economy and is almost universal in the food retail industry. Along with loyalty programs, it has led to the development of massive databases which accurately and proficiently track the purchasing habits of customers. Making use of this information is one of the most important efforts in the management of this sector to further increase profitability.

This thesis explores the application of several statistical techniques to extract specific information from two large databases of customer purchasing behavior at a major US grocery chain. In particular, we first focus on the impact of coupon use on brand loyalty in two commodity groups, pasta with sauce and pancake mix with syrup. Furthermore, we devise a graphical tool to visualize relationships between complementary commodities in order to aid retailers and brand managers accurate portrayal and measurement of the success of their specific products. Next we consider a larger database with complete purchasing information for 2500 frequent shopper households. Here we develop tests and models to evaluate the effect of direct marketing on customer loyalty and spend. Based on our results, we discuss the impact of direct marketing and coupons to retail firms as well as consumer products manufacturers.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I
INTRODUCTION AND GENERAL INFORMATION

The Elusive Grocery Shopper

There have been many studies constructed within academia, corporate, and even third-party statistical data gathering groups that have been centered on the grocery shopper psyche. While the current statistical assessment is of interest in these studies, the ability to predict the shopper's next movements and purchase patterns is equally (if not more) important. Such information could enable brand managers and grocery store owners to make effective buying and stocking choices. These effective buying and stocking choices would inevitably lead to greater productivity, reduced operating cost, increased profit, and ultimately reduced prices for the end consumer.

But how does one assess the grocery shopper's *current* state of mind?

Even more importantly, how does one assess the grocery shopper's *future* state of mind?

Does the grocery shopper *himself* know what he will buy tomorrow? Next week? Next year?

These kinds of questions haunt our brand managers and grocery store owners on a daily basis. To have intrinsic knowledge of the future, to have the keys to the grocery shopper's brain, to have the ability to adapt a buying schedule based on accurate and easy to understand data would simplify and solidify the grocery retail world.

The real questions are; where do we obtain this information, does it really exist, and how can it be extracted from a large database?

Data, Statistics, and More Data

The existence of large databases provides a huge amount of potential information to drive decision making. Unfortunately, much of this information may be of little value making it hard to know which data to trust and which data to stay away from.

Advertising and Coupon Data

Articles, journals, and independent studies (Supermarket News, Tech Solutions, Brandweek, etc.) reference grocery store shopping pattern data frequently. Cause and effect is asserted on the basis of simple percentages. There is no check and balance applied; no statistical consultant's thumb raised. The analyst's own intuition and experience accepts or rejects the null hypothesis that the data is trustworthy to use and abuse.

For example, consider the following ambiguous passage:

> "According to the Customer Focus 2004: Grocery Study released by Vertis (Baltimore), 71% of female chief grocery shoppers (i.e. those responsible for 60% or more of household grocery shopping) who read advertising inserts make lists and plan their grocery shopping efforts based on items viewed in advertising inserts or circulars. Therese Mulvey, vice president of marketing research at Vertis, notes, "Grocery marketers who want to have an impact with female chief grocery shoppers should consider the significant role this medium has in determining which items are purchased and where they purchase them."[1]

[1] www.preparedfoods.com June 2004

The suggestion within the excerpt of the above article implies that advertising inserts are the key to the female grocery store shopping psyche. As it is straight-forward and simple, we must ask ourselves, was this a posed question to the female shoppers? How well do they know themselves and their shopping actions? Furthermore, was this a self-assessment based on behavior recalled over the past month? Week?

As further examples, consider the following vague coupon claims:

1) Volume

> "According to new data from NCH Marketing Services, CPG (consumer packaged goods) marketers distributed $258 billion coupons in 2003, a 4% increase from 2002, according to NCH, a provider of coupon processing and marketing services. Redemption, meanwhile, slipped to 3.6 billion coupons, a 5.3% decrease from 2002… While overall redemption may be down, marketers SN polled reported good results from their individual efforts. "We've had very good redemption results," said Rob Lorys, vice president, consumer marketing, Georgia-Pacific Corp., Atlanta."[2]

2) Couponing Goals

> "Consumer packaged goods marketers remain active in couponing for a variety of reasons. Del Monte Foods Co., San Francisco, for instance, uses coupons to attract new consumers and reward existing ones, according to Melissa Murphy, company spokeswoman. Del Monte uses a mix of couponing methods, including FSIs, direct mail, in-store and on-pack.
>
> ConAgra Foods, Omaha, Neb., views couponing as an important way to drive trial on new products and get consumers to retry products after product improvements have been made, according to Chris Kircher, the company's vice president of communications. Couponing is also a way to suggest new usage occasions for products, Kircher added.

[2] Angrisani, Carol, "Coupon Confidence; Consumers May Not be Embracing Coupons Like They Used to, but Marketers Reamin Committed to the Promotional Tool", Supermarket News, April 5, 2004

ConAgra distributes coupons through a number of methods, including FSIs, direct mail, in-pack, in-store and the Internet. "Our methods vary according to specific brand and portfolio objectives," Kircher said.

ConAgra is moving away from the use of coupons as a price subsidy on existing brands in favor of equity-building programs, Kircher noted.

At Georgia-Pacific, meanwhile, the main goal of couponing is to stimulate trial of new products and encourage pantry-loading of its brands."[3]

3) The Internet

The Internet accounted for 0.2% of coupon distribution in 2003, on par with 2002. Despite the highly publicized cases of Internet coupon fraud in some markets last year, many marketers remain supportive of the Internet as a coupon distribution tool.

Georgia-Pacific is one of them. Internet couponing is an important part of the company's integrated marketing approach. One reason is that its target market is women with children, and this demographic spends a lot of time on the Internet, Lorys said.[4]

Do coupons lead to brand loyal customers? Does direct marketing provide an increase in profitability and/or frequency of purchases?

Procter and Gamble, with a reputation as a firm with a high quality marketing department, has taken heed of this data as noted in the following article:

"Procter and Gamble will initiate its first zero-couponing effort in western New York state, an effort the company says will function as a testing ground for its concerns over the viability of coupons. Company executives say that they believe the effectiveness of coupons in marketing has diminished substantially in recent years, and will investigate whether other media can fill the coupon's traditional role. Statistics indicate that approximately $3 billion of the $6.5 billion spent annually on coupons by industry goes toward administrative and production costs."[5]

---

[3] Ibid
[4] Ibid
[5] Tenser, James. :P&G Sets Zero-Couponing Test" Supermarket News, 46.n3, Jan 15, 1996

Despite such reports, internet coupons have recently spawned more cause and effect conclusions:

"Coupons, whether issued by manufacturers or by retailers themselves, can be an effective promotional tool. Widely distributed through mail, magazines and newspapers, they can promote repeat purchases, build brand loyalty, encourage trial, provide product exposure and attract new customers."[6]

A different trend, however, is being realized within the coupon world:

"It might come as a surprise to many that coupon-redemption rates are actually declining. When a marketing tool as old and powerful as coupons begins to slip, it's clearly time for brand managers and retailers alike to figure out what they're doing wrong. There's really only one question that needs to be asked here: What prompts a customer to redeem a coupon in the first place? It's a very simple question. What surprises me is that too many in our field don't know how to answer it."[7]

Brand Loyalty Data

First and foremost, in order to understand whether coupons lead to brand loyalty, the latter's definition needs to be addressed. What makes a customer brand loyal? How do we measure brand loyalty?

Once the definition is classified, the question of coupons building brand loyalty can be answered.

The Goal of this Thesis

It is unlikely that all of our grocery shopper psyche questions will be answered in the current Thesis. However, it will shed some light on the relationship between a coupon shopper and a brand loyal one. It will also provide insight into whether coupon circulation is truly beneficial for the brand manager and/or the grocery

[6] Amato-McCoy, Deena M. "Print and Save" Tech Solutions, October, 2005
[7] Meyers, Peter, "Redemption Revisition" Brandweek 48.38, October 22, 2007

store owner. Furthermore, it will provide suggestions for standardizing operational definitions of brand loyalty and coupon use, an important step in the measurement and evaluation of these concepts.

It will also look at the effect of direct marketing campaigns and promotions as provided from the grocery chain sector. It will encompass the overall outcome of a direct mail marketing campaign, whether positive or negative, and develop a model as to the future prediction and inference of this data.

Best of all, this Thesis will provide the gateway to further study of the grocery shopper psyche. It is but a doorway into the studies that may develop in the future. The answers obtained herein will develop into better questions; a nudge towards a better understanding of ourselves and our counterparts as we traverse the grocery aisles of tomorrow.

CHAPTER II
THE DUNNHUMBY DATA SET

The data referenced and analyzed throughout this Thesis was obtained from dunnhumbyUSA, a specialized provider of customer focused analytical and targeting services. Dunnhumby uses statistical research and analysis to provide companies with much needed help with marketing and business development. Meanwhile, some of the data obtained through these relationships are provided for academic research and exploration.

The dunnhumbyUSA grocery store data investigated in this Thesis was obtained via two Source Files: Carbo-Loading and The Complete Journey. Each contains household level transactions over a two year period, pooling together a picture of customer-related decisions and shopping patterns.

Carbo-Loading Source Files are a compilation of tables that have detailed transaction level information within four product categories: pasta, pasta sauce, syrup, and pancake mix.

The Complete Journey Source Files also house detailed transaction level information but on a much more complicated level: all product purchases are accounted for, not just within a select few categories. For this specific purpose, a select group of households were chosen for this data set.

The next couple of pages will highlight the details within each Source File and their intricate similarities and differences, as well as summary data of tables within each Source File.

Carbo-Loading Source File

The pasta, pasta sauce, syrup, and pancake mix household level transaction data was obtained through a loyalty card program of a leading US grocer. These transactions were monitored over a two year period and total over 5 million specific product purchases. These 5 million specific product purchases were documented across 387 unique stores of the leading US grocer. A total of 927 different products within the four commodities were recorded during this time period.

Furthermore, each product's location within a specific weekly mailer was documented and tracked over the monitored two year period.

Figure 1 represents the tables that are included in the Carbo-Loading Source File, and their relationships with one another. All of the tables are centered around the main transactions data table, which houses the product-level purchases.

The Complete Journey Source File

The Complete Journey data tables are represented in Figure 2. With a direct comparison to that of the data tables in the Carbo-Loading data, we can clearly see that The Complete Journey data is a little more complex. Over this two year period, 2500 households were chosen for this data's tracking purposes, which were classified as frequent shoppers of the grocery store chain. Instead of 927 different products within just 4 Carbo-Loading categories, 92,358 products were tracked over a two year period within the Complete Journey Source File.

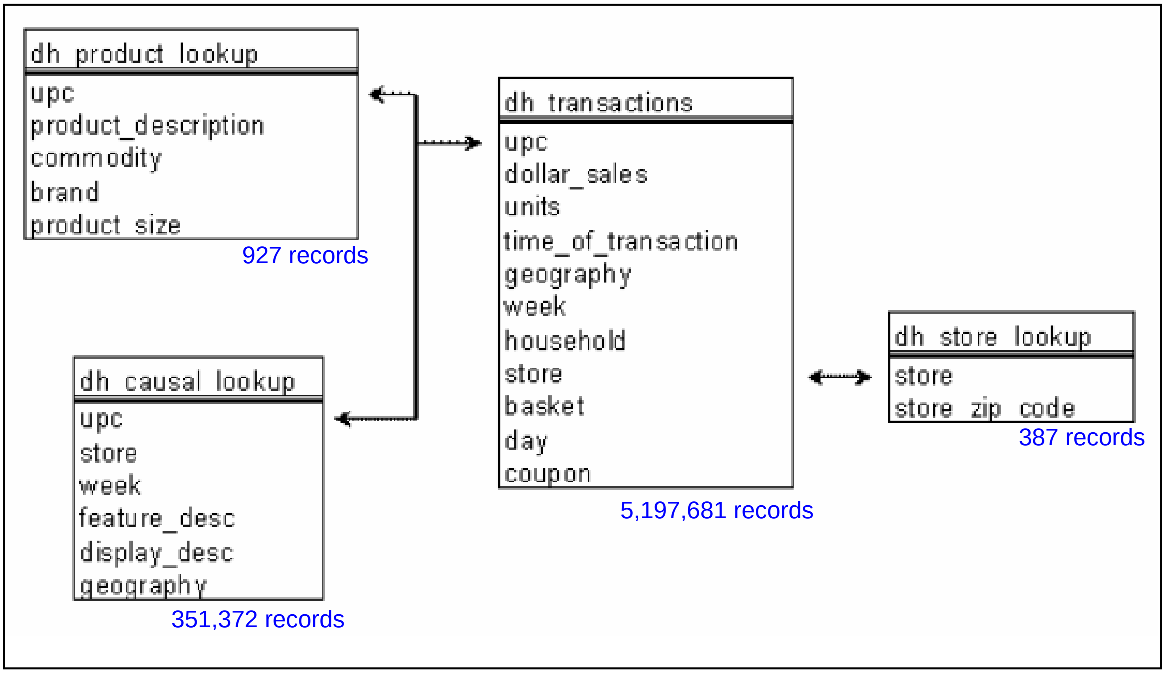

Figure 1. Carbo-Loading: Data Overview and Table Map.[8]

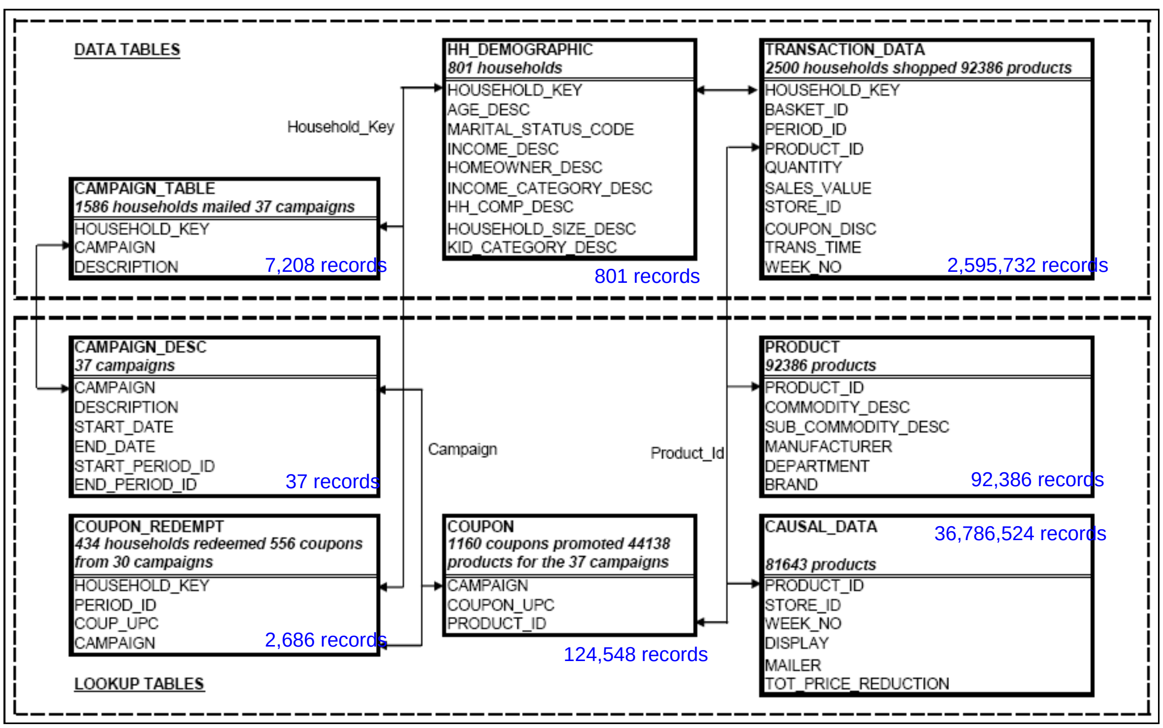

Figure 2. The Complete Journey: Data Overview and Table Map.[9]

[8] DunnhumbyUSA. Dunnhumby Source Files. USA. 2001.
[9] Ibid

All of these households' purchases are contained within the transactions table, available for a complete and thorough analysis of the shopper psyche.

Transactions

For the entire two year period in our data, 2500 frequent shoppers were tracked on the basis of the grocery transactions. Every single purchase is accounted for, on every single shopping trip. Price information, coupon use information is readily available. The quantity of a specific product is also measured.

From this data we can easily surmise the longitudinal aspect of the grocery shopper psyche. For the purposes of this Thesis, we will focus on the effects and strengths of direct mail marketing programs. However, more analysis and study will be questioned for future papers regarding this data.

Demographics

32% of our Complete Journey data has available demographic information. Out of this group we have the following information:

- estimated age range
- marital status (single or married)
- estimated income range
- homeownership (vs. renter)
- household composition (adults and kids)

Table 1 summarizes the age information of our households that have the demographics available. The bulk of the customers within our demographics section are within the 25 to 55 year range, composing almost 80% of our frequent shopper list. (The household age range data was obtained by querying the Source Files within SAS. All SAS code used within this Thesis going forward is available within the Appendix.)

Table 1. Household Age Range.

Estimated Age Range	*Number of Households*	*Percent*
19-24	46	5.74%
25-34	142	17.73%
35-44	194	24.22%
45-54	288	35.96%
55-64	59	7.37%
65+	72	8.99%

Table 2 summarizes the marital status information of our households that have the demographics available. Only 57% of our data has an identified marital status (A – married, B – single). The rest of the data is unknown.

Table 3 summarizes the homeownership information of our households that have the demographics available. 63% of our frequent shoppers are homeowners, 6% are renters (or probable renters) and the rest of the data is unknown.

Table 4 summarizes the household composition of our households that have the demographics available. The bulk of our frequent shoppers are either single (male or female), 2 adults with no kids, or 2 adults with kids. Only 9% of our data is unknown in this distribution.

Marketing

A total of 37 direct mailers were sent to our 2500 select households over the two year tracking period. The 37 campaigns promoted 1160 coupons for 44,138 distinct products. Out of these 37 campaigns, 434 households redeemed a total of 556 coupons from 30 of the campaigns. In other words, a little over 17% of our frequent shoppers actually used the coupons that were mailed directly to their places of residence.

Table 2. Household Marital Status.

Marital Status	*Number of Households*	*Percent*
A	340	42.45%
B	117	14.61%
U	344	42.95%

Table 3. Household Homeownership.

Homeowner Description	*Number of Households*	*Percent*
Homeowner	504	62.92%
Probable Homeowner	11	1.37%
Probable Renter	11	1.37%
Renter	42	5.24%
Unknown	233	29.09%

Table 4. Household Composition.

Household Composition	*Number of Households*	*Percent*
1 Adult Kids	47	5.87%
2 Adults Kids	187	23.35%
2 Adults No Kids	255	31.84%
Single Female	144	17.98%
Single Male	95	11.86%
Unknown	73	9.11%

CHAPTER III
CARBO-LOADING BASICS

Household Product Penetration

The household penetration of each product within each commodity within our Carbo-Loading data gives us a simple view of preferences within our grocery shopper population. There are a total 510,027 distinct households within this data set, and Table 5 shows how many distinct households within this total actually shopped in each of our four commodities.

Pasta

Table 6 shows the product/brand penetration of the Pasta commodity group. For the sake of brevity, only the top 5 products have been shown. It is surprising (as compared to the other commodities that will be mentioned in the next paragraphs) that the Private Label brand reigns within the pasta commodity. It looks like our grocery shoppers are more price conscious within this area and that the Private Label manufacturers have a solid market share of the sales.

Pasta Sauce

Table 7 shows the product/brand penetration of the Pasta Sauce commodity group. As before, for the sake of brevity only the top 5 products have been shown. Contrary to the Pasta commodity finding, the Private Label brand does not have control of the market share within this group; it holds the third spot in which only about 25% of the shoppers that do buy pasta sauce tend to buy the Private Label brand. Since Pasta and Pasta Sauce do go hand in hand, the Private Label brand managers may have an opportunity within this area to do cross-commodity marketing.

Table 5. Household Commodity Penetration.

Commodity	*Distinct Households*	*Percent of Total Population*
Pasta	411,601	80.70%
Pasta Sauce	358,600	70.31%
Syrups	256,250	50.24%
Pancake Mixes	130,580	25.60%

Table 6. Pasta Penetration.

Pasta	*Distinct Households*	*Percent of Total Population*
Private Label	267,358	64.96%
Barilla	125,579	30.51%
Mueller	91,128	22.14%
Creamette	86,672	21.06%
Private Label Premium	76,558	18.60%

Table 7. Pasta Sauce Penetration.

Pasta Sauce	*Distinct Households*	*Percent of Total Population*
Ragu	207,237	57.79%
Prego	118,132	32.94%
Private Label	90,754	25.31%
Hunt's	69,509	19.38%
Classico	58,865	16.42%

Pancake Mixes

Table 8 shows the product/brand penetration of the Pancake Mixes commodity group. As before, for the sake of brevity only the top 5 products have been shown. Just as in the Pasta Sauce commodity finding, the Private Label brand does not have control of the market share within this group; it holds the third spot in which only about 25% of the shoppers that do buy pancake mixes do buy the Private Label brand. There may be some opportunity within this commodity to grab a larger portion of the market share, but it seems that more studies need to be designed around shoppers' tastes in this area.

Syrups

Table 9 shows the product/brand penetration of the Syrup commodity group. As before, for the sake of brevity only the top 5 products have been shown. The Private Label reigns within this commodity, and therefore we can suggest to the Private Label brand managers that there are cross-commodity marketing opportunities within this area as well as within the Pasta and Pasta Sauce groups.

Products Purchased Together

Having insight into product penetration within each commodity shows brand managers cross-commodity marketing opportunities within their respective brands; however, are there other opportunities such as cross-product opportunities? What are the products/brands that are most commonly purchased together within two complementary categories? Let's investigate.

Pasta and Pasta Sauce

Table 10 highlights the top product brands commonly purchased together from the complementary categories of pasta and pasta sauce. The top cumulative

Table 8. Pancake Mixes Penetration.

Pancake Mixes	Distinct Households	Percent of Total Population
Aunt Jemima	56,103	42.96%
Hungry Jack	35,557	27.23%
Private Label	32,199	24.66%
Krusteaz	9,241	7.08%
White Lily	7,161	5.48%

Table 9. Syrup Penetration.

Syrups	Distinct Households	Percent of Total Population
Private Label	106,181	41.44%
Aunt Jemima	65,654	25.62%
Mrs Butterworth	31,809	12.41%
Karo	30,713	11.99%
Northwoods	29,456	11.50%

Table 10. Products Commonly Purchased Together [Pasta and Pasta Sauce].

Pasta Brand	Pasta Sauce Brand	%	Cumulative %
Private Label	Ragu	18.37%	18.37%
Private Label	Private Label	8.78%	27.15%
Private Label	Prego	8.09%	35.24%
Private Label	Hunt's	5.83%	41.07%
Barilla	Ragu	4.69%	45.76%
Mueller	Ragu	4.34%	50.10%
Creamette	Ragu	3.95%	54.05%
Ronzoni	Ragu	2.80%	56.86%
Private Label	Classico	2.65%	59.51%
Barilla	Prego	2.51%	62.03%
Private Label Premium	Ragu	2.27%	64.30%
Mueller	Prego	2.01%	66.30%
Barilla	Classico	1.92%	68.23%
Creamette	Prego	1.83%	70.05%

70% of the data is shown. Not surprisingly, the first couple of top spots within the pasta side reside with the Private Label brand. Almost 20% of the market share along with the Private Label pasta brand belongs to the Ragu pasta sauce.

Pancake Mixes and Syrups

Table 11 highlights the top product brands commonly purchased together from the complementary categories of pancake mixes and syrups. Once again, the top cumulative 70% of the data is shown. The Aunt Jemima is the favorite of our grocery shoppers within both the complementary commodities; however, the Private Label brand has almost the same amount of market share within this realm. The third spot goes to the mixture of Aunt Jemima pancake mix with that of the Private Label syrup. Cross-product promotional opportunities abound with the due of Aunt Jemima and the Private Label brand!

Table 11. Products Commonly Purchased Together [Pancake Mixes and Syrups].

Pancake Mix Brand	*Syrup Brand*	*%*	*Cumulative %*
Aunt Jemima	Aunt Jemima	13.65%	13.65%
Private Label	Private Label	13.05%	26.70%
Aunt Jemima	Private Label	8.32%	35.02%
Hungry Jack	Private Label	5.11%	40.13%
Hungry Jack	Aunt Jemima	4.40%	44.52%
Aunt Jemima	Mrs Butterworth	4.05%	48.57%
Aunt Jemima	Northwoods	3.03%	51.60%
Hungry Jack	Hungry Jack	2.82%	54.42%
Hungry Jack	Mrs Butterworth	2.44%	56.86%
Aunt Jemima	Log Cabin	2.32%	59.19%
Aunt Jemima	Hungry Jack	1.84%	61.02%
Hungry Jack	Northwoods	1.65%	62.67%
Private Label	Aunt Jemima	1.52%	64.19%
Hungry Jack	Log Cabin	1.42%	65.62%
Private Label	Northwoods	1.41%	67.03%
Krusteaz	Private Label	1.37%	68.40%
Private Label	Private Label Value	1.01%	69.41%
Krusteaz	Aunt Jemima	0.91%	70.32%

CHAPTER IV
CARBO-LOADING: DETAILED ANALYSIS AND DISCUSSION

After this preliminary study, we now restrict our focus to considering coupon usage and brand loyalty. Throughout this section both will be highlighted along with their intricate relationship with one another as based on the Carbo-Loading Source file from dunnhumbyUSA. Some questions will be answered, new ones will be posed. And the answer to whether a coupon brings about brand loyalty will be sought.

Coupon Use

There are a total of 510,027 distinct households within the Carbo-Loading data set. Out of this entire population, 42,028 are actual coupon users; meaning, that they used a coupon at one time or another during any one of their shopping trips during our 2 year tracking period. Translation: only about 8% of our population uses coupons. The marketing brand managers are reaching a small pool of customers. Table 12 shows the further breakdown of total coupon users by commodity. The percentage of coupon users varies drastically between that of our four commodities; the lowest being that of the pasta commodity group. It is interesting to note that this group has the highest market share belonging to the Private Label brand.

Out of all of our customers in our two year data file, do a certain number exist that purchased a commodity or a product for the first time with the use of a coupon? Did this result in additional purchases within that household of that commodity or product? Table 13 shows the breakdown of first time purchases via a coupon within our four commodities of Pasta, Pasta Sauce, Pancake Mix, and Syrups. Within the Pancake Mix and Syrup complementary commodities, the

Table 12. Coupon Usage

	Population	Coupon Users	% Coupon Users
Total	510,027	42,028	8.24%
Pasta	411,601	11,875	2.89%
Pasta Sauce	358,600	24,983	6.97%
Pancake Mix	130,580	4,067	3.11%
Syrup	256,250	13,935	5.44%

Table 13. First Coupon Usage

	Population	Coupon Users	% Coupon Users	First Coupon Usage	% out of Coupon Users
Pasta	411,601	11,875	2.89%	1,611	13.57%
Pasta Sauce	358,600	24,983	6.97%	4,901	19.62%
Pancake Mix	130,580	4,067	3.11%	920	22.62%
Syrup	256,250	13,935	5.44%	3,181	22.83%

percentage of first time coupon purchasers holds steady at about 23%. However, within our Pasta and Pasta Sauce complementary commodities we have significant variation. Pasta (the commodity reigned by the Private Label brand) has the least amount of first time coupon users, as expected.

Within the four commodities, the entire first time coupon user population made subsequent purchases of the product, with or without the use of the coupon.

Now let's take a look at specific brands within our commodities and the first time coupon purchaser data. Table 14 summarizes all brands by their particular commodity as ordered by the number of distinct households within our population that have first purchased the brand using a coupon. Every single one of our first purchase coupon users has made subsequent purchases of that brand, within that commodity.

Pasta Brands

Over half of all of our first purchase coupon users, within the Pasta commodity, bought the Barilla brand using a coupon first. Barilla enjoys the greatest market share within this category of coupon usage, followed not so closely by Ronzoni and San Giorgio and Mueller brands.

Pasta Sauce Brands

Over half of all of our first purchase coupon users, within the Pasta Sauce commodity, bought the Ragu brand using a coupon first. Ragu enjoys the greatest market share within this category of coupon usage, followed not so closely by Prego and Bertolli brands. Barilla should start cross-marketing here.

Pancake Mix Brands

Only about a third of all of our first purchase coupon users, within the Pancake Mix commodity, bought the Hungry Jack brand using a coupon first. Hungry

Table 14. First Coupon Usage Among Brands

Commodity	Brand	First Coupon Users	% out of First Commodity Coupon Usage
	Barilla	911	56.55%
	Ronzoni	439	27.25%
	San Giorgio	204	12.66%
	Mueller	197	12.23%
	No Yolks	133	8.26%
	Creamette	80	4.97%
	Private Label	26	1.61%
	Private Label Premium	26	1.61%
	Healthy Harvest	23	1.43%
	DaVinci	7	0.43%
	Barilla Plus	5	0.31%
	Hodgson Mills	2	0.12%
Pasta	Colavita	1	0.06%
	Ragu	2,958	60.36%
	Prego	934	19.06%
	Bertolli	692	14.12%
	Classico	200	4.08%
	Newman's	115	2.35%
	Private Label Premium	94	1.92%
	Hunt's	63	1.29%
	Emeril's	33	0.67%
	Barilla	28	0.57%
	Private Label	17	0.35%
Pasta Sauce	Chef Pizza	1	0.02%
	Hungry Jack	316	34.35%
	Aunt Jemima	127	13.80%
	Pioneer	47	5.11%
	Bisquick	12	1.30%
	White Lily	11	1.20%
	Krusteaz	3	0.33%
	Hodgson Mills	1	0.11%
Pancake Mixes	M W Flapstax	1	0.11%
	Northwoods	618	19.43%
	Aunt Jemima	476	14.96%
	Log Cabin	258	8.11%
	Hungry Jack	251	7.89%
	Kellogg	100	3.14%
	Alaga	38	1.19%
	Cozy Cottage	11	0.35%
	Cary's	10	0.31%
	Spring Tree	9	0.28%
	Private Label	7	0.22%
	Grandma Molases	2	0.06%
	Private Label Value	2	0.06%
	Smuckers	2	0.06%
	Maple Grove	1	0.03%
Syrups	Private Label Premium	1	0.03%

Jack enjoys the greatest market share within this category of coupon usage, followed not so closely by the Aunt Jemima brand; the others do not represent a significant portion of the market.

Syrup Brands

Only about a fifth of all of our first purchase coupon users, within the Syrup commodity, bought the Northwoods brand using a coupon first. Hungry Jack falling to a distant fourth within this first coupon use syrup commodity group may look at cross-marketing opportunities since it enjoys such a predominant placement within the Pancake Mix category.

To Be Loyal Or Not Be Loyal

Before the question of whether coupon marketing leads to brand loyalty can be addressed, we must first ask what it is to be brand loyal. There can be varying degrees of brand loyalty; completely brand loyal, somewhat brand loyal, not so brand loyal, and not at all brand loyal. We can also apply a Brand Loyalty Index (BLI) in which we measure the specific percentage of the brand that is bought by a distinct household; for example, if the Barilla brand is purchased 4 times within one household and the Prego brand is bought once in that same dwelling, the Barilla brand would receive an index of 0.8 while the Prego brand would receive a BLI of 0.2, within that specific household.

To calculate an overall BLI per brand, all of the households' BLI values would be averaged for each specific brand, within a specific commodity. The higher the index, the better the brand loyalty to that specific brand.

Table 15 shows a breakdown of complete brand loyalty of households by commodity. Complete brand loyalty is classified as the purchase of one brand

Table 15. Complete Brand Loyalty by Commodity

Commodity	Brand Loyal Customers	Total Commodity Shoppers	%
Pasta	59,781	411,601	14.52%
Pasta Sauce	83,937	358,600	23.41%
Pancake Mix	24,372	130,580	18.66%
Syrups	40,171	256,250	15.68%

only throughout the entire two year tracking period. Furthermore, the household must have bought the brand within the commodity class at least twice.

The commodity with the highest percentage of brand loyal customers is the Pasta Sauce commodity. This finding follows along with our findings of the brand penetration study; the lowest brand loyalty lies within the Pasta commodity, in which the largest brand market share belongs to the Private Label brand.

The next four sections will give detailed data for the Brand Loyalty Index (BLI) for each brand within a commodity. Within these sections we will have available the statistics of how many distinct households within our data set actually bought the product; this was an imperative add as the brands with a high index may have had a very low penetration number. The coupled information will give us a great overall view of how the brands measure up against one another.

Pasta

Table 16 gives the Brand Loyalty Index (BLI) for the top index brands within the Pasta commodity group. Our top index pasta brand is La Russa; however, only one distinct household purchased this brand within our data set! It is not a significant finding and should be ignored. The highlighted brands are ones with

Table 16. Pasta BLI by Brand.

Brand	BLI	Distinct Households Purchased	% of Total Population
La Russa	1.000	1	0.00%
Private Label	0.664	267,358	64.96%
Edd Og	0.583	1	0.00%
La Moderna	0.577	97	0.02%
Bionature	0.500	1	0.00%
Mueller	0.487	91,128	22.14%
Mlinotst	0.448	3	0.00%
Creamette	0.442	86,672	21.06%
Barilla	0.414	125,579	30.51%
Ronzoni	0.386	68,901	16.74%
Private Label Premium	0.379	76,558	18.60%
Hodgson Mills	0.327	18,936	4.60%
Castelna	0.324	5	0.00%
Kraft	0.320	1,557	0.38%
Eddie	0.315	246	0.06%
Dreamfield	0.310	6,456	1.57%
Private Label Value	0.309	20,801	5.05%
Darielle	0.306	21	0.01%
San Giorgio	0.300	42,559	10.34%
No Yolks	0.291	28,057	6.82%
Raos	0.286	37	0.01%
Dececco	0.274	3,111	0.76%
Healthy Harvest	0.274	8,054	1.96%
Mother's	0.273	100	0.02%
China Mandarin	0.265	7	0.00%
Colavita	0.265	507	0.12%
R&F	0.264	678	0.16%
DaVinci	0.264	5,158	1.25%
Sugar Buster	0.257	90	0.02%
Vita	0.250	1	0.00%
Pennsylvania Dutch	0.243	1,692	0.41%
Barilla Plus	0.243	6,849	1.66%
Defino	0.240	408	0.10%
Alessi	0.239	312	0.08%
Al Dente	0.230	863	0.21%
Notta	0.230	271	0.07%
Mrs Weiss	0.218	427	0.10%
Orzo	0.217	83	0.02%
Annie Chns	0.211	3	0.00%

significant household penetration; Private Label being the top one, and coming in with a BLI of 0.664. The interesting find is that Barilla, which currently maintains 30% of the household penetration, has a rather low BLI index. In other words, one third of our households chose Barilla but are not very loyal to the brand. Do coupons play a role here?

Pasta Sauce

Table 17 gives the BLI for the top index brands within the Pasta Sauce commodity group. Once again, the top index brands have very low penetration values and should be ignored. Ragu, Prego, Hunt's, and Private Label which are all high penetration brands all have high index values. It looks like our households choose to be brand loyal within this commodity more so than the Pasta commodity group.

Pancake Mix

Table 18 gives the BLI for the top index brands within the Pancake Mix commodity group. Our high penetration brands also have very strong BLI values – much more so than even the Pasta Sauce commodity group!

Syrup

Table 19 gives the BLI for the top index brands within the Syrup commodity group. The high penetration brands have pretty strong BLI values; not as strong as that of the Pancake Mix commodity group but very much comparable.

In addition to the BLI values, we can look within our commodity groups to test whether coupon usage does relate to brand loyalty.

Are Coupon Users Brand Loyal?

Brand managers use coupons to spur excitement. To propel the grocery shopper

Table 17. Pasta Sauce BLI by Brand.

Brand	BLI	Distinct Households Purchased	% of Total Population
Pomi	1.000	2	0.00%
Ferrara	0.750	3	0.00%
Enrico	0.714	2	0.00%
Ragu	0.684	207,237	57.79%
Prego	0.562	118,132	32.94%
Hunt's	0.554	69,509	19.38%
Private Label	0.531	90,754	25.31%
Dave's	0.530	15	0.00%
Joey's	0.500	1	0.00%
Classico	0.461	58,865	16.42%
Newman's	0.429	19,330	5.39%
Silver Palate	0.419	16	0.00%
San Marzano	0.412	52	0.01%
Bertolli	0.404	40,187	11.21%
Mayacmas	0.400	5	0.00%
Chef Pizza	0.377	1,596	0.45%
Raos	0.373	166	0.05%
Alessi	0.358	106	0.03%
Colavita	0.357	316	0.09%
Annarino	0.355	10	0.00%
Patsy's	0.354	384	0.11%
Emeril's	0.351	9,564	2.67%
Roselli	0.344	148	0.04%
RR	0.333	1	0.00%
Buitoni	0.333	463	0.13%
Cento	0.326	60	0.02%
Candoni	0.324	1,094	0.31%
Mom's	0.323	185	0.05%
Dell Amore	0.312	117	0.03%
Barilla	0.303	6,202	1.73%
Bellino	0.302	36	0.01%
Private Label Premium	0.295	16,894	4.71%
Brother's	0.266	59	0.02%

Table 18. Pancake Mix BLI by Brand.

Brand	BLI	Distinct Households Purchased	% of Total Population
Osem Bissli	0.876	53	0.04%
Aunt Jemima	0.834	56,103	42.96%
Lund Swede	0.833	2	0.00%
Bisquick	0.817	7,021	5.38%
Private Label	0.813	32,199	24.66%
White Lily	0.812	7,161	5.48%
M W Flapstax	0.794	1,115	0.85%
Hungry Jack	0.755	35,557	27.23%
Bruce's	0.725	498	0.38%
Pioneer	0.721	2,027	1.55%
Fastshake	0.721	24	0.02%
Hodgson Mills	0.719	4,287	3.28%
Krusteaz	0.717	9,241	7.08%
Maple Grove	0.708	2,000	1.53%
Classique	0.695	356	0.27%
Private Label Premium	0.681	2,731	2.09%
Mrs Butterworth	0.621	5,164	3.95%

Table 19. Syrup BLI by Brand.

Brand	BLI	Distinct Households Purchased	% of Total Population
Braswell	1.000	1	0.00%
Lyles	0.833	2	0.00%
Private Label	0.755	106,181	41.44%
Vermont Gold	0.750	2	0.00%
Aunt Jemima	0.689	65,654	25.62%
DaVinci	0.667	2	0.00%
Tree of Life	0.667	3	0.00%
Northwoods	0.640	29,456	11.50%
Mrs Butterworth	0.629	31,809	12.41%
Hungry Jack	0.627	18,766	7.32%
Howard's	0.617	11	0.00%
Alaga	0.603	2,135	0.83%
Private Label Premium	0.596	14,060	5.49%
Karo	0.596	30,713	11.99%
Log Cabin	0.589	23,590	9.21%

toward their specific brands and hope that once the grocery shopper tries their product, none other will match their needs, within that specific commodity. However, has the theory that coupon usage leads to brand loyalty ever been carefully investigated? To seek the answer to this question we can look at a Chi Square independence test between the two categories (coupon usage and brand loyalty) and ask whether there truly is an association between the two populations. If the null hypothesis of no association is not rejected, it would suggest that there is no conclusive link between the two. However, if the null hypothesis of no association is rejected, we may conclude that coupon users and brand loyal customers are associated with one another. In this case, we will look at the odds ratio measurement in order to determine which direction the association is the strongest.

The independence test designed to test the statistically significant association between coupon users and brand loyal customers is applied to a tabulated version of our data. A coupon user is defined as a household which has used a coupon within the specific commodity. A brand loyal customer is defined to have only purchased one brand within the specific commodity, and at least purchased this brand twice within our two year tracking period.

The following are the results of the independence tests per each commodity group.

Pasta

Figure 3 shows the results of the independence test for our Pasta commodity group. Since our p-value is so small, the null hypothesis is rejected and we can definitively conclude that there is a statistically significant association between a coupon user and a brand loyal customer. The direction of this significant association is determined by the Odds Ratio value; since this value is less than

coupon_user / brand_loyal

Frequency Percent Row Pct Col Pct	0	1	Total
0	340542 82.74 85.19 96.79	59184 14.38 14.81 99.00	399726 97.11
1	11278 2.74 94.97 3.21	597 0.15 5.03 1.00	11875 2.89
Total	351820 85.48	59781 14.52	411601 100.00

Statistics for Table 2 of coupon_user by brand_loyal
Controlling for commodity=pasta

Statistic	DF	Value	Prob
Chi-Square	1	888.2957	<.0001
Likelihood Ratio Chi-Square	1	1135.1184	<.0001
Continuity Adj. Chi-Square	1	887.5081	<.0001
Mantel-Haenszel Chi-Square	1	888.2935	<.0001

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	0.3046	0.2804	0.3309

Figure 3. Chi Square Independence Test: Pasta Group.

the value of 1 (which would imply equal odds), the odds of being a non-brand loyal customer are much greater for a coupon user than a non-coupon user. In other words, the odds of a brand loyal customer are much greater for a non-coupon user. In our Pasta commodity group, this suggests that coupon usage will not lead to brand loyalty although further studies are required to confirm this.

Pasta Sauce

Figure 4 shows the results of the independence test for our Pasta Sauce commodity group. Since our p-value is small again, the null hypothesis is rejected and we can definitively conclude that there is a statistically significant association between a coupon user and a brand loyal customer. The direction of this significant association is determined by the Odds Ratio value; since this value is again less than the value of 1 (which would imply equal odds), the odds of being a non-brand loyal customer are much greater for a coupon user than a non-coupon user. In other words, the odds of a brand loyal customer are much greater for a non-coupon user. In our Pasta Sauce commodity group, this suggests that coupon usage will not lead to brand loyalty.

Table 3 of coupon_user by brand_loyal
Controlling for commodity=pasta sauce

coupon_user (Frequency / Percent / Row Pct / Col Pct)	brand_loyal 0	brand_loyal 1	Total
0	254286 70.91 76.22 92.58	79331 22.12 23.78 94.51	333617 93.03
1	20377 5.68 81.56 7.42	4606 1.28 18.44 5.49	24983 6.97
Total	274663 76.59	83937 23.41	358600 100.00

Statistics for Table 3 of coupon_user by brand_loyal
Controlling for commodity=pasta sauce

Statistic	DF	Value	Prob
Chi-Square	1	370.0351	<.0001
Likelihood Ratio Chi-Square	1	389.1921	<.0001
Continuity Adj. Chi-Square	1	369.7372	<.0001
Mantel-Haenszel Chi-Square	1	370.0341	<.0001

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	0.7245	0.7011	0.7488

Figure 4. Chi Square Independence Test: Pasta Sauce Group.

Pancake Mix

Figure 5 shows the results of the independence test for our Pancake Mix commodity group. In this scenario, our p-value is not significant. The null hypothesis is not rejected and we cannot definitively conclude that there is a statistically significant association between a coupon user and a brand loyal customer. This is further reiterated by the value of the Odds Ratio; since its 95% confidence interval contains the value of 1, it implies equal odds for the outcome. In other words, the odds of a brand loyal customer are equally likely for a non-coupon user as they are for a coupon user. Even though these results differ from the past two commodity groups, we can again suggest from these results that coupon usage does not lead to brand loyalty within the Pancake Mix commodity group.

Syrup

Figure 6 shows the results of the independence test for our Syrups commodity group. Since our p-value is so small again, the null hypothesis is rejected and we can definitively conclude that there is a statistically significant association between a coupon user and a brand loyal customer. The direction of this significant association is determined by the Odds Ratio; since this value is again less than the value of 1 (which would imply equal odds), the odds of being a non-brand loyal customer are much greater for a coupon user than a non-

Table 1 of coupon_user by brand_loyal
Controlling for commodity=pancake mixes

coupon_user / brand_loyal

Frequency Percent Row Pct Col Pct	0	1	Total
0	102917 78.82 81.35 96.90	23596 18.07 18.65 96.82	126513 96.89
1	3291 2.52 80.92 3.10	776 0.59 19.08 3.18	4067 3.11
Total	106208 81.34	24372 18.66	130580 100.00

Statistics for Table 1 of coupon_user by brand_loyal
Controlling for commodity=pancake mixes

Statistic	DF	Value	Prob
Chi-Square	1	0.4785	0.4891
Likelihood Ratio Chi-Square	1	0.4759	0.4903
Continuity Adj. Chi-Square	1	0.4506	0.5020
Mantel-Haenszel Chi-Square	1	0.4785	0.4891

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	1.0284	0.9499	1.1135

Figure 5. Chi Square Independence Test: Pancake Mix Group.

Table 4 of coupon_user by brand_loyal
Controlling for commodity=syrups

coupon_user / brand_loyal

Frequency Percent Row Pct Col Pct	0	1	Total
0	203977 79.60 84.18 94.40	38338 14.96 15.82 95.44	242315 94.56
1	12102 4.72 86.85 5.60	1833 0.72 13.15 4.56	13935 5.44
Total	216079 84.32	40171 15.68	256250 100.00

Statistics for Table 4 of coupon_user by brand_loyal
Controlling for commodity=syrups

Statistic	DF	Value	Prob
Chi-Square	1	70.9375	<.0001
Likelihood Ratio Chi-Square	1	74.1290	<.0001
Continuity Adj. Chi-Square	1	70.7358	<.0001
Mantel-Haenszel Chi-Square	1	70.9372	<.0001

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	0.8059	0.7663	0.8474

Figure 6. Chi Square Independence Test: Syrups Group.

coupon user. In other words, the odds of being a brand loyal customer are much greater for a non-coupon user. In our Syrups commodity group, we can suggest that coupon usage will not lead to brand loyalty.

Graphical Measurement Tools

Since we have no proof that coupons bring about brand loyalty, what can brand managers focus on in order to improve their marketing efforts? Before starting to think about approaches to attract customers, brand managers may look at brand loyalty measurement tools in order to track their progress. An effective brand loyalty measurement tool can be very helpful in tracking their success during testing periods, both against prior performance and against other competing brands.

Pasta and Pasta Sauce

Figure 7 is one example of a graphical measurement tool that focuses on brand loyalty. It uses the Brand Loyalty Index (BLI) values as introduced earlier in this Thesis. If a brand is present within two complementary categories (in this case, the Pasta and Pasta Sauce commodities) it can be plotted on the represented graph. In this example, the Pasta BLI values are located on the x-axis while the Pasta Sauce BLI values are located on the y-axis. The higher the BLI value the better the success of the brand within the household preference. In our graph, the BLI values go from low to high, left to right for the Pasta group and low to high for the Pasta Sauce group. The coveted region for any brand in this graphical tool is the upper right hand quadrant; the Private Label brand is the only product that enjoys this position. All of the other brands have major room for improvement. Such improvements may include, but are not limited to, taking advantage of the cross-commodity opportunity; marketing a coupon or promotion that includes both commodity groups.

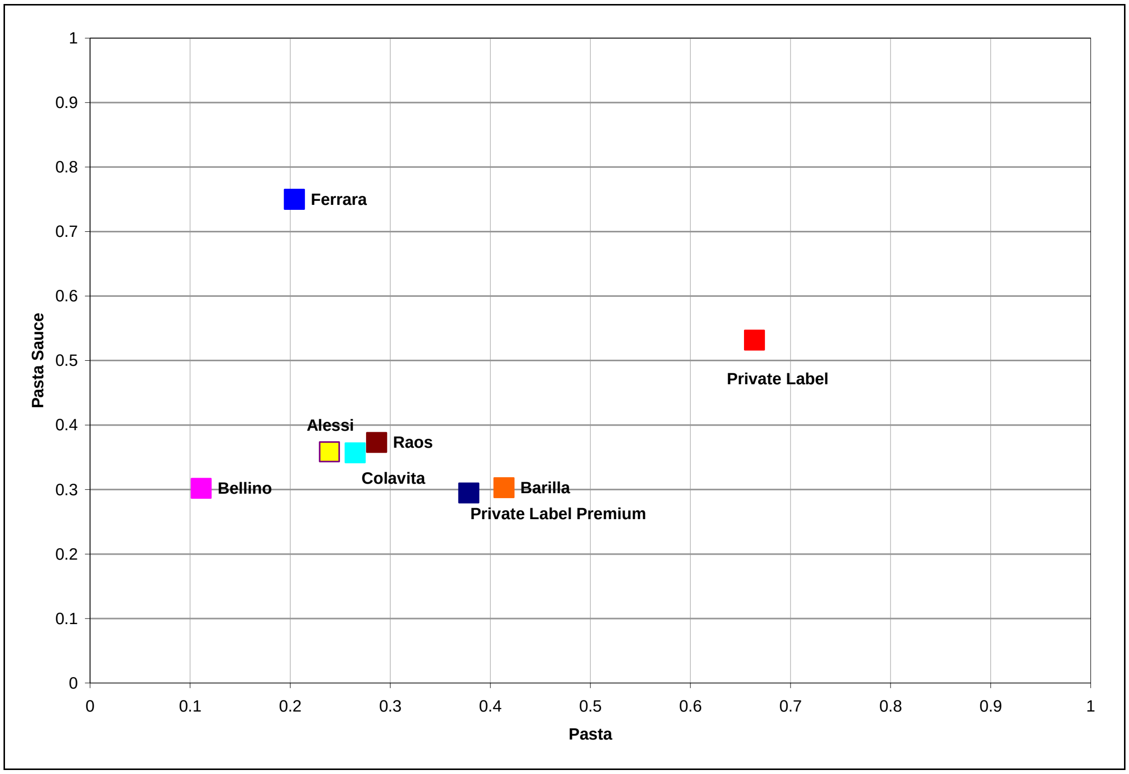

Figure 7. Graphical Brand Loyalty Tool: Pasta and Pasta Sauce.

Pancake Mix and Syrups

Figure 8 uses the same graphical brand loyalty measurement tool for the Pancake Mix and Syrup complementary categories. The Pancake Mix BLI values are located on the x-axis while the Syrup BLI values are located on the y-axis. The direction of the BLI values follow the same pattern as of that introduced in Figure 7. Again, the coveted region for any brand in this graphical tool is the upper right hand quadrant; all of the brands that cross into both commodities are enjoying this position – the brand managers are knowingly or unknowingly doing something right. Regardless, with the use of this tool they can track their progress over time, against their own track records as well as those of the competing brands.

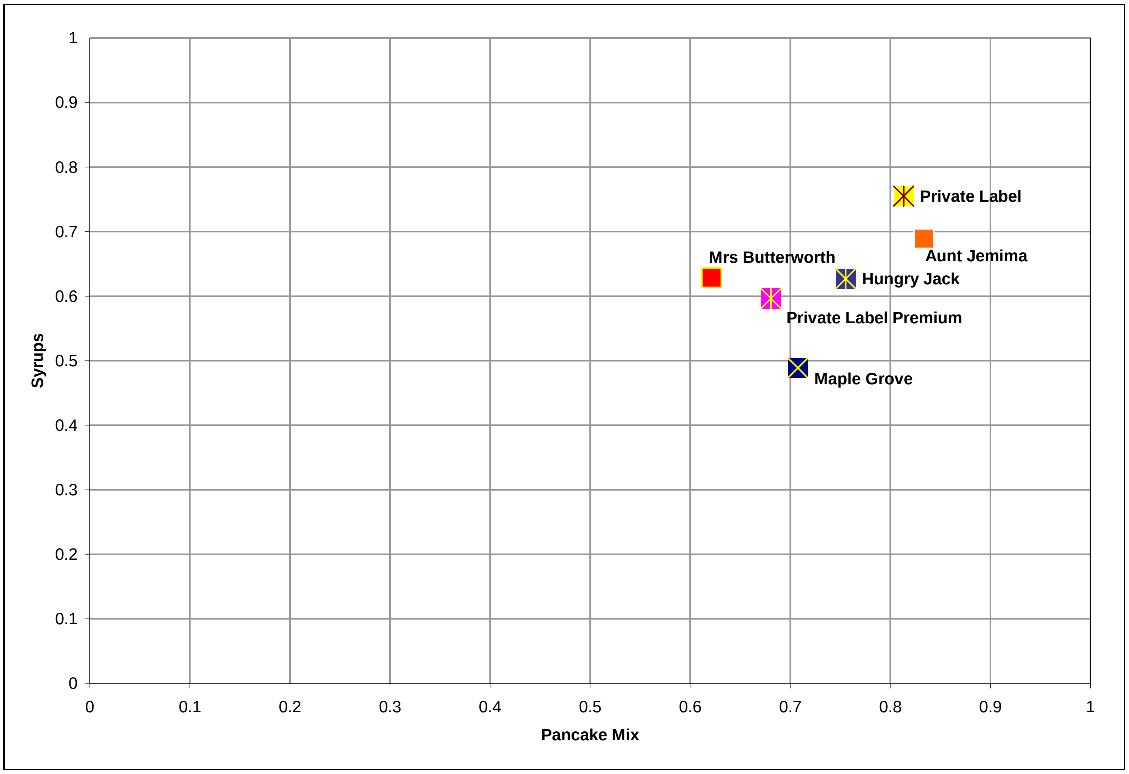

Figure 8. Graphical Brand Loyalty Tool: Pancake Mix and Syrups.

CHAPTER V
DATA SET 2: THE COMPLETE JOURNEY

Grocery store retailers have access to a sophisticated and more focused marketing tool: the direct mail marketing campaign. The direct mail marketing campaign analyzed below uses coupons to motivate shoppers to come through the door more often and to spend more, but does it in a focused manner; the coupons are tailored to the individual households, based on their historical spending habits and patterns.

Does Direct Marketing Really Work?

Within the Complete Journey data set, 2500 frequent shoppers were tracked over a two-year period. All transactions were accounted for. Out of this entire population, 1584 distinct households were chosen for the direct mail marketing programs throughout the two year tracking period. 434 of these chosen distinct households used coupons provided by the campaign at the major US grocery retailer; almost a third of the test population. This is a much higher redemption rate than that of the coupon user population within our first data set. Our goal in the analysis below is to decide whether the use of direct mail stimulated customers to visit stores more frequently and spend more.

Before and After

One way to discover whether there is a boost in results due to the direct marketing campaign is to take a look at the frequency of trips and the average spend of the chosen households, before and after their individual campaign start dates. Table 20 displays statistics for two tracking variables: the difference in the number of trips per week and the difference in the average spend per households as averaged across the entire 1584 distinct household population.

Table 20. Before vs. After Campaign Stats.

Variable	*Mean*	*Lower 95% CI*	*Upper 95% CI*
Trip Difference	0.1479	0.1031	0.1926
Spend Difference	6.9019	5.7041	8.0996

Since both variables have a positive average mean, and their specific 95% confidence intervals do not include the value of 0, we are 95% confident that there has been a positive increase in both the frequency of grocery shopping trips per week and the average spend per week after a campaign start date pertaining to a specific household.

Table 21 shows the same variable statistics but divided into two populations: the individual households that received the campaigns but opted not to use the coupons at all throughout the two-year tracking period and the households that chose to redeem the coupons at any point during the two-year tracking period. Again, both groups have a significantly positive response within both of the variables. Furthermore, the coupon-user group has a much higher mean in both the frequency of visits and average spend: this suggests that coupons specifically targeted to individuals based on their preferences and habits, that become a coupon user during the campaign period, have a higher propensity for increasing their trips and average spend at the grocery store.

A Repeated Measures Analysis

Since an increase in average spend is one of the most efficient ways to track profits, we can look at the campaign data from a slightly different perspective. Figure 9 shows a graph of the total average spend per quarter over all households that were included in the direct mail marketing program. The blue line with yellow triangles represents the total spend per quarter averaged across

Table 21. Before vs. After Campaign Stats (Coupon vs. Non-Coupon).

	Variable	*Mean*	*Lower 95% CI*	*Upper 95% CI*
Non-Coupon User	Trip Difference	0.0655	0.0139	0.1171
	Spend Difference	3.5023	2.2265	4.778
Coupon User	Trip Difference	0.3656	0.2792	0.4521
	Spend Difference	15.8865	13.2922	18.4809

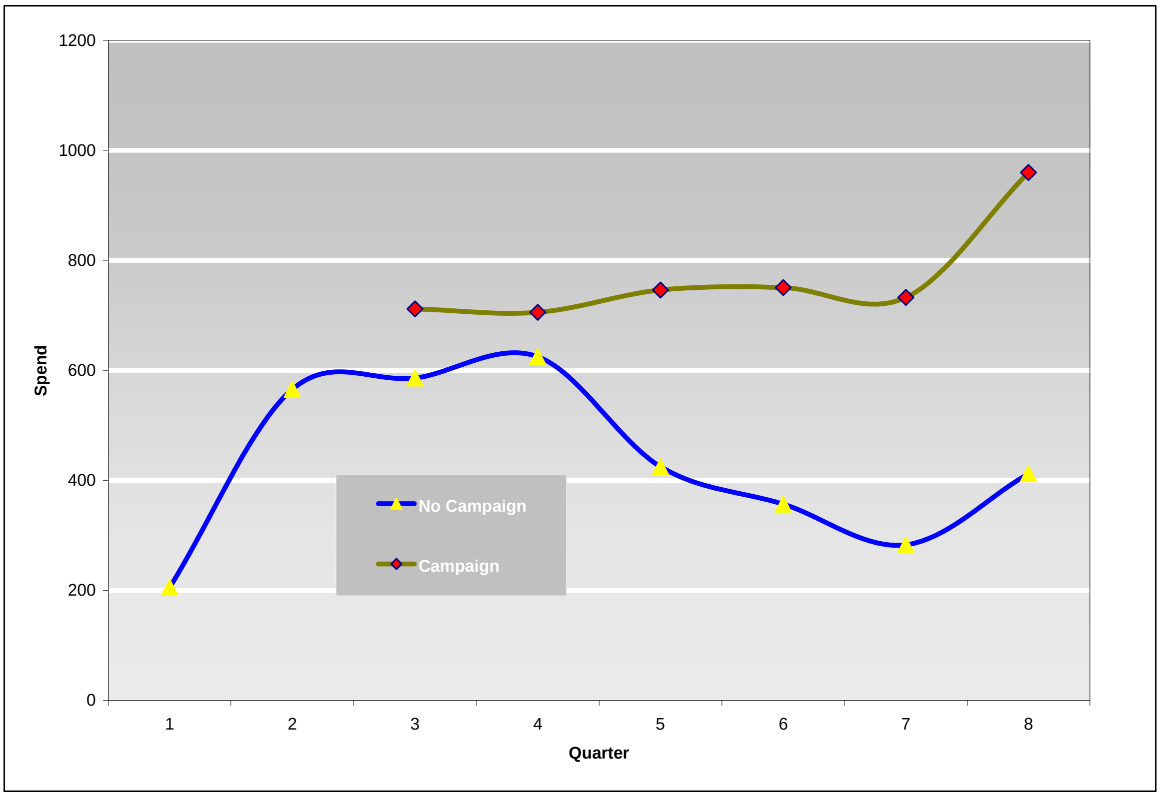

Figure 9. Total Spend by Quarter: Campaign vs. No Campaign.

all households that did not receive a campaign in that given quarter (but were candidates for direct mail campaigns in other quarters). The green line with red markers represents the total spend per quarter averaged across all households that received a campaign within the specific quarter. The campaigns did not start until the 3rd quarter in our 2-year tracking period; with the simple separation of data among the campaign/no campaign populations within a given quarter, a clear difference is displayed on this graph.

This data can be used to perform a Repeated Measures Analysis in order to study the statistical differences between the campaign/no-campaign populations as well as model their future behavior. Although our data is in the form of an observational study, it contains both fixed and repeated effects to enable us to use the mixed model design. The repeated measurement is the total spend per quarter. Our subjects are the distinct households, and the treatments are whether the household received a campaign during a particular quarter.

The questions of interest are:

1. Whether campaign mean spend and number of visits changes over time; meaning, whether a significant effect of time is present within the campaign/no-campaign populations.
2. How campaign differences change over time; meaning, whether the presence of a direct marketing campaign within a given quarter has a significant effect on grocery expenditures and/or a significant effect on the number of visits to the grocery retailer.

Translating to the statistical modeling terms, we are interested whether there is a time main effect (the quarter variable) and whether there is an interaction between the quarter and campaign variables.

However, before the actual model can be applied, an appropriate covariance structure for the repeated observations needs to be selected. The choices for our Repeated Measures Analysis are the Compound Symmetric, Autoregressive Order 1, and Unstructured. Table 22 displays the summary comparison results between these three covariance structures, using the Akaike Information Criterion (AIC) value. The reason for an appropriate covariance selection at this stage being that measurements in a repeated data set are more likely to be highly correlated at adjacent values rather than the measurements taken several time points apart. The results in Table 22 show that the Unstructured covariance matrix should be used to compile our model.

After several different modeling trials it was determined that the cubic regression model offered the best fit for our data. Figure 10 displays the results and significance of our cubic regression coefficients; all are deemed significant for our model.

Figure 11 displays the exact regression equations of our campaign/no-campaign models. The coefficients for each model are displayed. First and foremost, the signs of each of the coefficients, as compared between the two models, are completely opposite of one another which suggests that the two are indeed significantly different and behave on different levels. Furthermore, all coefficients are significant at predicting total spend for the No Campaign model; none of the coefficients, except the intercept, are significant at explaining total spend for the Campaign model. This further suggests that if a specific household is part of the campaign group, the propensity for increased spending is significant, with no other factors present. Direct marketing proves to be an effective tool, within this population.

Table 22. Covariance Structure Selection.

Covariance Structure	AIC	Effect Significance
Compound Symmetric	186965.3	All significant
Autoregressive Order 1	185766.1	All significant
Unstructured	185096.4	All significant

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
35	8295.12	<.0001

Type 1 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
campaign	1	3155	101.84	<.0001
quarter	1	3155	82.16	<.0001
quarter*campaign	1	3155	70.13	<.0001
quarter*quarter	1	3155	1109.31	<.0001
quarte*quarte*campai	1	3155	77.97	<.0001
quarte*quarte*quarte	1	3155	404.08	<.0001
quar*quar*quar*campa	1	3155	50.25	<.0001

Figure 10. Cubic Regression Effect Significance.

Solution for Fixed Effects

Effect	campaign	Estimate	Standard Error	DF	t Value	Pr > |t|
campaign	0	-293.87	20.3401	3155	-14.45	<.0001
campaign	1	957.82	220.72	3155	4.34	<.0001
quarter*campaign	0	647.14	21.1517	3155	30.60	<.0001
quarter*campaign	1	-164.73	128.31	3155	-1.28	0.1993
quarte*quarte*campai	0	-135.14	5.3557	3155	-25.23	<.0001
quarte*quarte*campai	1	32.8575	23.9549	3155	1.37	0.1703
quar*quar*quar*campa	0	8.3006	0.3885	3155	21.37	<.0001
quar*quar*quar*campa	1	-2.2678	1.4394	3155	-1.58	0.1152

Figure 11. Cubic Regression Coefficient by Model.

CHAPTER VI
CONCLUSIONS AND RECOMMENDATIONS

By and large, the data sets used in this Thesis were not analyzed exhaustively. There are an infinite number of analyses and models that can be developed from the transactions that were obtained by dunnhumbyUSA; many of which could be used to aid the retailers and brand managers of the grocery sector to further improve their operations and profit margins. Such analysis may also shed light on more general aspects of consumer behavior.

Within this Thesis, the grocery consumer's shopping patterns were investigated. Product penetration, common purchasing behavior, and coupon use were extracted from the databases, giving insight as to the preferences and purchasing styles of the customers of the large US grocery retailer.

A statistical approach suggests that coupons may not bring about brand loyalty in the average customer. As a result, brand managers may want to look beyond the sporadic distribution of coupons and make use of graphical brand loyalty measurement tools, and test new programs that delve into cross-commodity marketing campaigns.

Furthermore, it is suggested that brand managers partner with the grocery retail sector to take part in the creation and execution of designed direct marketing campaigns. The analysis and modeling of the data within this Thesis suggests that direct marketing campaigns significantly improve the frequency and spend by the average grocery store consumer.

All in all, further testing is suggested and encouraged. Randomization of the campaign treatments, full experimental design, and the non-biased application of

programs to grocery shoppers would help proved valid statistical results and conclusions. There are many longitudinal study models to choose from; many can be applied and compared based on their merit. This Thesis is just a gateway to the understanding of the grocery shopper's world.

LIST OF REFERENCES

1) www.preparedfoods.com June 2004

2) Angrisani, Carol, "Coupon Confidence; Consumers May Not be Embracing Coupons Like They Used to, but Marketers Reamin Committed to the Promotional Tool", Supermarket News, April 5, 2004

3) Tenser, James. :P&G Sets Zero-Couponing Test" Supermarket News, 46.n3, Jan 15, 1996

4) Amato-McCoy, Deena M. "Print and Save" Tech Solutions, October, 2005

5) Meyers, Peter, "Redemption Revisition" Brandweek 48.38, October 22, 2007

6) DunnhumbyUSA. Dunnhumby Source Files. USA. 2001.

APPENDIX

```
* Table 1. Household Age Range.;
select age_desc, count(*) as NumberofHouseholds, count(*)/801 as Percent
format = percent5.
from ia.hh_demographic
group by age_desc;
```

```
* Table 2. Household Marital Status.;
select marital_status_code, count(*) as NumberofHouseholds, count(*)/801 as
Percent format = percent5.
from ia.hh_demographic
group by marital_status_code;
```

```
* Table 3. Household Homeownership.;
select homeowner_desc, count(*) as NumberofHouseholds, count(*)/801 as
Percent format = percent5.
from ia.hh_demographic
group by homeowner_desc;
```

```
* Table 4. Household Composition.;
select hh_comp_desc, count(*) as NumberofHouseholds, count(*)/801 as
Percent format = percent5.
from ia.hh_demographic
group by hh_comp_desc;
```

```
* Table 5. Household Commodity Penetration.;
proc sql;

select count(distinct household)
from ia.dh_transactions;

quit;

proc sql;

select b.commodity, count(distinct a.household)
from ia.dh_transactions a, ia.dh_product_lookup b
where a.upc = b.upc
group by b.commodity;

quit;
```

```
* Table 6. Pasta Penetration.;
```

```
proc sql;

select b.brand, count(distinct a.household)
from ia.dh_transactions a, ia.dh_product_lookup b
where a.upc = b.upc and b.commodity = 'pasta'
group by b.brand
order by b.brand;

quit;
```

```
* Table 7. Pasta Sauce Penetration.;
proc sql;

select b.brand, count(distinct a.household)
from ia.dh_transactions a, ia.dh_product_lookup b
where a.upc = b.upc and b.commodity = 'pasta sauce'
group by b.brand
order by b.brand;

quit;
```

```
* Table 8. Pancake Mixes Penetration.;
proc sql;

select b.brand, count(distinct a.household)
from ia.dh_transactions a, ia.dh_product_lookup b
where a.upc = b.upc and b.commodity = 'pancake mixes'
group by b.brand
order by b.brand;

quit;
```

```
* Table 9. Syrup Penetration.;
proc sql;

select b.brand, count(distinct a.household)
from ia.dh_transactions a, ia.dh_product_lookup b
where a.upc = b.upc and b.commodity = 'syrups'
group by b.brand
order by b.brand;

quit;
```

```
* Table 10. Products Commonly Purchased Together [Pasta and Pasta Sauce].;
proc sql;

create table ia.Commodity_by_Trip as
select a.household, a.basket,
case when (b.commodity = 'pasta') then 1 else 0 end as pasta,
case when (b.commodity = 'pasta sauce') then 1 else 0 end as pasta_sauce,
case when (b.commodity = 'pancake mixes') then 1 else 0 end as
pancake_mixes,
case when (b.commodity = 'syrups') then 1 else 0 end as syrups,
b.brand, count(b.brand) as No_Purchases
from ia.dh_transactions a, ia.dh_product_lookup b
where a.upc = b.upc
group by a.household, a.basket
order by a.household, a.basket;

quit;

proc sql;

create table ia.Pasta_and_Sauce as
select household, basket, pasta, pasta_sauce, brand, no_purchases
from ia.Commodity_by_Trip
where pasta = 1 or pasta_sauce = 1;

quit;

proc sql;

create table ia.Pasta_and_Sauce_1 as
select distinct basket
from ia.Pasta_and_Sauce
group by basket
having sum(pasta) > 0 and sum(pasta_sauce) > 0;

quit;

proc sql;

create table ia.Pasta_and_Sauce_2 as
select *
from ia.Pasta_and_Sauce
```

```
where basket in (select distinct basket from ia.Pasta_and_Sauce_1);

quit;

proc sql;
create table ia.Pasta_3 as
select basket, brand
from ia.Pasta_and_Sauce_2
where pasta = 1
order by basket;

quit;

proc sql;
create table ia.Sauce_4 as
select basket, brand
from ia.Pasta_and_Sauce_2
where pasta_sauce = 1
order by basket;

quit;

proc sql;

create table ia.pasta_and_sauce_34 as
select a.basket, b.basket, a.brand as pasta_brand, b.brand as sauce_brand
from ia.Pasta_3 a, ia.Sauce_4 b
where a.basket = b.basket
order by a.basket, b.basket, a.brand, b.brand;

quit;

proc sql;

create table ia.pasta_and_sauce_final as
select pasta_brand, sauce_brand, count(basket)
from ia.pasta_and_sauce_34
group by pasta_brand, sauce_brand;

quit;
```

```
* Table 11. Products Commonly Purchased Together [Pancake Mixes and
Syrups].;
```

```
proc sql;

create table ia.Commodity_by_Trip as
select a.household, a.basket,
case when (b.commodity = 'pasta') then 1 else 0 end as pasta,
case when (b.commodity = 'pasta sauce') then 1 else 0 end as pasta_sauce,
case when (b.commodity = 'pancake mixes') then 1 else 0 end as
pancake_mixes,
case when (b.commodity = 'syrups') then 1 else 0 end as syrups,
b.brand, count(b.brand) as No_Purchases
from ia.dh_transactions a, ia.dh_product_lookup b
where a.upc = b.upc
group by a.household, a.basket
order by a.household, a.basket;

quit;

proc sql;

create table ia.Pancake_and_Syrup as
select household, basket, pancake_mixes, syrups, brand, no_purchases
from ia.Commodity_by_Trip
where pancake_mixes = 1 or syrups = 1;

quit;

proc sql;

create table ia.Pancake_and_Syrup_1 as
select distinct basket
from ia.Pancake_and_Syrup
group by basket
having sum(pancake_mixes) > 0 and sum(syrups) > 0;

quit;

proc sql;

create table ia.Pancake_and_Syrup_2 as
select *
from ia.Pancake_and_Syrup
where basket in (select distinct basket from ia.Pancake_and_Syrup_1);
```

```
quit;

proc sql;
create table ia.Pancake_3 as
select basket, brand
from ia.Pancake_and_Syrup_2
where pancake_mixes = 1
order by basket;

quit;

proc sql;
create table ia.Syrup_4 as
select basket, brand
from ia.Pancake_and_Syrup_2
where syrups = 1
order by basket;

quit;

proc sql;

create table ia.Pancake_and_Syrup_34 as
select a.basket, b.basket, a.brand as pancake_brand, b.brand as syrup_brand
from ia.Pancake_3 a, ia.Syrup_4 b
where a.basket = b.basket
order by a.basket, b.basket, a.brand, b.brand;

quit;

proc sql;

create table ia.Pancake_and_Syrup_final as
select pancake_brand, syrup_brand, count(basket)
from ia.Pancake_and_Syrup_34
group by pancake_brand, syrup_brand;

quit;
```

```
* Table 12. Coupon Usage.;
* Query for calculating number of total coupon users in data set;
proc sql;
```

```
create table ia.couponusers as
select household
from ia.dh_transactions
group by household
having sum(coupon) > 0;

quit;

proc sql;
select count(distinct household) from ia.couponusers;
quit;

/* 42,028 coupon users out of a total of 510,027 - roughly 8% of our population */

* Query for calculating number of total coupon users per commodity;
proc sql;

create table ia.couponusers as
select b.commodity, a.household
from ia.dh_transactions a, ia.dh_product_lookup b
where a.upc = b.upc
group by b.commodity, a.household
having sum(coupon) > 0;

quit;

proc sql;
select commodity, count(distinct household) from ia.couponusers
group by commodity;
quit;
```

```
* Table 13. First Coupon Usage.;
* Query for first pasta commodity purchased via coupon;
proc sql;

create table ia.coupon_used as
select commodity, household, day
from ia.dh_transactions a, ia.dh_product_lookup b
where a.upc = b.upc and commodity = 'pasta' and coupon = 1;

quit;

proc sql;

create table ia.coupon_used_1 as
```

```
select commodity, household, day
from ia.dh_transactions a, ia.dh_product_lookup b
where a.upc = b.upc and commodity = 'pasta' and coupon = 0;

quit;

proc sql;

create table ia.coupon_used_2 as
select a.commodity, a.household
from ia.coupon_used a, ia.coupon_used_1 b
where a.household = b.household
group by a.commodity, a.household, a.day
having min(a.day) < min(b.day);

quit;

proc sql;

select count(distinct household)
from ia.coupon_used_2;

quit;

* Query to calculate number of households that had subsequent purchases;
* of that commodity without a coupon;

proc sql;

select count(distinct a.household)
from ia.dh_transactions a, ia.coupon_used_2 b, ia.dh_product_lookup c
where a.household = b.household and a.upc = c.upc and b.commodity =
c.commodity
and a.coupon = 0;

quit;


* Query for first pasta sauce commodity purchased via coupon;
proc sql;

create table ia.coupon_used as
select commodity, household, day
```

```
from ia.dh_transactions a, ia.dh_product_lookup b
where a.upc = b.upc and commodity = 'pasta sauce' and coupon = 1;

quit;

proc sql;

create table ia.coupon_used_1 as
select commodity, household, day
from ia.dh_transactions a, ia.dh_product_lookup b
where a.upc = b.upc and commodity = 'pasta sauce' and coupon = 0;

quit;

proc sql;

create table ia.coupon_used_2 as
select a.commodity, a.household
from ia.coupon_used a, ia.coupon_used_1 b
where a.household = b.household
group by a.commodity, a.household, a.day
having min(a.day) < min(b.day);

quit;

proc sql;

select count(distinct household)
from ia.coupon_used_2;

quit;

* Query to calculate number of households that had subsequent purchases;
* of that commodity without a coupon;

proc sql;

select count(distinct a.household)
from ia.dh_transactions a, ia.coupon_used_2 b, ia.dh_product_lookup c
where a.household = b.household and a.upc = c.upc and b.commodity =
c.commodity
and a.coupon = 0;
```

```
quit;


* Query for first pancake mixes commodity purchased via coupon;
proc sql;

create table ia.coupon_used as
select commodity, household, day
from ia.dh_transactions a, ia.dh_product_lookup b
where a.upc = b.upc and commodity = 'pancake mixes' and coupon = 1;

quit;

proc sql;

create table ia.coupon_used_1 as
select commodity, household, day
from ia.dh_transactions a, ia.dh_product_lookup b
where a.upc = b.upc and commodity = 'pancake mixes' and coupon = 0;

quit;

proc sql;

create table ia.coupon_used_2 as
select a.commodity, a.household
from ia.coupon_used a, ia.coupon_used_1 b
where a.household = b.household
group by a.commodity, a.household, a.day
having min(a.day) < min(b.day);

quit;

proc sql;

select count(distinct household)
from ia.coupon_used_2;

quit;

* Query to calculate number of households that had subsequent purchases;
* of that commodity without a coupon;
```

```
proc sql;

select count(distinct a.household)
from ia.dh_transactions a, ia.coupon_used_2 b, ia.dh_product_lookup c
where a.household = b.household and a.upc = c.upc and b.commodity =
c.commodity
and a.coupon = 0;

quit;


* Query for first syrups commodity purchased via coupon;
proc sql;

create table ia.coupon_used as
select commodity, household, day
from ia.dh_transactions a, ia.dh_product_lookup b
where a.upc = b.upc and commodity = 'syrups' and coupon = 1;

quit;

proc sql;

create table ia.coupon_used_1 as
select commodity, household, day
from ia.dh_transactions a, ia.dh_product_lookup b
where a.upc = b.upc and commodity = 'syrups' and coupon = 0;

quit;

proc sql;

create table ia.coupon_used_2 as
select a.commodity, a.household
from ia.coupon_used a, ia.coupon_used_1 b
where a.household = b.household
group by a.commodity, a.household, a.day
having min(a.day) < min(b.day);

quit;

proc sql;
```

```
select count(distinct household)
from ia.coupon_used_2;

quit;

* Query to calculate number of households that had subsequent purchases;
* of that commodity without a coupon;

proc sql;

select count(distinct a.household)
from ia.dh_transactions a, ia.coupon_used_2 b, ia.dh_product_lookup c
where a.household = b.household and a.upc = c.upc and b.commodity =
c.commodity
and a.coupon = 0;

quit;
```

```
* Table 14. First Coupon Usage Among Brands.;
proc sql;

create table ia.coupon_used as
select a.upc, a.household, a.day, b.commodity, b.brand
from ia.dh_transactions a, ia.dh_product_lookup b
where a.upc = b.upc and a.coupon = 1;

quit;

proc sql;

create table ia.coupon_used_1 as
select a.upc, a.household, a.day, b.commodity, b.brand
from ia.dh_transactions a, ia.dh_product_lookup b
where a.upc = b.upc and a.coupon = 0;

quit;

proc sql;

create table ia.coupon_used_2 as
select a.commodity, a.brand, a.household
from ia.coupon_used a, ia.coupon_used_1 b
where a.upc = b.upc and a.household = b.household
group by a.upc, a.household, a.day
```

```
having min(a.day) < min(b.day);

quit;

proc sql;

create table ia.coupon_used_3 as
select commodity, brand, count(distinct household)
from ia.coupon_used_2
group by commodity, brand;

quit;

* Query to calculate number of households that had subsequent purchases;
* of that commodity without a coupon;
proc sql;

create table ia.coupon_subs as
select b.commodity, b.brand, count(distinct a.household)
from ia.dh_transactions a, ia.coupon_used_2 b, ia.dh_product_lookup c
where a.household = b.household and a.upc = c.upc and b.commodity =
c.commodity
and a.coupon = 0
group by b.commodity, b.brand;

quit;
```

```
* Table 15. Complete Brand Loyalty by Commodity.;
proc sql;

create table ia.bu_cu as
select a.household, b.commodity,
case when (count(distinct b.brand) = 1 and count(b.brand) > 1)
		then 1
	else 0
end as brand_loyal,
case when (sum(a.coupon) > 0)
		then 1
	else 0
end as coupon_user
from ia.dh_transactions a, ia.dh_product_lookup b
where a.upc = b.upc
group by a.household, b.commodity
order by a.household, b.commodity;
```

```
quit;

proc sql;

select commodity, count(distinct household)
from ia.bu_cu
where brand_loyal = 1
group by commodity;

quit;
```

```
* Table 16. Pasta BLI by Brand.;
* Table 17. Pasta Sauce BLI by Brand.;
* Table 18. Pancake Mix BLI by Brand.;
* Table 19. Syrup BLI by Brand.;
proc sql;

create table ia.blindex as
select a.household, b.commodity, b.brand, sum(a.units) as quantity
from ia.dh_transactions a, ia.dh_product_lookup b
where a.upc = b.upc
group by a.household, b.commodity, b.brand;

quit;

proc sql;

create table ia.blindex_1 as
select a.household, b.commodity, sum(a.units) as total_quantity
from ia.dh_transactions a, ia.dh_product_lookup b
where a.upc = b.upc
group by a.household, b.commodity;

quit;

proc sql;
create table ia.blindex_2 as
select a.household, a.commodity, a.brand, a.quantity/b.total_quantity as bli
from ia.blindex a, ia.blindex_1 b
where a.household = b.household and a.commodity = b.commodity
group by a.household, a.commodity, a.brand;

quit;
```

```
proc sql;
create table ia.blindex_3 as
select commodity, brand, avg(bli) as bli
from ia.blindex_2
group by commodity, brand;

quit;
```

```
* independence test: using previously created table ia.bu_cu ;
* Figure 3. Chi Square Independence Test: Pasta Commodity.;
* Figure 4. Chi Square Independence Test: Pasta Sauce Commodity.;
* Figure 5. Chi Square Independence Test: Pancake Mix Commodity.;
* Figure 6. Chi Square Independence Test: Syrup Commodity.;

proc freq data = ia.bu_cu;
tables commodity*coupon_user*brand_loyal/ chisq riskdiff cmh; run;
```

```
* Table 20. Before vs After Campaign Stats. ;
* Table 21. Before vs After Campaign Stats (Coupon vs Non-Coupon). ;

* all campaigns by start date ;
proc sql;

create table ia.a as
select a.household_key, a.campaign, b.start_day
from ia.campaign_table a, ia.campaign_desc b
where a.campaign = b.campaign
group by a.household_key, a.campaign, b.start_day
order by a.household_key;

quit;

* distinct households with the earliest week of campaign start ;
proc sql;

create table ia.aa as
select distinct household_key, min(start_day) as frst_start_day,
round((min(start_day)/7),1) as frst_start_week, count(distinct campaign) as
campaigns
from ia.a
group by household_key;

quit;
```

```
* 1584 households mailed a campaign ;
* these 1584 households received 30 different campaigns - varied between
households ;

* statistics before campaign;
proc sql;

create table ia.b_1 as
select distinct a.household_key, count(distinct a.basket_id) as total_trips,
sum(a.sales_value) as total_spend
from ia.transaction_data a, ia.aa b
where a.household_key = b.household_key and a.day < b.frst_start_day
group by b.household_key;

quit;

proc sql;

create table ia.b_2 as
select distinct a.household_key, a.total_trips/b.frst_start_week as avg_trips,
a.total_spend/b.frst_start_week as avg_spend
from ia.b_1 a, ia.aa b
where a.household_key = b.household_key
group by a.household_key
order by a.household_key;

quit;

* statistics after campaign;
proc sql;

create table ia.c_1 as
select distinct a.household_key, count(distinct a.basket_id) as total_trips,
sum(a.sales_value) as total_spend, b.campaigns, b.frst_start_week
from ia.transaction_data a, ia.aa b
where a.household_key = b.household_key and a.day >= b.frst_start_day
group by b.household_key;

quit;

* 1581 households have data after campaign dates ;
```

```
proc sql;

create table ia.c_2 as
select distinct household_key, total_trips/(104 - frst_start_week) as avg_trips,
total_spend/(104 - frst_start_week) as avg_spend, campaigns
from ia.c_1
group by household_key
order by household_key;

quit;

* combining data for differences in stats ;

proc sql;

create table ia.d as
select distinct a.household_key, b.avg_trips - a.avg_trips as diff_avg_trips,
b.avg_spend - a.avg_spend as diff_avg_spend, b.campaigns
from ia.b_2 a, ia.c_2 b
where a.household_key = b.household_key
group by a.household_key
order by a.household_key;

quit;

* adding the coupon user variable ;
proc sql;

create table ia.d_1 as
select *, 1 as coupon_user
from ia.d
where household_key in (select distinct household_key from
ia.coupon_redempt);

quit;

proc sql;

create table ia.d_2 as
select *, 0 as coupon_user
from ia.d
where household_key not in (select distinct household_key from
ia.coupon_redempt);
```

```
quit;

proc sql;

create table ia.d_3 as
select *
from ia.d_1
union
select *
from ia.d_2;

quit;

proc means data = ia.d_3 mean clm alpha = 0.05;
        var diff_avg_trips diff_avg_spend;
run;

proc sort data = ia.d_3 out = ia.d_4;
        by coupon_user;
run;

proc means data = ia.d_4 mean clm alpha = 0.05;
        var diff_avg_trips diff_avg_spend;
        by coupon_user;
run;
```

```
* Table 22. Covariance Structure Selection. ;

* all distinct households ;
proc sql;

create table ia.z as
select distinct household_key
from ia.transaction_data
group by household_key;
quit;

* all campaigns by start date ;
proc sql;

create table ia.zz as
select a.household_key, a.campaign, b.start_day, round((b.start_day/7),1) as
frst_start_week
```

```
from ia.campaign_table a, ia.campaign_desc b
where a.campaign = b.campaign
group by a.household_key, a.campaign, b.start_day
order by a.household_key;

quit;

* quarter uno ;

proc sql;

create table ia.zz_1 as
select distinct b.household_key, 'Qtr1' as quarter, sum(a.sales_value) as
total_spend
from ia.transaction_data a right join ia.z b on a.household_key =
b.household_key
where a.week_no between 1 and 13
group by b.household_key;

quit;

proc sql;
create table ia.zz_2 as
select distinct household_key, count(distinct campaign) as campaign
from ia.zz
where frst_start_week between 1 and 13
group by household_key;
quit;

proc sql;
create table ia.zz_uno as
select *
from ia.zz_1 a full join ia.zz_2 b on a.household_key = b.household_key
where a.household_key <> . or b.household_key <> .;
quit;

* quarter dos ;

proc sql;

create table ia.zz_1 as
select distinct a.household_key, 'Qtr2' as quarter, sum(a.sales_value) as
total_spend
```

```
from ia.transaction_data a
where a.week_no between 14 and 26
group by a.household_key;

quit;

proc sql;
create table ia.zz_2 as
select distinct household_key, count(distinct campaign) as campaign
from ia.zz
where frst_start_week between 14 and 26
group by household_key;
quit;

proc sql;
create table ia.zz_dos as
select *
from ia.zz_1 a full join ia.zz_2 b on a.household_key = b.household_key
where a.household_key <> . or b.household_key <> .;
quit;

* quarter tres ;

proc sql;

create table ia.zz_1 as
select distinct a.household_key, 'Qtr3' as quarter, sum(a.sales_value) as
total_spend
from ia.transaction_data a
where a.week_no between 27 and 39
group by a.household_key;

quit;

proc sql;
create table ia.zz_2 as
select distinct household_key, count(distinct campaign) as campaign
from ia.zz
where frst_start_week between 27 and 39
group by household_key;
quit;

proc sql;
```

```
create table ia.zz_tres as
select *
from ia.zz_1 a full join ia.zz_2 b on a.household_key = b.household_key
where a.household_key <> . or b.household_key <> .;
quit;

* quarter quatros ;

proc sql;

create table ia.zz_1 as
select distinct a.household_key, 'Qtr4' as quarter, sum(a.sales_value) as
total_spend
from ia.transaction_data a
where a.week_no between 40 and 52
group by a.household_key;

quit;

proc sql;
create table ia.zz_2 as
select distinct household_key, count(distinct campaign) as campaign
from ia.zz
where frst_start_week between 40 and 52
group by household_key;
quit;

proc sql;
create table ia.zz_quatros as
select *
from ia.zz_1 a full join ia.zz_2 b on a.household_key = b.household_key
where a.household_key <> . or b.household_key <> .;
quit;

* quarter cinqos ;

proc sql;

create table ia.zz_1 as
select distinct a.household_key, 'Qtr5' as quarter, sum(a.sales_value) as
total_spend
from ia.transaction_data a
where a.week_no between 53 and 65
```

```
group by a.household_key;

quit;

proc sql;
create table ia.zz_2 as
select distinct household_key, 'Qtr5' as quarter, count(distinct campaign) as
campaign
from ia.zz
where frst_start_week between 53 and 65
group by household_key;
quit;

proc sql;
create table ia.zz_cinqos as
select *
from ia.zz_1 a full join ia.zz_2 b on a.household_key = b.household_key
and a.quarter = b.quarter
where a.household_key <> . or b.household_key <> .;
quit;

* quarter setes ;

proc sql;

create table ia.zz_1 as
select distinct a.household_key, 'Qtr6' as quarter, sum(a.sales_value) as
total_spend
from ia.transaction_data a
where a.week_no between 66 and 78
group by a.household_key;

quit;

proc sql;
create table ia.zz_2 as
select distinct household_key, count(distinct campaign) as campaign
from ia.zz
where frst_start_week between 66 and 78
group by household_key;
quit;

proc sql;
```

```
create table ia.zz_setes as
select *
from ia.zz_1 a full join ia.zz_2 b on a.household_key = b.household_key
where a.household_key <> . or b.household_key <> .;
quit;

* quarter sieven ;

proc sql;

create table ia.zz_1 as
select distinct a.household_key, 'Qtr7' as quarter, sum(a.sales_value) as
total_spend
from ia.transaction_data a
where a.week_no between 79 and 91
group by a.household_key;

quit;

proc sql;
create table ia.zz_2 as
select distinct household_key, count(distinct campaign) as campaign
from ia.zz
where frst_start_week between 79 and 91
group by household_key;
quit;

proc sql;
create table ia.zz_sieven as
select *
from ia.zz_1 a full join ia.zz_2 b on a.household_key = b.household_key
where a.household_key <> . or b.household_key <> .;
quit;

* quarter ocho ;

proc sql;

create table ia.zz_1 as
select distinct a.household_key, 'Qtr8' as quarter, sum(a.sales_value) as
total_spend
from ia.transaction_data a
where a.week_no between 92 and 104
```

```
group by a.household_key;

quit;

proc sql;
create table ia.zz_2 as
select distinct household_key, count(distinct campaign) as campaign
from ia.zz
where frst_start_week between 92 and 104
group by household_key;
quit;

proc sql;
create table ia.zz_ocho as
select *
from ia.zz_1 a full join ia.zz_2 b on a.household_key = b.household_key
where a.household_key <> . or b.household_key <> .;
quit;

* el combinationez ;

proc sql;
create table ia.zz_top as
select * from ia.zz_uno
union
select * from ia.zz_dos
union
select * from ia.zz_tres
union
select * from ia.zz_quatros
union
select * from ia.zz_cinqos
union
select * from ia.zz_setes
union
select * from ia.zz_sieven
union
select * from ia.zz_ocho;
quit;

proc sql;

create table ia.distinct_campaign as
```

```
select distinct household_key
from ia.zz;

quit;

proc sql;

create table ia.zz_justcampaingpeeps as
select *
from ia.zz_allhouseholds
where household_key in (select household_key from ia.distinct_campaign);

quit;

proc sql;

create table ia.zz_zz as
select household_key, quarter, total_spend,
case when (campaign > 0) then 1 else 0 end as campaign
from ia.zz_justcampaingpeeps;

quit;

proc sql;

create table ia.zz_zzz as
select *
from ia.zz_zz
where household_key in (select distinct household_key from
ia.coupon_redempt);

quit;



proc mixed data = ia.zz_zz;
	class quarter household_key campaign;
	model total_spend = quarter campaign campaign*quarter;
	random household_key(campaign);
run;

proc sql;
create table ia.zz_zz1 as
```

```
select household_key,
case when quarter = 'Qtr1' then 1
          when quarter = 'Qtr2' then 2
          when quarter = 'Qtr3' then 3
          when quarter = 'Qtr4' then 4
          when quarter = 'Qtr5' then 5
          when quarter = 'Qtr6' then 6
          when quarter = 'Qtr7' then 7
          when quarter = 'Qtr8' then 8
          else 0 end as quarter,
total_spend, campaign
from ia.zz_zz;
quit;

proc mixed data = ia.zz_zz1;
         class campaign household_key quarter;
         model total_spend = campaign quarter campaign*quarter;
         repeated / type=cs sub=household_key(campaign) r rcorr;
run;


proc mixed data = ia.zz_zz1;
         class campaign household_key quarter;
         model total_spend = campaign quarter campaign*quarter;
         repeated / type=ar(1) sub=household_key(campaign) r rcorr;
run;

proc mixed data = ia.zz_zz1;
         class campaign household_key quarter;
         model total_spend = campaign quarter campaign*quarter;
         repeated / type=ar(1) sub=household_key(campaign) r rcorr;
         random intercept / sub=household_key(campaign);
run;


proc mixed data = ia.zz_zz1;
         class campaign household_key quarter;
         model total_spend = campaign quarter campaign*quarter;
         repeated / type=un sub=household_key(campaign) r rcorr;
run;
```

```
* Figure 10.  Cubic Regression Effect Significance. ;
proc mixed data = ia.zz_zz1;
         class campaign household_key;
```

```
        model total_spend = campaign quarter quarter*campaign quarter*quarter
                                                      quarter*quarter*campaign
quarter*quarter*quarter
                                                      quarter*quarter*quarter*campaign/
htype=1;
        repeated / type=un sub=household_key(campaign);
run;
```

```
* Figure 11.  Cubic Regression Coefficient by Model. ;
proc mixed data = ia.zz_zz1;
        class campaign household_key;
        model total_spend = campaign quarter*campaign
quarter*quarter*campaign
                                                      quarter*quarter*quarter*campaign /
noint s htype=1;
        repeated / type=un sub=household_key(campaign);
run;
```

VITA

Milena Hanna Chotard was born Milena Hanna Dydak in Warsaw, Poland in 1976. Her family has a strong history of teaching, and a University professorship in Nigeria for her step-father, Stan Kasprzyk, was their ticket to exit Poland during the 1980s conflicts between the Communist Party and the Solidarity movement. Milena's family emigrated to Yola, Nigeria in 1983. At the time, Nigeria was governed by a military junta, or dictatorship, resulting in widespread violence, tortures, and murders in Yola. In 1985, Milena's family fled from Nigeria to Toronto Canada, where Milena spent her teen years. In 1995, Milena enrolled in the University of Tennessee, where her father, Jerzy Dydak, had been a professor of mathematics since 1984. Milena graduated with two degrees from the University of Tennessee in Finance and Statistics in December, 1999.

Upon graduating, Milena accepted a position with Circuit City, Inc's (NYSE: CC) corporate office in Richmond, VA as a financial analyst. Milena rapidly advanced through the company over the next three years, with direct management responsibility for a team of six employees, and for project budgets of over $30 million. Milena obtained her Black Belt certification and was one of the pilot Black Belts for the Circuit City Six Sigma program. As a Black Belt, she performed statistical analysis including regression, chi square, hypothesis testing, normality testing, f-test, t-test, and variance analysis, while leading project teams. In September of 2003, Milena accepted an offer as a Portfolio Manager of an $8 billion portfolio of auto loans at Sun Trust Banks, Inc. (NYSE: STI) Corporate Office in Richmond, VA. In 2004, Milena accepted yet another offer with Portfolio Recovery Associates (NYSE: PRAA) in Norfolk, VA, where she was responsible for all statistical testing within the company. Milena's efforts there generated over $2 million in bottom line profits.

In 2006, Milena returned to the University of Tennessee as a graduate teaching assistant, and in pursuit of a Masters in Statistics. She taught Statistics 201 ("Intro to Statistics", over 100 students) in the Fall, 2007 and Spring, 2008 semesters. Milena expects to graduate with honors in August of 2008.

Milena is married to Christopher Chotard, has an 18 month old daughter, Lauren Chotard, and a son due in August of 2008. Milena enjoys running and has completed two marathons. Her other interests include the outdoors, travel, and spending time with her family.