SEVEN SUMMITS
OF MARKETING RESEARCH
Decision-Based Analytics for Marketing’s Toughest Problems
Seven Summits of Marketing Research

Decision-Based Analytics for Marketing’s Toughest Problems

Greg M. Allenby
Ohio State University
Columbus, Ohio

Jeff D. Brazell
The Modellers Inc.
Salt Lake City, Utah

February 12, 2016
Contents

The Basics iii

1 Market Definition 1

2 Market Segmentation 43

3 Customer Satisfaction 71

4 Product Analysis 117

5 Pricing Analysis 145

6 Advertising Analysis 177

7 Optimization 205

Appendix A: Questionnaires 219

Appendix B: Review of Linear Algebra 257

Appendix C: Review of Statistics 279

Bibliography 313
The Basics

The basic premise of marketing is to make things that people will want to buy. The reason is that it is easier to change the product to fit the person than to change the person to fit the product. People lead busy lives, often with so many tasks and pursuits on their minds that they don’t have the time or inclination to change or be changed. Changing a person means changing the patterns in which they allocate the resources they have at their disposal, including their time, money and attention. Changing the opinions, habits and preferences of an individual is possible, but requires such a huge investment that it is often prohibitively costly. Instead, marketing operates by trying to understand what might be of interest to individuals within the context of their lives as it currently exists, and directs firms to be relevant to these tasks and pursuits.

Marketing also operates by understanding the capabilities of the firm, and attempts to bend the current production process to produce goods and services that could be produced profitably. In the short run, firms have a relatively fixed set of skills, and the addition of skills for the purpose of producing something new creates organizational challenges that can be cost prohibitive for the venture to succeed. The acquisition and retention of capabilities is the subject of business strategy, not marketing strategy. In our analysis, we take the view that marketing’s advice to the firm is bound within strategic decisions and systems that are costly to change in the short run, just as it is costly to change individuals.

Marketing must therefore act in a manner that profitably relates two sets of goal-directed behaviors, or decisions. We assume that both firms and individuals are goal-directed in the use of their resources as they seek to improve their well-being by allocating resources and taking actions that make sense to them at the time they do it. This behavior may not appear “rational” to an outside observer, but the goal of marketing analysis is to try to make sense of consumer behavior and to guide firms in the best use of their limited resources. In this sense, it is never appropriate to claim that the actions of a person or a firm are “irrational” because this only means that the analyst has failed to make sense of these behaviors. We assume that consumer tastes and pursuits may change
over time, that people and firms act on information that we, as analysts, may not observe and that our data and models are incomplete. For these reasons and others, error terms are present in our models of consumer behavior, and our analysis is intended to guide the decisions of a firm but not be definitive – managerial judgement and creativity will always play important roles in marketing decision making.

We believe that research in marketing must embrace the idea of being decision-based and not being perfectly rational to us. We present a set of decision-based analytics for understanding consumer behavior and for exploring the consequences of decisions made by a firm and its competitors. We show that analysis based on what people have done in the past, or what they would do now, is more informative than analysis based on what they currently think or feel. We find that data collected on fixed-point scales (e.g., a 7-point scale) is both time-consuming to collect and creates problems in analysis when people interpret the scale points differently. We propose the use of simple scales that ask respondents to indicate the presence or absence of a behavior or decision. The strength of effect is then estimated from the data instead of being obtained directly from the respondent.

The use of data reflecting decisions is a common theme of the analysis in this book. It guides us in the development of our models and helps us answer the question of “what do we learn from the data?” Our experience is that analysis in marketing does not often embrace the meaning and behavioral implication of its models. Consider, for example, a standard regression model:

\[ y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_k x_{k,i} + \varepsilon_i \]

which is used extensively in marketing research. Here “\( y \)” denotes an outcome variable, the “\( x \)” variables are the \( k \) different inputs, “\( i \)” indexes the different observations of the system, and the term “\( \varepsilon \)” is the error term that is typically assumed to be distributed Normal with a variance equal to \( \sigma^2 \). The regression coefficients \( \beta_j, j = 0, \cdots, k \) indicate the expected change in the output variable (\( y \)) for a change in the input variable (\( x \)) holding fixed all other inputs. The regression model literally implies that the input variables are weighted by the regression coefficients, and then added up along with the error term to obtain the value of the output variable. This represents a gross approximation of how we know marketing really works. We know, from years of collective study and application, that buying decisions are staged, that some variables (e.g., advertising) make other variables (e.g., pricing) more effective and that some variables have mediating effects while others have moderating effects on the output variable. None of these considerations are reflected in the regression model, and so it must be modified before it can be used for marketing analysis. Possibly the worst approximation is that the regression
model does not reflect goal-directed, decision-based behavior of the individuals in the
context of their lives.

Consider what we will learn from applying the standard regression model to marketing
data. The first thing to realize is that it does not give special meaning to any particular
value of the outcome \( (y) \) or input \( (x) \) variables. A data value of zero (e.g., \( y = 0 \)) is
treated the same in the model as a value of one, or any other value. But, intuitively,
if our data comes from a goal-directed process, we should expect our model to treat an
outcome or an input value of zero differently than a positive value for the variable. The
reason is the values of the variables have meaning, and zero is typically used to represent
“not” – not buying, not doing, not considering, not remembering, etc., whereas positive
values represent the opposite. Sometimes all we know is that a person does or does
not already “do” something, and at other times we observe the extent to which they do
something. Examples include the amount of time spent engaging in a specific activity,
or the amount of money spent in a specific product category. If we observe the extent to
which individuals allocate goal-directed resources, then we learn much more than if we
simply know that no allocation is made.

A natural question to ask at this point is whether goal-directed behavior is a reason-
able assumption. There are many instances where the assumption of being goal-directed
seems untenable. An example is an individual who makes a spur-of-the-moment choice,
such as when standing in the checkout line of a grocery store. In this case, it is not clear
what goal an individual is pursuing. One story for the behavior of impulsive purchases
is that an individual wants something simply because they see it, or because a neighbor
has one. While many of the things we buy can be categorized as impulsive purchases,
our assumption of goal-directed behavior is intended to pertain to individuals responding
to their needs in a more thoughtful and directed manner.

This book attempts to provide a formal way of dealing with the discreteness of mar-
keting data that arise from decisions on the part of individuals and firms, acknowledging
the fact that research in marketing is a progressive pursuit that starts with the definition
of a market and ends with a goal of optimizing the marketing mix. Our treatment of
these subjects is not meant to be exhaustive, but instead is meant to provide one way of
engaging in theoretically grounded analysis. Within each of the topics we examine in this
book, there is vast literature that we will barely touch upon. However, even though the
breadth and depth of our coverage is limited, our hope is that the material we present
will serve as a foundation for additional in-depth treatment for each of the topics.

Figure 1 provides an overview of the material in this book. Chapter 4, Product
Analysis, serves as the focal construct for our analysis. The first three chapters of the
book (Market Definition, Market Segmentation and Customer Satisfaction) present anal-
ysis helpful for determining the competitive brands and variables of use in conducting
product analysis. The last three chapters (Pricing Analysis, Advertising Analysis and Optimization) enhance and extend basic product analysis. Thus, while marketing as a discipline is concerned with a number of different decision variables, Product Analysis is seen as the most important.

Figure 1: Overview of Material

A brief summary of the material contained in each chapter is as follows:

1. Market Definition – to establish the boundaries of analysis.
2. Market Segmentation – to understand the needs of potential customers.
4. Product Analysis – to relate needs to wants.
5. Price Analysis – to translate wants into demand.
6. Advertising Analysis – to incorporate brand belief and consideration effects.
7. Optimization – to coordinate product, price and promotion decisions.

Chapters 1 and 2 introduce the reader to exploratory analysis in marketing, Chapter 3 deals introduces predictive analysis useful for assessment, and Chapters 4 through 7 examine analysis associated with intervention, or change to the marketing mix. Exploration,
prediction and intervention span the spectrum of analyses conducted in marketing. Data from two national surveys are used to illustrate these forms of analyses. The survey questionnaires are provided in Appendix A, and contain data used in various formats encountered in marketing research:

- Pick any/J format – activities, brands used, media consumed, channels frequented, needs, general attitudes.
- 7-point Likert scale – brand beliefs and overall satisfaction for brands used.
- Counts – advertising exposures.
- Binary responses – purchase intentions and likely to recommend.
- Multinomial responses – choice-based conjoint analysis.
- Volumetric responses – anticipated demand for product concepts.

We discuss analysis of these data in the context of a brand manager wishing to reposition a brand to maximize inroads into a market target. Exercises provided at the end of each chapter assume that students are assigned to teams to carry out their analysis, with class discussion intended to center around results of their analyses. The surveys are of two product categories, Ice Cream and Florida Vacations, with data that are used to motivate and illustrate the analysis.

Why We Wrote This Book

This book brings together two recent developments from the academic and practitioner communities. Bayesian statistical models and Markov chain Monte Carlo methods have become increasingly visible in the academic literature because of their ability to produce respondent-level coefficients for complex models. Choice simulators and spreadsheet analysis have likewise been increasingly used by practitioners to predict the effects of alternative actions. We bring together these developments and propose a set of practical solutions to some commonly encountered problems in marketing research. We believe that all practitioners should be comfortable with spreadsheet analysis of the type discussed in this book, and that many will benefit from a greater understanding of the assumptions and methods used to obtain the coefficients they employ. Likewise, academic readers should be familiar with the more technical material in this book and would benefit from seeing how model parameter estimates can be used to inform substantive marketing problems. Our book is intended to provide an understanding of the
theory underlying marketing research methods, and an appreciation for the analysis that can ensue once parameter estimates are available. Software accompanying our book can be downloaded from:

http://themodellers.com/AnalyticsTool

It is our experience that analysis in marketing is often conducted with an emphasis on prediction, not on inference. Spreadsheets have made prediction popular, but often at the expense of sound inference. Many of the procedures used in marketing research analysis can be thought of as recipes for conducting analysis, with multiple techniques used in sequence to produce insights. When asked for the justification of these procedures, advocates usually defend their methods in vague terms such as “it seems to work well.” When pressed for greater justification, the best that is offered is that the results seem to predict well.

It is our view that inference is under-appreciated in marketing analytics as results are used to guide management to make things that people will want to buy. Inference involves the estimation of effect sizes and the identification of the manner in which variables affect choices. Management is interested in knowing which product attributes are highly valued by which consumers, and which attributes give a brand superior standing in the marketplace. Management wants to know which competitors it faces, how it might effectively make inroads in attracting new customers and how to best respond to competitive initiatives such as price cuts or a new product introduction. All of these scenarios have inference being equally important, and often more important than prediction. This is not to say that prediction isn’t important, just that it is less important than inference when management is considering interventions for improving the state of their brand.

We believe that analysis is coherent when it can be shown to provide correct inference about an assumed data-generating mechanism such as a consumer’s decision process. The data-generating mechanism is referred to as the “likelihood” in statistical analyses, and provides an explicit formula for how the data are thought to arise – i.e., the likelihood provides a prescription for simulating the data if the model parameters are known to us. Data for the regression model described previously can be simulated by assuming a set of independent variables \(x\) and a set of parameter values \(\{\beta, \sigma^2\}\) and simulating values of the dependent variable \(y\). Then, given the data \(\{y, x\}\) it should be the case that an analysis procedure can recover the parameter values with accuracy up to that implied by the parameter’s error bounds. This demonstration of data simulation (or forecasting) and parameter recovery is necessary for coherent inference to be present in any actual data analysis. Unfortunately, most procedures used in the analysis of marketing research data fail this test. The reason is that multi-step procedures are rarely explicit about the
assumed data-generating mechanism, or likelihood. Without clarity about the likelihood, there can be no confidence that an analysis procedure “works.”

All of the procedures proposed in this book are explicit about the formulation of the likelihood and can be shown to recover true parameter values in simulation experiments. We take a Bayesian orientation to our analysis of marketing research problems, although this orientation is not usually brought to the attention of the reader. The Bayesian orientation adheres to what is known as the likelihood principle, which states that all of the information in the data about the parameters of a model are contained in the model likelihood. As a result, the analysis and methods proposed in this book are coherent in that they provide true inferences about underlying model parameters.

We hope you enjoy working through the topics in the book, and that you find value in the analysis presented. Our intent is to offer a self-contained treatment of topics that provides a benchmark for data collection and analysis, as well as for assessing alternative approaches that are discussed throughout the book. The appendices to the book contain technical material for readers who want to understand the inner-workings of the methods we will describe.

We are indebted to many people for making this book possible. We thank our families for their support and encouragement. We thank our institutions, Ohio State University and The Modellers, Inc., for providing material resources and feedback. We also thank Intercept Survey Inc. for fielding the surveys and collecting data, as well as our colleagues at various universities and organizations for their support, including Alex Varbanov and Michael Thompson from Procter and Gamble who provided critical and helpful feedback, Marc Dotson from The Modellers for careful editing, and Michelle Petrel for her constructive and encouraging comments. Helpful and extensive comments were also provided by current and former students, including Jaehwan Kim and his students at Korea University, and Ohio State doctoral students John Howell, Sanghak Lee, Tatiana Dyachenko. Thanks also go to the programming team at the Modellers, particularly Todd Humphries, Mike Smith and David Guell for their diligent work producing the interactive decision tool (IDT). Matt Madden deserves special recognition for the excellent tutorials produced for the IDT.

We are particularly thankful to the Ohio State students contributing vignettes found throughout the book. The student’s initials are found at the end of each text box.
Chapter 1

Market Definition

Establishing the boundaries of analysis.

A market is the place where transactions occur, and is defined in terms of the people, contexts, products, media and means of exchange. In this chapter we learn how to analyze past marketplace decisions – past purchases, consumption contexts and media usage – to understand the competitive landscape. Our analysis is based on discrete data describing discrete behaviors, such as the brands a respondent has recently used. We show how to model this data with a latent, unobserved multivariate normal distribution, and how to employ the data reduction technique known as principal component analysis to produce competitive maps that help establish the boundaries of analysis.

1.1 Introduction

There are two sides to every market - the producer side and the consumer side. For every product category, there are one or more activities for which the category is potentially useful. For every activity, there are one or more corresponding product categories. Understanding the relationships that exist among and between categories and activities is the first of the seven summits of marketing analysis.

A market is the arena in which transactions take place. It involves people finding help for the tasks and pursuits of their lives, and firms offering related goods and services with the intent of making a profit. Market analysis helps firms understand their current customers, and helps identify opportunities for new customers and the potential demand for new offerings.
There are multiple ways of approaching this analysis, including the analysis of substitution patterns based on changes in prices and product availability and the timing of purchases. In legal proceedings, the identification of the boundaries of a market are defined in terms of local monopoly power and the ability of firms to profit from small but significant price increases. Our goal is to understand patterns in consumer behavior — the brands people have purchased, where they purchased them, how they learned about them and the contexts of consumption — so that we may establish the boundaries of our analysis. Not all brands, all purchase outlets, all advertising media and all contexts are of equal interest to a firm. Market definition serves to narrow the focus of a firm’s offerings.

### 1.2 Dimensions of a Market

The central issue of marketing analysis is determining the answer to the question “What shall we offer to whom?” The challenge in answering this question is that neither the offering nor the target consumers are fixed. Both are variables in the decision. As we consider aspects of analysis for market definition, it is important to realize that a market establishes the outer boundaries of all our analysis in marketing. That is, we conduct analysis on the structure of a market to help identify a subset of individuals to serve — individuals who may already be participating in the market in some way. The process of defining a market helps us to take the first step toward understanding whom we intend to target. The second and final step occurs when we segment the market, which is discussed in Chapter 2.

As firms engage in business with their customers, answers to the following questions are either implicitly or explicitly answered (see Fennell and Allenby (2003)):

1. What will we offer?
2. In what broad range of price?
3. To whom will we offer it?
4. How will we let them know about our offering?
5. How and where will we engage in exchange with them?
6. With whom will we compete?
Prior to engaging in any marketing research, firms will have general answers to the first two questions – what they intend to offer and at what general price level. The purpose of marketing research analysis is to provide specific answers to these and the rest of the questions on the list. This is done within the current constraints and expertise of the firm. If a firm does not currently have expertise in a particular type of manufacturing or production, or in a particular type of channel, it is doubtful that a particular marketing venture will yield sufficient revenue to warrant movement into these new areas of competition. Mergers and acquisitions are topics of business strategy, not marketing strategy. Marketing strategy operates within the limits of existing expertise and competencies, and seeks to modify current capabilities of the firm in a direction that would profitably serve some, but not all, people.

Defining a market involves setting boundaries, or limits, on the set of answers to the questions listed above. It identifies i) the general type of product to be offered, including key features; ii) the general level of price; iii) the set of activities for which the product might be used; iv) the channels used by consumers to acquire competitive offerings; v) media used by consumers in the category; and vi) other brands similar on these dimensions.

The definition of a market is not established by an external governing body or a set of predetermined specifications. While many companies may be classified as shoe manufacturers, Nike’s market is different from that of Kenneth Cole when markets are thought of in terms of types of shoes: casual shoes, sandals, work/safety shoes, athletic shoes and subcategories within each such as running shoes.

Every company defines its market as it deems best by establishing an arena of exchange. This definition is guided by the competencies of the firm and its ability to make competitive inroads in the marketplace. Defining one’s market helps a firm guide its analysis by focusing attention on the important elements of competition, which can be different for different brands. (BK)

The result of market definition analysis is the identification of prospects, brands and other elements of a market for further analysis. A prospect is a person with the willingness and ability to potentially part with their money and other resources for the right to acquire and use a firm’s offering. Prospects need not be current customers of a firm, or a current customer of the competitors in a product category. Prospects include people who find the current array of offers deficient and decide to do without purchased
solutions to their problems. Not all dog owners, for example, find the current array of dog food offerings acceptable, and may instead feed their dogs table scraps. Similar non-participation occurs in all product markets.

1.3 Analysis for Market Definition

In practice, questionnaires use various questions to qualify people for inclusion, and these questions tend to screen out individuals who are not currently “in” the product market (see questions S1-S8 in the Ice Cream survey and questions S1-S10 in the Florida Vacation survey in Appendix A). Respondents are qualified for inclusion in a survey if certain criteria are met, such as past purchase/participation in the product category, or demographic requirements such as age, gender and income. These requirements typically relate to the ability and willingness of respondents to participate (i.e., purchase, use, consider) in the product category. The reason unqualified people are screened out of participation is because their responses are qualitatively different from the responses from individuals who are qualified. It makes no sense, for example, to ask respondents about aspects of a service they have not encountered in the past.

Defining a market is similar to the process of deciding who to include in a questionnaire in that it operates by exclusion. Its goal is to screen out objects of analysis — respondents, alternative products, distribution outlets and media — that are not of immediate concern to the firm. Doing so focuses attention on the elements of a market within which a firm chooses to participate. Market definition is therefore an active decision made by a firm, and is not something that firms are expected to agree upon. Two firms operating in the same product market (e.g., blue jeans) may select a different set of competitors as being in their market, and may choose to compete using different media because their strengths and competencies differ.

We will conduct analysis on the following subset of market-defining variables:

2. Variables describing the context of product usage.
3. Media consumption habits.
4. Channel usage.

Our data for market definition are collected using a “Pick any/J” format to reflect past decisions. An example of this format is provided in Figure 1.1, where respondents indicate their past brand purchases. Multiple items can be selected, and the goal of
1.3. ANALYSIS FOR MARKET DEFINITION

analysis is to understand the prevalence and response patterns of the selected items. Market definition analysis is used to indicate aspects of a market that are potentially important for analysis, and aspects that are probably not important.

Figure 1.1: Pick any/J Data Collection Format

Pick any/J data are obtained by having respondents indicate the items in a list that apply to them. J refers to the number of different choice alternatives. Examples are provided in the questionnaires used to collect the data analyzed in this book (see Appendix A). The data from each respondent is a list of variables, each taking on just two values: “1” if the item was selected and “0” if the item was not selected. Thus, Pick any/J data are extremely discrete. It is tempting to cross-tabulate these data and conduct simple analysis on the $2 \times 2$ tables that arise. This procedure quickly becomes unwieldy as the number of response categories increase. For $J=10$ response categories, the number of tables is $(10 \times 9)/2 = 45$, and for $J=20$ response categories the number of tables grows to $(20 \times 19)/2 = 190$. It is therefore useful to consider a more formal approach to reduce the high dimensionality of the analysis.

Modeling Pick any/J Data

A model for data involves two things – a story of how the data are generated and an assumption of a statistical distribution that gives rise to the data being random. Our data generating story involves latent, or unobserved, variables. Latent variables are common aspects of models in every discipline. Examples include the concept of a personality in
These variables are not directly observed but are thought to exist to explain things that are observed such as the constancy of behavior, the attraction of objects, the creation of compounds and choices in the marketplace. The purpose of assuming the existence of these latent variables is to provide an explanation of the observed data. For example, knowing a person’s utility function allows one to predict behavior under various competitive scenarios. Our goal for Pick any/J data is to parsimoniously represent patterns present in the responses.

We assume that the observed pick any/J data are censored realizations of a continuous variable that are distributed multivariate normal (see Appendix C) across the population of respondents (Edwards and Allenby (2003)). The advantage of employing the multivariate normal distribution is that it can flexibly reflect pair-wise associations among the variables, with properties that are also useful for understanding a wide variety of problems in marketing research. Respondents are assumed to indicate usage of a brand or particular medium, or their participation in an activity, if there is sufficient interest or value that prompts their memory to indicate a “yes.” We formalize this notion through a statistical model:

\[
\text{if } x_j = 1 \text{ then } z_j > 0 \text{ and if } x_j = 0 \text{ then } z_j \leq 0
\]

where \( x_j \) indicates the \( j^{th} \) element of the observed data, and \( z_j \) is a corresponding latent unobserved variable. We assume the latent variables \( z \) are distributed multivariate normal:

\[
z \sim \text{Normal} \left( \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_J \end{bmatrix}, R = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1J} \\ r_{21} & 1 & & r_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ r_{J1} & r_{J2} & \cdots & 1 \end{bmatrix} \right) \]

where \( R \) is a correlation matrix. A correlation \( r_{ij} \) describes the association between two variables, with positive values indicating a positive association where both variables tend to be large at the same time. A negative correlation indicates that a high value for one variable is usually accompanied by a low value for the other. The advantage of assuming the errors are correlated will become clear in the following sections(s).

Table 1.1 reports estimates of the correlation matrix for Ice Cream and Florida Vacation destination usage based on the raw Pick any/J data. The top portion of the table reports correlations for Ice Cream, and the bottom portion reports correlations for Florida Vacations.
### Table 1.1: Correlation Estimated Based Directly on Pick any/J Data

<table>
<thead>
<tr>
<th></th>
<th>Dreyer’s Blue Bunny</th>
<th>Blue Bell</th>
<th>Breyers</th>
<th>Ben and Jerry’s</th>
<th>Haagen-Dazs</th>
<th>Store Brand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dreyer’s Blue Bunny</td>
<td>1.000</td>
<td>-0.041</td>
<td>-0.100</td>
<td>0.003</td>
<td>0.058</td>
<td>0.060</td>
</tr>
<tr>
<td>Blue Bell</td>
<td>-0.041</td>
<td>1.000</td>
<td>0.036</td>
<td>0.020</td>
<td>-0.063</td>
<td>-0.090</td>
</tr>
<tr>
<td>Breyers</td>
<td>-0.100</td>
<td>0.036</td>
<td>1.000</td>
<td>-0.081</td>
<td>-0.056</td>
<td>-0.061</td>
</tr>
<tr>
<td>Ben and Jerry’s</td>
<td>0.003</td>
<td>0.020</td>
<td>-0.081</td>
<td>1.000</td>
<td>-0.047</td>
<td>-0.004</td>
</tr>
<tr>
<td>Haagen-Dazs</td>
<td>0.058</td>
<td>-0.063</td>
<td>-0.056</td>
<td>-0.047</td>
<td>1.000</td>
<td>0.309</td>
</tr>
<tr>
<td>Store Brand</td>
<td>0.060</td>
<td>-0.090</td>
<td>-0.061</td>
<td>0.004</td>
<td>0.309</td>
<td>1.000</td>
</tr>
<tr>
<td>Magic Kingdom</td>
<td>-0.022</td>
<td>-0.071</td>
<td>-0.004</td>
<td>0.005</td>
<td>-0.227</td>
<td>-0.146</td>
</tr>
<tr>
<td>Epcot Animal Kingdom</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hollywood Studios</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Universal Studios</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Islands of Adventure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Busch Gardens</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We find that the estimates based on the multivariate normal distribution are much larger than those based on the raw data, as seen in table 1.2, sometimes almost twice as large. Correlations based directly on the raw binary (0/1) data will lead to an underestimation of correlations because the latent variables are censored. The formula for calculating the correlation of two variables \((x, y)\) is:

\[
r_{x,y} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}
\]

When the true variables governing behavior are continuous, but an analyst only observes a censored version of the true latent variables, then large deviations of the latent variables from their means (\(\bar{x}\) and \(\bar{y}\)) are replaced with smaller deviations based on the binary data, resulting in smaller calculated values of \(r_{x,y}\). The difference in estimates of the correlation coefficients illustrates the importance of using formal statistical models for analyzing
marketing data, particularly when the data are discrete. Correlation coefficients should never be calculated directly from binary raw data.

**Principal Components**

Our goal is to reduce the complexity of analyzing the raw data collected on a Pick any/J format by providing pictures of the associations. The first step in this process is using a statistical model to obtain an estimate of the correlation matrix, $R$. The second step is to use principal components analysis to produce maps based on the correlation matrix. Principal component analysis explains the correlation matrix through a few linear combinations of the original variables $z' = (z_1, z_2, \cdots, z_J)$ using the mathematics of linear algebra (see Appendix B).

Principal component analysis provides a visual representation of the correlation matrix, where similar “objects” are plotted close to each other:
1.3. ANALYSIS FOR MARKET DEFINITION

Figure 1.2: Competitive Brand Map for Florida Vacations (S9).

Figure 1.2 indicates a high degree of substitution among the Disney offerings – Magic Kingdom, Epcot, Animal Kingdom and Hollywood Studios. Respondents attending one of these theme parks almost certainly also attend the others. A high rate of substitution is also present between Universal Studios and Universal’s Islands of Adventure. These high rates of substitution correspond to high pair-wise correlations estimated from the Pick any/J data for responses to the question (S9) “Which of the following theme parks have you visited in the past five years?”
Principal component analysis is a statistical technique that uses an orthogonal transformation to decrease the dimensionality of a set of observations. The resulting principal component maps are useful to marketers because they give valuable insight into how to define the competitive landscape, context of consumption, need states, and channel choices. There is an intrinsic trade off with principal component analysis: the valuable macro-level insights come at the expense of being able to resolve the distinct, individual data points. For example, consider the segment of a painting on the left below.

You can clearly identify each of the individual specks of color: auburn reds and rich cremes. These are like individual pick any data points. Interesting and valuable on their own but give little insight to the big picture. Now look at the full painting on the right. You can make out the full picture but have lost the ability to see the individual brush strokes you were able to see before. This is how principal component analysis works. You are taking a step back from the canvas so you see the whole composition. Both viewpoints are valuable and instructive but what you can learn from them is substantively different. (KE) Painting Credit: Maximillien Luce, Morning, Interior, 1890.

In principal components analysis, analysis begins by decomposing the correlation matrix into the product of three matrices:

\[
R = QLQ' \quad \text{and} \quad L = \begin{bmatrix}
\lambda_1 & 0 \\
& \ddots \\
0 & \ddots & \lambda_J
\end{bmatrix}
\]

with eigenvalues \( \lambda_1, \lambda_2, \ldots, \lambda_J \) forming a diagonal matrix and associated eigenvectors \( Q = [q_1, q_2, \ldots, q_J] \) forming a square matrix. We consider summaries of the data using
1.3. ANALYSIS FOR MARKET DEFINITION

linear combinations of the latent variables $z$ with weights $\ell$:

$$
y_1 = \ell_1'z = \ell_{11}z_1 + \ell_{21}z_2 + \cdots + \ell_{J1}z_J \\
y_2 = \ell_2'z = \ell_{12}z_1 + \ell_{22}z_2 + \cdots + \ell_{J2}z_J \\
\vdots \\
y_J = \ell_J'z = \ell_{1J}z_1 + \ell_{2J}z_2 + \cdots + \ell_{JJ}z_J
$$

Then

$$
\text{Var}(y_i) = \ell_i' R \ell_i \quad \text{and} \quad \text{Cov}(y_i, y_j) = \ell_i R \ell_j
$$

The principal components are those linear combinations $y_1, y_2, \ldots, y_J$ whose variances are as large as possible, while being orthogonal (uncorrelated) to each other. The first principal component is the linear combination $\ell_1'z$ that maximizes $\text{Var}(\ell_1'z)$ subject to $\ell_1' \ell_1 = 1$. The second principal component is the linear combination $\ell_2'z$ that maximizes $\text{Var}(\ell_2'z)$ subject to $\ell_2' \ell_2 = 1$ and $\text{Cov}(y_1, y_2) = \ell_1 R \ell_2 = 0$, etc. The solution to the problem is to set the weights $\ell$ equal to the eigenvectors of $R$, i.e., $(\ell_1, \ell_2, \ldots, \ell_J) = (q_1, q_2, \ldots, q_J)$. A proof of this claim is provided in Appendix B.

In general, we are interested in reducing the dimension of the original multivariate variables and yet continue to account for the majority of variance. A useful relationship for determining the amount of explained variance is:

$$
\text{tr}(R) = \text{tr}(QLQ') = \text{tr}(Q'QL) = \text{tr}(L) = \lambda_1 + \lambda_2 + \cdots + \lambda_n
$$

Therefore the proportion of the total population variance due the $k^{th}$ principal component is:

$$
\frac{\lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_n}
$$

Principal component analysis can provide a low-dimensional approximation to the correlation matrix by ignoring eigenvectors $(q)$ associated with small eigenvalues $\lambda$:

$$
R = \lambda_1 q_1 q_1' + \lambda_2 q_2 q_2' + \cdots + \lambda_r q_r q_r' + \cdots + \lambda_n q_J q_J' \\
\hat{R} = \lambda_1 q_1 q_1' + \lambda_2 q_2 q_2' + \cdots + \lambda_r q_r q_r'
$$

where $\hat{R}$ is an estimate of the correlation matrix $R$. To illustrate, consider the estimated correlation matrix for the Florida Vacations usage question examined above:
An approximation to the correlation matrix in Table 1.3 based on just one eigenvector $\{q_1\}$ is displayed in table 1.4. There are both similarities and differences in these two sets of correlations. The Disney offerings retain their high inter-correlations and there is a difference between Busch Gardens and the remainder of offerings. But, the Universal offerings (i.e., Universal Studios and Islands of Adventure) are not well distinguished from the Disney offerings. The approximation based on $\{q_1\}$ accounts for 65% of the total population variance.

The addition of a second eigenvector in the approximation in Table 1.5 serves to separate out the Universal offerings. The correlation in usage between Universal Studios and Hollywood Studios drops from 0.563 in table 1.4 to 0.375 in table 1.5. The estimated correlations reported in table 1.5 are based on the first two eigenvectors $\{q_1, q_2\}$ that

### Table 1.3: Model-Based Correlation Estimates

<table>
<thead>
<tr>
<th>Magic Kingdom</th>
<th>Epcot Kingdom</th>
<th>Animal Kingdom</th>
<th>Hollywood Studios</th>
<th>Universal Studios</th>
<th>Islands of Adventure</th>
<th>Busch Gardens</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>0.831</td>
<td>0.804</td>
<td>0.670</td>
<td>0.520</td>
<td>0.609</td>
<td>0.354</td>
</tr>
<tr>
<td>0.831</td>
<td>1.000</td>
<td>0.825</td>
<td>0.713</td>
<td>0.499</td>
<td>0.538</td>
<td>0.403</td>
</tr>
<tr>
<td>0.804</td>
<td>0.825</td>
<td>1.000</td>
<td>0.722</td>
<td>0.463</td>
<td>0.537</td>
<td>0.280</td>
</tr>
<tr>
<td>0.670</td>
<td>0.713</td>
<td>0.722</td>
<td>1.000</td>
<td>0.366</td>
<td>0.519</td>
<td>0.195</td>
</tr>
<tr>
<td>0.520</td>
<td>0.499</td>
<td>0.463</td>
<td>0.366</td>
<td>1.000</td>
<td>0.770</td>
<td>0.436</td>
</tr>
<tr>
<td>0.609</td>
<td>0.538</td>
<td>0.537</td>
<td>0.519</td>
<td>0.770</td>
<td>1.000</td>
<td>0.480</td>
</tr>
<tr>
<td>0.354</td>
<td>0.403</td>
<td>0.280</td>
<td>0.195</td>
<td>0.436</td>
<td>0.480</td>
<td>1.000</td>
</tr>
</tbody>
</table>

### Table 1.4: Principal Component Approximation: $\hat{R} = \lambda_1 q_1 q_1'$

<table>
<thead>
<tr>
<th>Magic Kingdom</th>
<th>Epcot Kingdom</th>
<th>Animal Kingdom</th>
<th>Hollywood Studios</th>
<th>Universal Studios</th>
<th>Islands of Adventure</th>
<th>Busch Gardens</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.791</td>
<td>0.793</td>
<td>0.770</td>
<td>0.694</td>
<td>0.642</td>
<td>0.710</td>
<td>0.465</td>
</tr>
<tr>
<td>0.793</td>
<td>0.794</td>
<td>0.771</td>
<td>0.695</td>
<td>0.643</td>
<td>0.711</td>
<td>0.465</td>
</tr>
<tr>
<td>0.770</td>
<td>0.771</td>
<td>0.749</td>
<td>0.676</td>
<td>0.624</td>
<td>0.691</td>
<td>0.453</td>
</tr>
<tr>
<td>0.694</td>
<td>0.695</td>
<td>0.676</td>
<td>0.609</td>
<td>0.563</td>
<td>0.623</td>
<td>0.408</td>
</tr>
<tr>
<td>0.642</td>
<td>0.643</td>
<td>0.624</td>
<td>0.563</td>
<td>0.520</td>
<td>0.576</td>
<td>0.377</td>
</tr>
<tr>
<td>0.710</td>
<td>0.711</td>
<td>0.691</td>
<td>0.623</td>
<td>0.576</td>
<td>0.637</td>
<td>0.417</td>
</tr>
<tr>
<td>0.465</td>
<td>0.466</td>
<td>0.453</td>
<td>0.408</td>
<td>0.377</td>
<td>0.417</td>
<td>0.273</td>
</tr>
</tbody>
</table>
1.3. ANALYSIS FOR MARKET DEFINITION

jointly account for 78% of the total population variance:

Table 1.5: Principal Component Approximation: $\hat{R} = \lambda_1 q_1^T + \lambda_2 q_2^T$

<table>
<thead>
<tr>
<th>Magic Kingdom</th>
<th>Epcot Kingdom</th>
<th>Animal Kingdom</th>
<th>Hollywood Studios</th>
<th>Universal Studios</th>
<th>Islands of Adventure</th>
<th>Busch Gardens</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.830</td>
<td>0.838</td>
<td>0.835</td>
<td>0.771</td>
<td>0.548</td>
<td>0.635</td>
<td>0.345</td>
</tr>
<tr>
<td>0.838</td>
<td>0.847</td>
<td>0.847</td>
<td>0.786</td>
<td>0.533</td>
<td>0.623</td>
<td>0.320</td>
</tr>
<tr>
<td>0.835</td>
<td>0.847</td>
<td>0.857</td>
<td>0.804</td>
<td>0.467</td>
<td>0.565</td>
<td>0.252</td>
</tr>
<tr>
<td>0.771</td>
<td>0.786</td>
<td>0.804</td>
<td>0.763</td>
<td>0.375</td>
<td>0.473</td>
<td>0.169</td>
</tr>
<tr>
<td>0.548</td>
<td>0.533</td>
<td>0.467</td>
<td>0.375</td>
<td>0.749</td>
<td>0.759</td>
<td>0.669</td>
</tr>
<tr>
<td>0.635</td>
<td>0.623</td>
<td>0.565</td>
<td>0.473</td>
<td>0.759</td>
<td>0.784</td>
<td>0.651</td>
</tr>
<tr>
<td>0.345</td>
<td>0.325</td>
<td>0.252</td>
<td>0.169</td>
<td>0.669</td>
<td>0.651</td>
<td>0.646</td>
</tr>
</tbody>
</table>

Adding a third eigenvector to the approximation of the correlation matrix $R$ in Table 1.6 results in estimates that are marginally better than that obtained using two eigenvectors, accounting for 88% of the total population variance. We see that the diagonal elements are closer to one, as they should be, and that correlations between Busch Gardens and the Universal offerings is lower. Further enhancements to the approximation from adding additional eigenvectors does little to improve the approximation, suggesting that consumer use of the seven offerings may be adequately explained by three underlying factors.

Table 1.6: Principal Component Approximation: $\hat{R} = \lambda_1 q_1^T + \lambda_2 q_2^T + \lambda_3 q_3^T$

<table>
<thead>
<tr>
<th>Magic Kingdom</th>
<th>Epcot Kingdom</th>
<th>Animal Kingdom</th>
<th>Hollywood Studios</th>
<th>Universal Studios</th>
<th>Islands of Adventure</th>
<th>Busch Gardens</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.833</td>
<td>0.848</td>
<td>0.838</td>
<td>0.769</td>
<td>0.525</td>
<td>0.618</td>
<td>0.378</td>
</tr>
<tr>
<td>0.848</td>
<td>0.880</td>
<td>0.856</td>
<td>0.777</td>
<td>0.461</td>
<td>0.568</td>
<td>0.430</td>
</tr>
<tr>
<td>0.838</td>
<td>0.856</td>
<td>0.860</td>
<td>0.802</td>
<td>0.447</td>
<td>0.550</td>
<td>0.282</td>
</tr>
<tr>
<td>0.769</td>
<td>0.777</td>
<td>0.802</td>
<td>0.765</td>
<td>0.394</td>
<td>0.487</td>
<td>0.142</td>
</tr>
<tr>
<td>0.525</td>
<td>0.461</td>
<td>0.447</td>
<td>0.394</td>
<td>0.902</td>
<td>0.878</td>
<td>0.441</td>
</tr>
<tr>
<td>0.618</td>
<td>0.568</td>
<td>0.550</td>
<td>0.487</td>
<td>0.878</td>
<td>0.876</td>
<td>0.474</td>
</tr>
<tr>
<td>0.378</td>
<td>0.432</td>
<td>0.282</td>
<td>0.142</td>
<td>0.441</td>
<td>0.474</td>
<td>0.984</td>
</tr>
</tbody>
</table>

As the number of eigenvectors used in the approximation increases, the closer the estimate of the correlation matrix $\hat{R}$ is to the true value reported in table 1.2.
Eigenvectors provide a way to view a situation from multiple angles. One way to understand them is to think about them as a series of pictures. You may want to take a photograph of something but aren’t sure which angle shows you the best view. In this case, you are not looking for the most picturesque view or aesthetically pleasing view, rather, you are seeking to understand how each object in your picture frame relates to each other. How close are they to one another? Is one larger than the next or is it just a deceptive angle you’re using that causes it to appear that way? For example, outside the Louvre in France, you can see this:

At first glance, you see statues surrounded by shrubs. Some couples visiting the gardens may even suspect this to be a cozy place to escape the crowds. It looks like a nice way to gain a little privacy. Like with eigenvectors, this is one way to view the situation (much like plotting the first two principal components). What happens when you view the same scene from a different angle?

Whoa! What a difference an eigenvector makes! From this angle, you can see the cozy spot is not so cozy after all. The bushes do not envelop the statues and they are shaped into points. Seeing the scene from this angle gives you a totally different picture just like eigenvectors do. (EW)
1.3. ANALYSIS FOR MARKET DEFINITION

Pictorial Representation

At first, the idea of the components included on a principal component plot can seem very arbitrary and confusing. You are ultimately able to label each axis of a plot as you see fit. For example, think about a recent job interview you completed. Most likely, the company interviewed you and a few other applicants. Although we hope this doesn’t happen, the interviewer could create a principal component plot and place each interviewee on the plot. However, the component on each axis could be a number of things. Take the principal component plot below as an example:

Here, the x-axis could be work experience and the y-axis could be labeled as desired salary. However, the x-axis could also be labeled as years of prior experience with the y-axis representing management potential. Overall, it is up to you as a marketer to determine what the component units of measurement should be. There is no standard axis label, which is a new, and sometimes frustrating concept. (SS)

A pictorial representation of the original data can be obtained with the eigenvectors serving as coordinates for the original variables $z$. Thus, variable $z_1$ has coordinate values \{q_{1,1}, q_{2,1}, \cdots, q_{r,1}\} and variable $z_n$ has coordinate values \{q_{1,n}, q_{2,n}, \cdots, q_{r,n}\} in the above approximation. A two dimensional representation of the variables is provided by the first two eigenvectors, with the proportion of the original variable explained by $(\lambda_1 + \lambda_2)/J$ for Pick/any J data. If two $z$ variables are plotted near each other, then they are explained
similarly by the $y$ variables. A proof of this claim is provided in Appendix B.

It is desirable to scale each column of $Q$ (the eigenvectors) by the eigenvalues so that distances in the resulting plots are proportional to the amount of “explained” variance. This improves the interpretation of the plots. The scaled eigenvectors are often referred to as the component loadings, and are related to the original eigenvectors by the transformation:

$$Q^* = Q L^{1/2} \quad \text{or} \quad q^*_i = \sqrt{\lambda_i} q_i$$

Finally, we scale the plotting characters by the mean $\mu$ of the statistical model. This allows us to quickly distinguish variables with a high incidence of selection versus those that are infrequently selected. Greater attention should be allocated to the variables (e.g., activities, brands used, media consumed) associated with a response ($x^*_j = 1$) as opposed to those with a non-response ($x^*_j = 0$).

In marketing, there are endless variables to analyze to maximize profits and growth. Trudging through the data can be as daunting as evaluating a crime scene with hidden clues. First you consider the data you have available, followed by setting parameters for your research. Lastly, you conduct your analysis. The data are the clues visible in the scene. The parameters are the quarantined area you wish to investigate and the analysis is the exhaustive review of the site.

Similarly in marketing, the data can be the competitors you are up against (Pick any/J data). The parameters are the aspects of these competitors you wish to review such as consumption and context. The analysis is the exhaustive review of gathered information.

As you look at your crime scene, there are many vantage points from which to consider your findings. Looking at the crime scene from the front may reveal certain clues. If you look from the side, you may find other clues. If you look from above, yet more clues.

The same is true as you review your marketing research. Different vantage points reveal different findings. These vantage points are called eigenvectors. The material in this chapter will help you achieve a better understanding of how to analyze data and unearth the clues to improve your marketing research. (AG)
1.4 Analysis of Questionnaire Data

Ice Cream

Tables 1.7–1.10 display the proportion of respondents answering “yes” to the Pick any/J questions describing aspects of the Ice Cream market. With Pick any/J data, respondents can respond to multiple response items, and as a result, the response probabilities do not need to add up to 100%. The specific question from which the data are taken is indicated in parenthesis – e.g., data from question S8 were used to generate the proportions displayed in Table 1.7. The questionnaires used for data collection are provided in Appendix A.

Table 1.7: Ice Cream Brands Purchased (S8)

<table>
<thead>
<tr>
<th>Brand</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dreyer’s</td>
<td>0.200</td>
</tr>
<tr>
<td>Blue Bunny</td>
<td>0.261</td>
</tr>
<tr>
<td>Blue Bell</td>
<td>0.186</td>
</tr>
<tr>
<td>Breyers</td>
<td>0.531</td>
</tr>
<tr>
<td>Ben and Jerry’s</td>
<td>0.369</td>
</tr>
<tr>
<td>Haagen-Dazs</td>
<td>0.271</td>
</tr>
<tr>
<td>Store Brand</td>
<td>0.531</td>
</tr>
</tbody>
</table>

Table 1.7 indicates that Breyers and the Store Brand of ice cream have the highest market penetration with more than half of the respondents indicating that they have purchased these brands within the last six months. Blue Bell, Dreyer’s and Blue Bunny have the smallest market penetration. Ben and Jerry’s and Haagen-Dazs are premium, upscale brands and it is interesting to note there exists a fairly large difference in the penetration of these brands. Blue Bunny and Blue Bell are of somewhat lower quality, and they too exhibit large differences in their penetration rates. The purpose of our analysis here and throughout this book is to identify why these differences exist and to suggest actions that can be taken to improve the performance of the brands.

Table 1.8: Ice Cream Consumption Contexts (S6)

<table>
<thead>
<tr>
<th>Context</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Routine Dessert</td>
<td>0.824</td>
</tr>
<tr>
<td>Special Occasion</td>
<td>0.541</td>
</tr>
<tr>
<td>Entertain Guests</td>
<td>0.393</td>
</tr>
<tr>
<td>Social Gathering</td>
<td>0.155</td>
</tr>
<tr>
<td>Special Treat</td>
<td>0.582</td>
</tr>
<tr>
<td>Other</td>
<td>0.020</td>
</tr>
</tbody>
</table>

Table 1.8 reports on the context of consumption. Ice cream is used by almost everyone as a dessert or snack, with 82.4% of respondents saying that this consumption context
is a routine part of their lives. More than 50% of respondents indicated that ice cream is used for special events, and nearly 40% of the respondents use ice cream to entertain guests in their homes.

A question that immediately arises is whether a respondent’s preferred brand of ice cream changes in different consumption contexts. It is important to remember, though, that market definition analysis does not provide specific answers to questions such as this. Instead, it provides general answers that are made more specific through additional analysis. For now, it is best to raise issues such as this in one’s mind for future investigation. We will return to the issue of quantifying the influence of context on brand preferences in Chapter 4 when we discuss Product Analysis.

Consumer purchase decision are affected by consumer consumption contexts and needs, with needs changing as contexts change. The same is true in sports. The fourth hitter in a baseball lineup is normally a power hitter. Many people think of him as the home run hitter for the team. However, if the game winning run is on third base, and there are no outs, the home run hitter may be asked to hit a sacrifice fly to the outfield, not a homerun. Or, if there is a runner on first and no outs, the coach might ask the power hitter to bunt the ball in order to move the runner into scoring position at second base. Good plays in baseball and good purchases in marketing respond to the situation at hand. (LL)

Table 1.9: Ice Cream Food and Nutrition Media (Q12)

<table>
<thead>
<tr>
<th></th>
<th>TV</th>
<th>Radio</th>
<th>Newspaper</th>
<th>Magazine</th>
<th>Website</th>
<th>Social Media</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.333</td>
<td>0.386</td>
<td>0.551</td>
<td>0.145</td>
<td>0.126</td>
<td>0.106</td>
</tr>
</tbody>
</table>

Table 1.9 describes where respondents typically get information about food and nutrition. More than half the respondents obtain food information from the newspaper, and only 10.6% obtain information from social media. It is also rare for individuals to obtain information by visiting a company’s website. TV and radio advertising is expected to reach about one third of the population. It is not clear, however, if the same people are reached by both of these advertising mediums.

Finally, Table 1.10 indicates that traditional supermarkets such as Albertsons, Safeway and Kroger are the preferred channel for purchasing ice cream. Nearly all respondents
1.4. ANALYSIS OF QUESTIONNAIRE DATA

Table 1.10: Ice Cream Purchase Channel (Q6)

<table>
<thead>
<tr>
<th>Traditional Supermkt</th>
<th>Natural Organic</th>
<th>Specialty Supermkt</th>
<th>Discount Mass Merchandise</th>
<th>Wholesale Club</th>
<th>Convenience Store</th>
<th>Ice Cream Store</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.928</td>
<td>0.048</td>
<td>0.065</td>
<td>0.356</td>
<td>0.108</td>
<td>0.087</td>
<td>0.100</td>
</tr>
</tbody>
</table>

(92.8%) report that they typically purchase ice cream at this venue. In addition, 35.6% of respondents indicate they purchase ice cream at discount mass merchandise outlets such as Walmart and Target.

The responses displayed in Tables 1.7–1.10 indicate the current state of individual variables that describe the ice cream market. Marketing strategy and the definition of one’s market, however, are often more concerned with the relationship among these variables – e.g., brands that are jointly purchased by individuals, consumption contexts that reflect a pattern of common behavior and channels of distribution that are unique.

The challenge of analyzing co-occurring responses is that the responses for the items reported above are not mutually independent. An analysis of the purchase information for the seven brands in Table 1.7 requires 21 (seven choose two) different two-way tables to describe the co-purchase relationships present in the data. Our hierarchical model applied to the brand use data reveals the following estimates of $\mu$ and $R$:

Table 1.11: Estimated Mean ($\mu$) for Ice Cream Brands (S8)

<table>
<thead>
<tr>
<th>Dreyer’s Blue Bunny</th>
<th>Blue Bell</th>
<th>Breyers</th>
<th>Ben and Jerry’s</th>
<th>Haagen-Dazs</th>
<th>Store Brand</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.835</td>
<td>-0.639</td>
<td>-0.885</td>
<td>0.081</td>
<td>-0.330</td>
<td>-0.610</td>
</tr>
</tbody>
</table>

The correlation matrix shows evidence of demand patterns that point to subsets of brands that compete with each other at different levels of intensity. The correlation between Ben and Jerry’s ice cream and Haagen-Dazs has a positive correlation of 0.447, while both brands are negatively correlated with the Store Brand.

Principal component analysis provides a parsimonious alternative to the analysis of cross-tabulated data, and provides a low-dimensional summary of the correlation matrix:
Table 1.12: Estimated Correlation Matrix ($R$) for Ice Cream Brands (S8)

<table>
<thead>
<tr>
<th>Dreyer’s Blue Bunny</th>
<th>Blue Bell</th>
<th>Breyers</th>
<th>Ben and Jerry’s</th>
<th>Haagen-Dazs</th>
<th>Store Brand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>-0.074</td>
<td>-0.212</td>
<td>0.007</td>
<td>0.097</td>
<td>0.102</td>
</tr>
<tr>
<td>-0.074</td>
<td>1.000</td>
<td>0.074</td>
<td>0.034</td>
<td>-0.111</td>
<td>-0.162</td>
</tr>
<tr>
<td>-0.212</td>
<td>0.074</td>
<td>1.000</td>
<td>-0.141</td>
<td>-0.089</td>
<td>-0.110</td>
</tr>
<tr>
<td>0.007</td>
<td>0.034</td>
<td>-0.141</td>
<td>1.000</td>
<td>-0.067</td>
<td>-0.004</td>
</tr>
<tr>
<td>0.097</td>
<td>-0.111</td>
<td>-0.089</td>
<td>-0.067</td>
<td>1.000</td>
<td>0.477</td>
</tr>
<tr>
<td>0.102</td>
<td>-0.162</td>
<td>-0.110</td>
<td>-0.004</td>
<td>0.477</td>
<td>1.000</td>
</tr>
<tr>
<td>-0.042</td>
<td>-0.113</td>
<td>0.000</td>
<td>0.009</td>
<td>-0.350</td>
<td>-0.243</td>
</tr>
</tbody>
</table>

Figure 1.3: Competitive Brand Map for Ice Cream (S8).
1.4. ANALYSIS OF QUESTIONNAIRE DATA

Figure 1.3 displays the eigenvector plot for brand purchases. The first two eigenvalues account for 42% of the total variation in the latent normal distribution ($\lambda_1 = 1.81, \lambda_2 = 1.25$) that gives rise to the Pick any/J data. Brands with a higher response incidence are plotted with a larger font, and brands with co-occurring responses are plotted closer to each other than brands for which positive responses tend not to be jointly present. The figure shows Ben and Jerry’s and Haagen-Dazs being located near each other, indicating a heightened pattern of competition. Blue Bell and Blue Bunny show a similar (but not as close) association.

Figure 1.4: Consumption Context Map for Ice Cream (S6).

Figure 1.4 is an eigenvector plot for consumption contexts. The first two eigenvalues account for 64% of the total variation of the latent normal distribution ($\lambda_1 = 2.61, \lambda_2 =$
The routine consumption context is plotted apart from those contexts for which ice cream is used to signify a special event. This suggests that individuals use ice cream for either a routine dessert or to signify a special occasion, but not both. There is also some evidence that rewarding oneself with an ice cream treat forms another distinct form of consumption. Further evidence of the presence of this third type of consumption can be validated, in part, by examining eigenvector plots involving additional components.

Figure 1.5: Media Map for Ice Cream (Q12).

![Eigenvector Plot](image)

Figure 1.5 is an eigenvector plot for food and nutrition-related media consumption. The first two eigenvalues account for 58% of the total variation of the latent normal distribution ($\lambda_1 = 2.28, \lambda_2 = 1.20$). Traditional media advertising (TV, newspaper and radio) are consumed similarly by the respondents in the survey, and newer media (web-
sites and social media) are not yet used to acquire information in the Ice Cream product market. The co-location of traditional media suggest that using one of these vehicles is sufficient for message placement.

Figure 1.6: Channel Map for Ice Cream (Q6).

Finally, Figure 1.6 shows that the dominant distribution channels of the traditional supermarket and discount mass merchandise store are not located near each other, implying that respondents either shop for ice cream at one or the other but not both. These two venues therefore reach different respondent groups. The first two eigenvectors used to construct this plot account for 55% of the total variation of the latent normal distribution ($\lambda_1 = 2.41, \lambda_2 = 1.46$).
Summary of Analysis

Our analysis of the variables associated with the Ice Cream market – brand purchase (S8), consumption context (S6), media consumption (Q12) and channel consumption (Q6) – indicates the possible presence of submarkets where demand is heightened among certain brands, possibly in response to different consumption contexts. We find evidence that the Ben and Jerry’s and Haagen-Dazs brands, and the Blue Bell and Blue Bunny brands, are closer substitutes to each other than other brands, and that respondents distinguish between routine and special occasions of ice cream consumption. Analysis also indicates the dominance of traditional media and traditional channels of distribution. Thus, for the purpose of market definition, it might be profitable to concentrate on the brand-context interaction while only considering traditional media and traditional distribution channels.

Florida Vacations

Tables 1.13–1.16 display the proportion of respondents answering “yes” to the Pick any/J questions describing aspects of the Florida Vacation market. The Florida Vacation questionnaire is provided in Appendix A.

<table>
<thead>
<tr>
<th>Table 1.13: Florida Vacation Parks Visited (S9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magic Kingdom</td>
</tr>
<tr>
<td>0.459</td>
</tr>
</tbody>
</table>

Table 1.13 reports the proportion of respondents who have visited each of the theme parks in the past five years. Disney’s Magic Kingdom and Epcot Center have the highest incidence rates, with 45.9% of respondents indicating they have recently been to the Magic Kingdom. Universal Studios and Universal’s Islands of Adventure, which includes the new Wizarding World of Harry Potter, have a lower rate of incidence. Busch Gardens, which is located in Tampa, Florida and not in Orlando, was visited by 18.6% of the respondents in the last five years. Some of the lower brand usage may be due to additional costs of travel.
Table 1.14: Florida Vacation Consumption Contexts (S7)

<table>
<thead>
<tr>
<th>Annual Vacation</th>
<th>Family Reunion</th>
<th>Special Occasion</th>
<th>School Break</th>
<th>Holiday</th>
<th>Just to Get Away</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.496</td>
<td>0.318</td>
<td>0.531</td>
<td>0.238</td>
<td>0.446</td>
<td>0.824</td>
</tr>
</tbody>
</table>

Table 1.14 indicates that, while consumption contexts vary, the dominant reason for vacationing in Florida is “just to get away,” with 82.4% of respondents indicating this as a consumption context. The next highest context reported was 53.1% for special occasions such as a birthday or anniversary. Overall, response rates for all of the consumption contexts are high, implying that all are potentially viable as platforms for repositioning efforts.

Table 1.15: Florida Vacation Travel Media (Q11)

<table>
<thead>
<tr>
<th>TV</th>
<th>Radio</th>
<th>Newspaper</th>
<th>Magazine</th>
<th>Website</th>
<th>Social Media</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.280</td>
<td>0.062</td>
<td>0.171</td>
<td>0.309</td>
<td>0.889</td>
<td>0.316</td>
</tr>
</tbody>
</table>

Table 1.15 reports on how respondents typically get information about travel and vacation packages. Information is obtained through websites for 88.9% of respondents, while only 6.2% of respondents report reliance on the radio. This may be due to complexities associated with purchase and consumption of a theme park vacation.

Suppose you are trying to contact a close friend to see when they will be at a restaurant. You might consider how they tend to communicate and the fact that they may be rushing to pay a cab as they come to meet you. These factors leads you to chose texting them because they can see it quickly, and your message “Are you close? At booth in back” can reach them quickly. Your message, medium, and timing might be quite different if you are calling your elderly grandmother to wish her a happy birthday – a different context, a different need and a different channel of communication. An email to either “customer” would have been a waste of time. (JO)
Table 1.16: Florida Vacation Purchase Channel (Q6)

<table>
<thead>
<tr>
<th>Hotel/Airline Website</th>
<th>Search Engine</th>
<th>Discount Travel Site</th>
<th>Travel Agent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.646</td>
<td>0.587</td>
<td>0.411</td>
<td>0.161</td>
</tr>
</tbody>
</table>

Table 1.16 indicates that nearly all respondents who book travel to a Florida theme park use the Internet, with 64.6% having gone directly to the hotel and airline website, and more than half having used search engine sites such as Travelocity and Orbitz or discount travel sites such as Priceline. Only 16.1% of respondents call a travel agent to make reservations.

Tables 1.13–1.16 indicate that Disney has a dominant market position with its four theme parks, that at some time in the last five years people just “want to get away” and end up in Florida (although more than 50% go to Florida for more specific reasons), the website is the dominant form of learning about the theme parks and few people rely on traditional travel agents to make reservations.

Estimates of the mean (µ) and correlation matrix (R) from our hierarchical model for Pick any/J data are provided in Tables 1.17 and 1.18.

Table 1.17: Estimated Mean (µ) for Florida Vacation Brands (S9)

<table>
<thead>
<tr>
<th>Magic Kingdom</th>
<th>Epcot Kingdom</th>
<th>Animal Kingdom</th>
<th>Hollywood Studios</th>
<th>Universal Studios</th>
<th>Islands of Adventure</th>
<th>Busch Gardens</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.094</td>
<td>-0.220</td>
<td>-0.594</td>
<td>-0.650</td>
<td>-0.588</td>
<td>-1.099</td>
<td>-0.888</td>
</tr>
</tbody>
</table>

The estimate of the means (µ) corresponds with the reported proportion of respondents saying they have visited the park within the past five years as reported in Table 1.13. That is, smaller proportions correspond to smaller (more negative) mean values, and higher proportions correspond to more positive mean values.

The correlation matrix shows a high degree of co-consumption of the brands, particularly within the Disney and Universal trademarks. A respondent attending Magic Kingdom is very likely to attend Epcot and Animal Kingdom, and to a lesser extent Hollywood Studios. The correlations associated with the co-occurrence of theme park visits is often 0.80 and above. A respondent reporting that they attend Universal Studios is similarly inclined to attend the Islands of Adventure, the other Universal park.
1.4. **ANALYSIS OF QUESTIONNAIRE DATA**

Table 1.18: Estimated Correlation Matrix ($R$) for Florida Vacation Brands (S9)

<table>
<thead>
<tr>
<th>Magic Kingdom</th>
<th>Epcot Kingdom</th>
<th>Animal Kingdom</th>
<th>Hollywood Studios</th>
<th>Universal Studios</th>
<th>Islands of Adventure</th>
<th>Busch Gardens</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>0.831</td>
<td>0.804</td>
<td>0.670</td>
<td>0.520</td>
<td>0.609</td>
<td>0.354</td>
</tr>
<tr>
<td>0.831</td>
<td>1.000</td>
<td>0.825</td>
<td>0.713</td>
<td>0.499</td>
<td>0.538</td>
<td>0.403</td>
</tr>
<tr>
<td>0.804</td>
<td>0.825</td>
<td>1.000</td>
<td>0.722</td>
<td>0.463</td>
<td>0.537</td>
<td>0.280</td>
</tr>
<tr>
<td>0.670</td>
<td>0.713</td>
<td>0.722</td>
<td>1.000</td>
<td>0.366</td>
<td>0.519</td>
<td>0.195</td>
</tr>
<tr>
<td>0.520</td>
<td>0.499</td>
<td>0.463</td>
<td>0.366</td>
<td>1.000</td>
<td>0.770</td>
<td>0.436</td>
</tr>
<tr>
<td>0.609</td>
<td>0.538</td>
<td>0.537</td>
<td>0.519</td>
<td>0.770</td>
<td>1.000</td>
<td>0.480</td>
</tr>
<tr>
<td>0.354</td>
<td>0.403</td>
<td>0.280</td>
<td>0.195</td>
<td>0.436</td>
<td>0.480</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Respondents who visit Busch Gardens have a lower correlation with the other theme parks, likely because it is located in Tampa, Florida while the other parks are located in the Orlando area.
Principal component analysis provides a low-dimensional summary of the correlation matrix.

Figure 1.7: Competitive Brand Map for Florida Vacations (S9).

Figure 1.7 indicates a high degree of substitution among the Disney offerings – Magic Kingdom, Epcot, Animal Kingdom and Hollywood Studios – as discussed above. Respondents attending one of these theme parks almost certainly also attend the others. There also appears to be some overlap with the Universal offerings and Busch Gardens, although the nearness of these parks on the map may be diminished when considering higher-order components (e.g., the plot of the second versus the third eigenvectors). The first two eigenvectors used to construct this plot account for 78% of the total variation of the latent normal distribution ($\lambda_1 = 4.38, \lambda_2 = 1.01$). This implies that the princi-
pal component plot accurately represents the latent correlations. The correlation matrix from which Figure 1.7 is constructed is displayed in Table 1.18.

Figure 1.8: Consumption Context Map for Florida Vacations (S7).

Figure 1.8 is an eigenvector plot for the consumption contexts associated with Florida vacations. The first two eigenvectors account for 50% of the variability of the latent normal distribution ($\lambda_1 = 1.77, \lambda_2 = 1.24$). Family reunions and holidays are contexts for which some respondents turn to Florida as the venue for celebration. Similarly, school breaks and annual vacations are plotted near each other, reflecting a possibly different purpose for the Florida theme parks. The contexts “just to get away” and “special occasion” are somewhat separated from the other two groups, and an examination of additional components (e.g., the second and third eigenvectors) would help reveal if
these reasons for going to Florida for vacation are sufficiently unique to warrant different behavioral responses.

Figure 1.9: Media Map for Florida Vacations (Q11).

Figure 1.9 is an eigenvector plot that describes where respondents get information about travel and vacations packages. The overwhelming location, as discussed above, are company websites accessed through the Internet. The first two eigenvectors account for 62% of the variability of the latent normal distribution ($\lambda_1 = 2.45, \lambda_2 = 1.30$). An interesting aspect of this figure is that “website” and “social” are not plotted near each other, despite their common reliance on the Internet. These media channels appear to cater to different respondent groups, offering an opportunity to leverage Internet expertise.
to broaden one's market. Traditional media such as TV and radio are plotted near each other, as are newspapers and magazines.

Figure 1.10: Channel Map for Florida Vacations (Q6).

Figure 1.10 is an eigenvector plot describing the resources people use to make their reservations. The first two eigenvectors account for 73% of the variability of the latent normal distribution ($\lambda_1 = 1.62, \lambda_2 = 1.29$). There appear to be two dominant methods of making reservations: one involving direct access to company websites and another seeking information using third-party search engines that lead to price deals, usually at the cost of not being able to guarantee getting the exact dates or hotel room wanted. The use of travel agents is a third response group that is of relatively minor importance.
Summary of Analysis

A preliminary analysis of variables associated with the Florida Vacation market – theme parks visited (S9), the context of the visit (S7), media used (Q11) and purchase channel (Q6) – indicates a high degree of co-consumption of theme parks organized under the same trademark. Respondents who visit one Disney park tend to visit the other Disney parks, as do those who visit a Universal Studio park. There are varied reasons for visiting the theme parks, with some redundancy between “annual vacation” and “school break,” as well as “holiday” and “reunion.” It may be best to re-define these contexts in a different way, possibly in terms of the visit being child-focused versus adult-focused. Finally, the analysis of media and distribution channels speaks to the importance of the Internet and differences between directed search, third-party search and the use of the Internet for posting comments and evaluation of the parks. For the purpose of market definition, the analysis indicates redundancies of some of the questionnaire items (e.g., holiday and family reunion) and diminished importance of others (e.g., TV and radio).

1.5 Technical Illustration (Optional)

We illustrate the technique of principal components using a simple example of bivariate normal data:

$$z \sim \text{Normal} \left( \mu = \begin{bmatrix} -2 \\ 2 \end{bmatrix}, R = \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix} \right)$$

Two thousand observations are simulated from this distribution and plotted in Figure 1.11. These data are centered at the mean value $\mu$ and show a clockwise tilt corresponding to a correlation of $\rho = 0.5$. Positive correlation indicates a positive association, implying that when one variable is above its mean, the second variable is expected to be above its mean too. Not all points have to adhere to this regularity, but as the correlation coefficient increases so does the regularity of this rule.

The censoring of these latent data are indicated by the dashed lines in the graph. Recall that the observed data are not multivariate normal, but are observed as Pick any/J format where a “1” indicates that the unobserved continuous variable takes on a value greater than zero. Thus, for two variables $x_1$ and $x_2$ there are four possible outcomes that are identified by the regions in the figure. Since $\mu_2 > \mu_1$, there will be many more “1” responses for the second questionnaire item $x_2$ than $x_1$. The values of $\mu$ and $R$ lead to simulated data for which it is rare to observe $x_1 = 1$ and $x_2 = 0$. In fact, none of the 2000 simulated values take on this outcome. There are 58 realizations
of \( x_1 = 1 \) (2.9\%) and 1946 realizations of \( x_2 = 1 \) (97.3\%). A tabulation of the observed (simulated) data is provided in Table 1.19.

**Table 1.19: Observed Data**

<table>
<thead>
<tr>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>54</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1888</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>58</td>
</tr>
</tbody>
</table>
Since the data are distributed multivariate normal, we can replace the plot with a plot of probability density contours that describe what we would expect of the data if we knew the true values of $\mu$ and $R$ (see Figure 1.12). These values will be estimated in our analysis of real data, but for the purpose of illustration assume that these values are known to us. The contour lines indicate the probable location of the data mass that is indicated by each label. Ten percent of the data are expected to be located within the smallest contour, 30% of the data within the next contour, and 99% of the data within the outermost contour of the plot. The sample size of the simulated data is $N = 2000$, and we expect to have 20 observations (1% of 2000) outside the outermost contour. We also expect that there will be some variation in the actual number of these outliers across simulated datasets.

Figure 1.12: Density Contours for Simulated Data
When working with real data, we do not get to observe the values of the simulated data $z$ plotted in blue. All that we get to observe is the number of observations in each region of the plot $x$ as shown in Table 1.19. The process of model estimation, described in Appendix C, leads us to estimates of the mean ($\mu$) and correlation matrix ($R$) that locates the multivariate normal distribution so that the mass of the distribution best matches the observed data. We know from the data that 1888/2000 observations, or 94.4% of the mass of the distribution should lie in the region associated with $x_1 = 0$ and $x_2 = 1$, and that 0% of the mass should lie in the region identified as $x_1 = 1$ and $x_2 = 0$. Model estimation involves finding model parameters that lead to predictions that are in close agreement with the observed data, with predictions based on the implied distribution of the data that comes from the model. In this case, predictions are calculated in terms of the proportion of the mass of the multivariate normal distribution falling within the boundaries used to censor the data.

An eigenvalue analysis of the correlation matrix reveals the following decomposition for $R = QLQ'$:

$$Q = \begin{bmatrix} \sqrt{0.5} & -\sqrt{0.5} \\ \sqrt{0.5} & \sqrt{0.5} \end{bmatrix}, \quad L = \begin{bmatrix} 1.5 & 0 \\ 0 & 0.5 \end{bmatrix}$$

with the first eigenvector $q_1$ is the first column of $Q$, and the second eigenvector $q_2$ is the second column of $Q$. Thus, the first principal component has a slope 1.0, and the second principal component has slope $-1.0$.

The eigenvectors are plotted along with the density contours in Figure 1.4. Recall that the goal of principal component analysis is to find linear combinations of the original variables that maximally explain variation in the original variables while being orthogonal (perpendicular, uncorrelated) with each other. We see that the first eigenvector $q_1$ corresponds to the major axis of the density ellipse, and is the combination of the original variables $z_1$ and $z_2$ that has maximum variance. The second eigenvector $q_2$ corresponds to the minor axis of the ellipse, and is the combination of original variables that produces minimum variance. The proportion of total variance in the $z$ data due to the first principal component is:

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{1.5}{2.0} = 0.75$$

and due to the second principal component is:

$$\frac{\lambda_2}{\lambda_1 + \lambda_2} = \frac{0.5}{2.0} = 0.25$$
Therefore, if it were necessary to simplify the representation of the $z$ data, it would be best to keep the first principal component $\sqrt{5}(z_1 + z_2)$ rather than the second $\sqrt{5}(-z_1 + z_2)$ because it retains more of its original variability.

Figure 1.13: Density Contours and Principal Components

The relationship between the principal components and the original $x$ variables is displayed in Figure 1.5. The axes of the plot are now the principal components, and the points are the variables with coordinates equal to the scaled eigenvectors:

$$Q^* = QL^{1/2} = \begin{bmatrix} 0.87 & -0.50 \\ 0.87 & 0.50 \end{bmatrix}$$

Since the original variables have positive correlation, the coordinate points for $z_1$ and $z_2$ are generally in the same neighborhood. They are also not distinguished by the first
principal component (i.e., they share the same coordinate value of 0.87). It is the second principal component that distinguishes the two points, with \( z_1 \) having a coordinate value of \(-0.50\) on the \( q_2 \) axis, and \( z_2 \) having a value of 0.50. As the correlation between \( z_1 \) and \( z_2 \) increases, the coordinate points converge to each other, and as the correlation coefficient goes to zero the coordinate values are maximally distant from each other.

The plotting characters in Figure 1.5 are scaled by a factor proportional to the mean of the normal distribution, i.e., \( \mu \). The data are simulated with \( \mu_1 = -2 \) and \( \mu_2 = 2 \), implying that there would be more of the observed Pick any/J variables equal to one for \( x_2 \) than \( x_1 \). Through the analysis discussed in this book, we believe it is more useful to focus attention on products, channels and media that are used rather than those that are not used, and our goal is to draw attention to these positive responses by making the plotting characters larger for questionnaire items that are selected more often.

Figure 1.14: Eigenvector Plot

A question often asked of the principal components is “What are they?” Principal
components analysis is a mathematical technique for representing high dimensional data, where the first component is the combination of the original variables ($z$) that has greatest variance, with other components having greatest variance subject to being at right angles, or orthogonal, to earlier ones. They represent a rotation of the coordinate system used to plot the data to identify dimensions that don’t contribute much to explaining anything, and also have a property that they can be used to plot the original variables in the new “space.” Beyond that, principal components do not convey substantive meaning. They do not automatically indicate, for example, the reasons behind why variables are plotted close to each other, such as brands sharing a common trademark or being similarly priced. This information is present in the thoughts and knowledge of the analyst, but is not contained in the correlation matrix from which the components are derived. It is therefore important to remember that the labeling, or naming of the components is an art and not a science.

1.6 Summary

Market definition focuses the attention of the analyst on some aspects of a market while ruling out other aspects. Markets are defined on multiple dimensions, and the purpose of defining a market for a brand is to establish boundaries of analysis. This boundary-setting task is present in even the simplest of market calculations such as computing a brand’s market share. The denominator of the market share fraction contains the competitive brands in the market, not all of the brands. Luxury cars, for example, are not included when calculating the market share of a minivan offering. Within the context of a survey, market definition takes two forms: i) the elimination of variables because of low response rates and ii) the elimination of variables because of their redundancy with other variables in the survey. Our analysis of the Ice Cream and Florida Vacation markets revealed many instances of both that helps us better frame research questions and possible actions.

Detecting low response rates and item redundancies in surveys is simplified with the use of data collected in a Pick any/J format. Response rates are measured as the proportion of respondents with affirmative responses to a question, and item redundancy is detected through an eigenvector analysis of a latent normal distribution. We propose use of a hierarchical model that assumes the Pick any/J data arise as a censored realization of an underlying continuous variable distributed multivariate normal (see Appendix C). The latent multivariate normal distribution separately measures the response rate and the co-incidence of response, with the response rate reflected in the mean of the multivariate normal distribution and co-incidence reflected in the distribution’s correlation
matrix. Without the separation of these two effects, two questionnaire items with high response rates will naturally have a high rate of co-incidence, leading to inflated estimates of redundancy unless some adjustment is made to the observed data. In addition, since Pick any/J data are discrete, some form of modeling is needed to obtain a measure of association such as a correlation coefficient. Binary data are known to deflate estimated correlation coefficients.

The eigenvector plots developed in this chapter provide a way to visualize high dimensional data so that attention can be focused on a subset of variables that are important to a brand. Every brand resides in a market with other brands, some of which compete at a heightened level. Consumption contexts are similarly competitive, as are media and distribution channels. The goal of market definition analysis is to broadly assess a brand’s current position in the marketplace, and to identify aspects of exchange that might be changed to improve a brand’s performance.

A football game can end one of two ways: the home team wins or the away team wins. Many factors work together in leading to this, such as the individual capabilities of players, amount of practice, tactic deployment, morale of the team, and many others. Unfortunately, for an ordinary game-watcher these factors are usually not easy to observe. Everyone can look at the score and understand which team wins the game, but it takes some knowledge and experience to tell specifically what contributed to the victory of the winning team.

This is exactly what marketing researchers are trying to achieve. It is simple to observe the products people buy, but what really matters is understanding the factors that collectively lead to purchases and how these factors are related to each other. (BH)
1.7 Homework

1. Download and install the interactive decision tool (IDT) software on your PC from:

   http://themodellers.com/AnalyticsTool

   The software does not currently run on a Mac, unless you use Bootcamp, Parallels or VMware to emulate a PC.

2. Form working groups of two or three people. You will be working with these groups throughout the term for the purpose of homework and the term project. Try to find balance within your group so that it has both qualitative and quantitative expertise.

3. Select (or be assigned) a brand of ice cream or a Florida vacation destination. Once confirmed by your instructor, this will be your brand for the term. Each brand will be assigned to only one group, so you may be assigned your second or third choice. The final term project will consist of making specific recommendations for re-positioning your brand in the marketplace in response to a recent market development.

4. Conduct research on your brand by visiting company websites and news sites, and become familiar with your brand’s current product, pricing, distribution and advertising initiatives. Be prepared to discuss current challenges and opportunities facing your brand.

5. Take either the Ice Cream or the Florida Vacation survey. The Ice Cream survey is provided at:

   www.interceptsveys.com/Survey1071/default.aspx?RID=Test&VID=1

   The Florida Vacation survey is provided at:

   www.interceptsveys.com/Survey1070/default.aspx?RID=Test&VID=1

   Do not let the screening questions in the beginning of the survey terminate your session too early. If needed, provide false answers to the initial questions so that you are allowed to take the entire survey.
6. Provide descriptive names of the axes for each map in Tab 1 (Market Definition) of the IDT. These names should characterize the placement of items along each axis, and provide a summary of what you learned from the graph. Be prepared to discuss your answers in class. There are three axes that need to be named for each of the graphs: Competitive Brand, Context, Media, Channel and Needs.

7. Each student is to maintain a journal for the class that records their thoughts about the following issues:

   a) What did you learn from this chapter?
   b) Where are you confused?
   c) What do you believe to be true from the class material?
   d) What do you question or not totally believe?
   e) What learning can you apply to your brand?
   f) What next piece of analysis would you conduct?

8. What vignette topic is needed to better explain, illustrate or apply the chapter material? Where would it be located? What would you say?
Chapter 2

Market Segmentation

Understanding the needs of potential customers.

In this chapter we develop a conceptual understanding of market segmentation and examine alternative ways of conducting market segmentation analysis. Keeping with our theme of using decision-theoretic measures, we collect information about the reasons people might buy a product using a Pick any/J question format that asks people to decide if a need is applicable to them or not. These data are analyzed by forming clusters, or segments, of individuals with similar reasons for purchase. Firms can create products of high value by focusing attention on those individuals with a unique set of needs that can be profitably served by the organization. These needs are often not reflected by the average consumer.

2.1 Introduction

No two people are alike. We differ in what we want and why we want it. We differ in the way we react to price changes, in where we shop, and in the meaning that goods and services have in our lives. For some, an evening out at a restaurant with a loved one is a chance to explore the depths of thought and learn new things about a person. For others, it’s a relief from the responsibility of cooking a meal. None of us have ever met another person who is remotely like us.

The fact that people are different from each other and are heterogeneous in their preferences and sensitivities to marketing variables is one of the reasons that marketing exists as a discipline. If it were not for heterogeneity, there would be no reason for the
variety of offerings we see in the marketplace (see Allenby et al. (1998)). There would be no reason for temporary price cuts, or for marketing strategies that attempt to create local monopolies by catering to the needs of a targeted group of people. The fact that people are different is what drives the diversity of offerings that are for sale.

Have you ever received a gift that became that “thing” you use all time, but never thought you needed? Someone close to you, who knows your life, everyday activities, motivations and attitudes knew that “thing” would be just what you would want. They didn’t necessarily know if you already had one, but they really thought about what may add value to your life. This is the most thoughtful way to gift giving, and you will always remember the occasion and the thoughtful person who gave it to you. In much the same way, businesses use market segmentation to create new products that consumers will want to buy, even if they don’t already know it. (BB)

Market segmentation is marketing’s response to heterogeneity. Analysis for market segmentation involves two things:

1. A conceptualization or definition of what is meant by the word “market.”

2. A means of taking this heterogeneous entity and breaking it into a small number of concrete groups, or “segments.”

Market segmentation refers to a broad set of tools applied by organizations to analyze and anticipate the diversity of responses to its actions. This analysis is used to predict the response of customers to a mailing for a direct marketing firm, consumers purchasing patterns in a product category for a retailer, and prospects who may or may not be current consumers for a firm considering a new product introduction. The term “market” is sometimes implicitly defined by the analysis that is conducted, and, unfortunately, often times is not defined at all.

Throughout our discussion, “market” will refer to a set of qualified prospects (potential buyers) whom we wish to understand. The word market is therefore used more narrowly than it was in Chapter 1, focusing on just one of the six dimensions (To whom will we offer it?) that defines the place where business is conducted. Our use of the word “market” allows us to concentrate on understanding the needs of potential customers.

The term “segment” is also not clearly defined in business analysis. Sometimes it refers to groups of individuals defined in terms of their needs, and at other times it is
defined in terms of some aspect of their wants, such as preferences for specific products (e.g., the loyal segment) or sensitivities to specific marketing variables (e.g., the deal-prone segment). While the attractiveness of market segments is often clearly articulated, such as being stable, reachable and profitable, the variables used to define them are often seen as unimportant so long as the segments obtained at the end of the procedure are reachable and profitable. Methodologies for segmenting are deemed to work when they produce good segments (see Fennell and Allenby (2004)).

Our goal in this chapter is to provide more explicit structure to thinking about deriving segments. Our hope is to find relatively homogeneous groups of potential customers with similar needs, if they exist.

Martin Luther King Jr. once said, “The true measure of a man is not how he behaves in moments of comfort and convenience but how he stands at times of controversy and challenge.” Meaning, you don’t know the true values of a person until you’ve witnessed their behavior in extreme circumstances. The same holds true in market segmentation. A marketer can explore the needs and attitudes of typical customers, but doing so will lack the insight needed to create brands of high value unless attention is also paid to atypical individuals.

Statistics such as correlations, means and standard deviations can be unreliable in market segmentation analysis because they provide an average measure of association across everybody in the sample. The extremes at the tails of the heterogeneity distribution often hold much greater promise for marketing analysis because they reflect people with the strongest and truest desires, who are most likely to respond favorably under competitive pressures and stress. (DP)

Why Conduct Market Segmentation Analysis?

A logical question to ask at this point is “Why segment at all?” If there are no two people who are truly alike, then the creation of groups of people who are similar in some way seems like a waste of time. Market segments are created and used by firms for three reasons:

1. For purposes of discussion.

2. Because the range of actions available to a firm is often limited to a few alternatives.
3. Because they provide a means for understanding extremes.

While it is possible for an analyst to understand and characterize a distribution of heterogeneity using mathematical notation, it is often not possible for other individuals to have the same notions about the diversity of response without having some concrete examples available for discussion. Many of these discussions occur in email messages, elevator rides and office hallways, and it is therefore necessary for people to have specific images of respondents in one’s mind to assess the worthiness of proposed actions. When communication is concrete, as it usually is, market segmentation is used to make an abstract, heterogeneous population real and easy to talk about.

Competitive pressures and organizational inertia often limit the range of responses available to management. When this occurs, it is useful to characterize the people most affected by each of the actions under consideration, rather than characterize them apart from the actions. Segments names such as “Brand Loyals” and “Deal Friendly” or possibly “Program Respondents” are used to identify groups of respondents who are expected to be responsive to specific marketing programs. Since the programs being explored are limited in number, so are the groups that are expected to respond.

### Why do we need marketing segments? Why not leave everyone as individuals?

Why not leave everyone as individuals? One reason why we need market segments is the same reason why the U.S. created states. With states, political leaders could easily identify and discuss specific territories within the United States. Also, the federal government did not have the resources or infrastructure to provide services all the way down to the individual tax payer, so the federal government routed tax dollars to the states and allowed them to manage specific services. Furthermore, there were large differences among southern states and northern states, and by creating states, these states were able to represent their unique differences within the federal government. (KP)

Finally, we shall see that market segmentation analysis benefits greatly by understanding and characterizing extreme respondents. Firms are most interested in customers who are driven to the product category by a unique set of needs. These individuals will hopefully be either the most satisfied or least satisfied with particular offerings, or are those who are either most likely to switch brands or those who are least likely to switch. They are often the most or least price sensitive. These extreme respondents are not well characterized by model parameters such as means, covariances and regression coefficients. The creation of market segments facilitates the analysis of extreme respondents who are characterized has being unique in some important way, allowing firms to estab-
lish local monopolies that yield above-average returns by catering to the needs of specific individuals.

In a conversation at work, your boss says to you, “There are two kinds of people in this world: those that are willing to work hard, and those that have no place in the workforce.” On your way home from the office, you call your mother and she says to you, “There are two kinds of people in this world: those that know how to drive and those that think they know how to drive.” Once you get home, you turn on the television and an actor says, “There are two kinds of people in this world: the quick and the dead.” Each distinction is a gross generalization, but useful for illustrating a point of view. (AG)

2.2 Approaches to Market Segmentation

There are two approaches to conducting market segmentation analysis – one upstream from the marketplace and one downstream, or in the marketplace. Originally, market segmentation was developed around the upstream definition, where the goal was to develop an in-depth understanding of the consumer. The central task of upstream analysis is to answer the question “Where is the individual coming from?” This information is useful in guiding firms to make things that people would want to buy by being relevant in the tasks and pursuits of their lives.

Understanding where people come from involves studying aspects of their lives and uncovering the significance of variables that are not directly under the control of marketing. This includes the motivations, perceptions, and attitudes of individuals in the context of their everyday lives, before they come to the marketplace. “Market” is defined here in a broad sense to include individuals who may or may not make purchases in a specific product category. Analysis of upstream markets requires the use of customized surveys and research studies that develop unique variables for analysis.

During the last 30 years, the availability of syndicated and marketplace data has shifted the focus of analysis in marketing to a downstream orientation. Instead of developing an understanding of where people are coming from, market segmentation has moved to an analysis of where people are going to. Bar coding, scanning and data warehousing technologies have made available vast amounts of data describing marketplace transactions, and firms are interested in making good use of these data. As a result, analysis has shifted from studying people in the context of their lives to studying people...
in the context of their purchases, and much progress has been made in developing models that accurately describe the marketplace wants of individuals. Marketplace information is useful for making people want to buy, as opposed to making things that people will want to buy (see Allenby et al. (2002)).

There are advantages and disadvantages in the movement from an upstream to a downstream orientation for analysis. Upstream analysis is more strategic, allowing for the measurement of variables that may or may not be acted upon by management. Downstream analysis is more tactical, measuring a different set of variables that are often immediately actionable.

If you are familiar with finance, you may find this very similar to the two basic approaches of stock valuation: fundamental analysis and technical analysis. Fundamental analysis focuses on the company that issues the stock. Analysts examine the company’s financial strength, growth potential, and management behavior in hope to predict the company’s financial performance in the future. They then derive the value of the stock from the expected stream of profit of the company. This approach is similar to the upstream approach of market segmentation, where the marketing researchers try to predict consumers’ purchasing decisions from their personalities and backgrounds.

Technical analysis, on the other hand, analyzes a stock’s performance in the market. Analysts specializing in this valuation method examine the past performance of a stock, patterns in the movement of the stock price, and attempt to predict its future performance using advanced mathematical tools. This methodology is similar to downstream market segmentation. (BH)

Upstream research allows exploration of constructs such as unmet demand (see Chandukala et al. (2011b)), while downstream research is rooted in terms of current demand. Finally, upstream analysis is capable of identifying and predicting major shifts in demand, while downstream analysis is better adapted at modifying the timing and location of demand that is already present. Our analysis of market segments will take on an upstream orientation. The reason for this is that the downstream analysis of segments is better accomplished within the context of specific marketing variables operating within specific models of downstream actions. We discuss such downstream analysis in Chapters 4-7, which deal with product, pricing and promotional analysis. This chapter concerns itself with upstream analysis based on motivations, or needs, that arise out of the tasks and interests in a person’s life. Thus, the “market” we seek to segment is not based in
2.3. BASIS VARIABLES

the marketplace. Instead, it corresponds to the activities in the context of a person’s life for which a product offering might be relevant.

You might be wondering which approach is better: upstream analysis or downstream analysis. The answer is that both are necessary if you want to perform successful segmentation.

If your focus is upstream analysis only, it may become too theoretical. With your head in the clouds, the competition is going to create the perfect shopping experience, maximize customer purchases, increase referrals, and steal your market share, sending you upstream into hot water with your manager.

If your focus is downstream analysis only, you are ignoring why the customer needs you in the first place. You can have the greatest sales pitch in the world, but if nobody walks in your store, because they do not need your product, you will be downstream without a paddle. (WBJ)

Our definition of “market” includes anyone with the potential willingness and ability to engage in transactions pertaining to the product category. Associated with every product category are one or more activities outside of the marketplace. For blue jeans, the markets may be farming, painting, light yard work, horseback riding, informal wear, and chic designer wear for trendy parties (see Figure 2.1). For automobiles, the activities include commuting to work, driving kids to a soccer game, travel associated with vacations, and hauling supplies. Defining a market in terms of one or more associated activities is useful for understanding where a person is coming from.

2.3 Basis Variables

Firms use a variety of variables to describe prospects for their offerings. Effective variables describe the conditions in which the brand might be useful. It is useful to think of marketplace offerings as instruments that embody benefits to users, which are in some ways matched to the problems these same users face. The description of these problems form what are known as the “basis” variables for marketing segmentation.

Basis variables provide a description of individuals at the lowest level of aggregation. The issues faced by an individual can be thought of in terms of collections of these variables, much in the same way that basis variables in mathematics are used as the building
blocks to represent high-dimensional data. The variables \( \{ x, y \} = \{1, 0\} \) and \( \{0, 1\} \), for example, form the basis for representing any real two-dimensional array. The choice of basis variables for market segmentation therefore determines what can and cannot be explained by our analysis.

Consider, for example, the potential use of demographic variables as a basis variable for market segmentation. Demographics such as age, gender and income describe an individual irrespective of the context of response. If an individual is 40 years old, they are 40 years old at home, in the office, in the gym, while driving, playing golf, and in all activities in which they engage. Demographic variables cannot be used to explain something at a finer level of detail, such as an individual’s desire to be a good parent. All that we know from demographic variables is that a person is a parent, not the ways in which they want to exhibit good parenting.
2.3. BASIS VARIABLES

Figure 2.2: General and Specific Basis Variables

<table>
<thead>
<tr>
<th>General Descriptors</th>
<th>Domain-specific Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>• <strong>Product use</strong></td>
<td>• <strong>Brand choice</strong></td>
</tr>
<tr>
<td>• General Demographics and Psychographics</td>
<td>• Concerns and Interests</td>
</tr>
<tr>
<td>• Domain of activity</td>
<td>• Context for activity</td>
</tr>
<tr>
<td>• Resource allocation</td>
<td>• Resource deployment</td>
</tr>
<tr>
<td>• Useful for defining a market</td>
<td>• Useful for identifying diverse wants within a market</td>
</tr>
<tr>
<td>• Used in general segmentation research</td>
<td>• Used in market segmentation research</td>
</tr>
</tbody>
</table>

Sarah is married, has two kids and has a low income. Every week, she shops at Healthy Organic Foods, a local organic store known for selling organic and specialty foods but at relatively high prices when compared to the nearby conventional supermarkets. Given Sarah’s low income, many might assume that she would never shop at an expensive healthfood store. However, Sarah’s kids are gluten intolerant and suffer from many food allergies, and Sarah wants to make sure her family eats the right foods to stay healthy.

Sarah’s demographics don’t in any way reveal her motivations for shopping at Healthy Organic Foods. Her decision to shop there demonstrates that solely using demographics is insufficient for identifying and defining market segments. In fact, the motivations of a person can never be read from a demographic, making demographics of limited value for understanding why people do what they do. (ND)
CHAPTER 2. MARKET SEGMENTATION

The deficiency of information present in demographic variables cannot be overcome by including additional demographic variables in an analysis. That is, regardless of the amount of demographic data available, it is not possible to infer a person’s needs without conducting some upstream analysis. This view is counter to a commonly held view that with sufficient demographic and purchase history data a firm can engage in one-to-one marketing where needs are inferred (see Fennell et al. (2003)).

By understanding where a person is coming from, successful market segmentation analysis provides guidance in understanding the ways in which a prospect might find an offering useful. It answers the question of why people might prefer one brand over another, as opposed to analysis at the level of the product category. Returning to our earlier example, demographic variables can be used to successfully point to product class participation, but are not successful in predicting which brand a person prefers. Knowing that a person is old and rich, for example, identifies them as prospects for the luxury car market, but provides little help in explaining preferences for specific brands within that market. Figure 2.2 compares the use of general and specific basis variables.

Ideal basis variables provide content-rich descriptions of how people interact with their environment in activities associated with a product category. If the category is shampoo, possible contexts might be used after exercising, before bed, or after being exposed to hair treatments and chemicals. For beer, possible contexts might be in formal or informal social settings including backyard barbeques and dinners with clients. For frozen dinners, the basis variables might include concerns about offering healthy nutrition to a busy family, and their desire to customize the meal to individual tastes.

A useful framework for basis variables is advocated by Fennell in her work (Fennell (1978), Yang et al. (2002), Fennell and Allenby (2006)). She considers seven classes of variables pertaining to motivating conditions associated with the activities people engage in the course of their lives.

Motivations to move away from an undesirable state of being:

1. To solve an immediate problem.
2. To prevent a potential problem.
3. To maintain the current situation, routine.

Motivations to move toward a desirable state of being:

4. To explore interest opportunities.
5. To enjoy sensory pleasure.

Complex motivations from interacting with the marketplace offerings:

7. Ineffective/frustrating outcome.

The first five motivations are simple and the last two are complex in the sense that they reflect the existence of a primary motivation that drives a person to action and a secondary effect coming from the acquisition and use of specific product offerings.

Consider, for example, some of the reasons that people grow flowers from seed, as discussed by Fennell and Allenby (2002). Is the individual struggling to keep some order in the garden that now overwhelms them? Do they grow flowers to keep an altar adorned throughout the year? Do they move in circles where only the latest varieties are admired? Is raising flowers a source of absorbing interest, endless fascination and intellectual challenge? Do they consider the garden a work of art, or is doing the garden as much a matter of routine as doing the dishes?

If differentiating between upstream and downstream analysis is making your head swim, consider thinking about these terms a mother making dinner choices for her family. In upstream analysis, we observe mom in her home environment, before she interacts with our product. Like many busy people, we may find that she’s pressed for time and tired. Perhaps she is worried that she won’t have enough ideas for dinner that week, or that she isn’t taking good care of her family unless she provides them something nutritious and delicious. Or perhaps she wants to avoid something painful – a lengthy preparation time, kids complaining that they don’t like the food, and a lengthy clean up.

In downstream analysis, we follow our mom into the marketplace, at the grocery store, where her needs turn into wants as she makes purchase decisions. Here we look at what drives those decisions. Is she swayed by coupons or price reduction signs that allow her to serve better dinners without high costs? Did advertising emphasizing “Ready to Eat,” or packaging announcing “Less Salt” or “Five Servings of Vegetables” pull her in? Or, has she been won over by the little stand at the grocery store that hands out free samples that let her know that the offering is tasty and delicious? It’s in the downstream analysis that we enter the shopping scene and connect with our customer as she makes purchase decisions. (LG)

Figure 2.3 provides examples of motivating conditions for the two product categories studied in this book – Ice Cream and Florida Vacations. The distinguishing feature of these statements is that they are concretely related to the associated activity and span
<table>
<thead>
<tr>
<th>Motivational Class</th>
<th>Ice Cream</th>
<th>Florida Vacations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Immediate Problem</td>
<td>Ice cream helps me relax and enjoy life. Ice cream is a wholesome treat.</td>
<td>I go on vacation to relax and enjoy life.</td>
</tr>
<tr>
<td></td>
<td>Ice cream gives me something fun to do with family and/or friends.</td>
<td>I want a wholesome vacation experience.</td>
</tr>
<tr>
<td></td>
<td>Ice cream provides relief from regular life.</td>
<td>I go on vacations to have fun with family and/or friends.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vacations provide relief from regular life.</td>
</tr>
<tr>
<td>2. Problem Prevention</td>
<td>Ice cream helps make an event special. Ice cream reminds me of my childhood.</td>
<td>Vacations are where lasting memories are made.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vacations remind me of my childhood.</td>
</tr>
<tr>
<td>3. Routine</td>
<td>I’m not picky – ice cream is ice cream.</td>
<td>When I travel on vacation I don’t have strong feelings about where to go.</td>
</tr>
<tr>
<td>4. Mental Exploration</td>
<td>I am amazed at how ice cream tastes are created.</td>
<td>I am amazed at how the vacation attractions work.</td>
</tr>
<tr>
<td>5. Sensory Stimulation</td>
<td>I enjoy trying new flavors of ice cream. I love having lots of flavors to</td>
<td>I enjoy trying new attractions when I’m on vacation.</td>
</tr>
<tr>
<td></td>
<td>choose from.</td>
<td>I love having a variety of attractions when I’m on vacation.</td>
</tr>
<tr>
<td>6. Costly</td>
<td>Serving ice cream takes too much effort. Ice cream is too expensive.</td>
<td>I would love to go to Florida for vacation, but I just don’t have the time.</td>
</tr>
<tr>
<td></td>
<td>Ice cream contains artificial flavors and ingredients.</td>
<td>I would love to go to Florida for vacation, but it’s not financially possible.</td>
</tr>
<tr>
<td></td>
<td>Ice cream is too fattening.</td>
<td>I would love to go to Florida for vacation, but it’s too planned out and artificial.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I would love to go to Florida for vacation, but I’m worried about my family’s safety.</td>
</tr>
<tr>
<td>7. Ineffective</td>
<td>Ice cream is good for entertaining. Everyone loves ice cream.</td>
<td>Florida vacations help me enjoy my family and friends more.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Everyone loves going on a Florida vacation.</td>
</tr>
</tbody>
</table>
the range of motivating conditions that prompt people to action. The advantage of a framework such as Fennell’s is that it helps ensure that the span of the basis variables is complete. Another distinguishing feature of the statements in Figure 2.3 is that, for classes 1-5, they describe prospects in the context of their everyday lives, not in the context of a shopping mall or the Internet.

Numerous methods have been proposed for developing basis variables for market segmentation. In addition, all firms use some set of adjectives for describing their current and potential customers, which could be considered as basis variables. Often times, the variables used to describe customer segments is a mixture of motivations and descriptions of marketplace and non-marketplace behaviors. The advantage of using non-marketplace descriptors is that they can provide a clearer picture of where customers are coming from that is not tainted by the performance of the products currently available. Describing behavior entirely in terms of people’s reactions to current offerings results in basis variables that are too dependent on existing technologies and current product positioning strategies.

Finally, it is important to mention that the ideal unit-of-analysis in marketing is not the household or a person, but a person engaged in a specific occasion of product use. Needs, or motivating conditions, occur as individuals engage in behaviors related to the potential use of the product. As the context of the behavior changes (e.g., private versus social consumption), the associated needs are also likely to change. It is therefore important to clearly articulate to the respondent the context for which responses are sought. It makes a difference if a respondent is brushing their teeth before they go to bed versus go out on a date, or if the family meal is being prepared in the middle or end of the week. For the Ice Cream and Florida Vacation studies, we explicitly collect information in the context of use so that analysis can be done at the person-context level, e.g., ice cream at birthday parties, or the use of Florida vacation destinations for inter-generational family gatherings with grandparents and grandchildren.
The study of human behavior and the formulation of predictions for future behavior pose a challenge for both psychologists and marketers alike. Similar to clinical psychology, marketing is most interested in the extremes of human behavior, motives, and opinions. This interest in the extremes can then be used to better describe and quantify other individuals that may be similar in traits, needs, and resulting wants. The real challenge occurs when attempting to quantify, or assign appropriate numbers to, human behavior that is naturally qualitative.

The quantitative/qualitative debate can best be described through the changing stages of water. As water changes from ice to liquid water and to steam, it can be measured through a quantitative reported temperature. However, the temperature alone does not tell the entire story. Water goes through qualitative changes as well that are not reflected in the quantitative report. If you needed to explain what ice was to someone, could you describe it thoroughly by just saying it was ‘water that is less than 32F’? Likewise, does describing vapor as ‘water that is greater than 212F’ really capture and describe the qualitative depth of what evaporated water is? This, in essence, is the challenge of explaining complex human behavior, whether in marketing or psychology, through quantitative measurement alone.

In clinical psychology, disorders are studied in both quantitative and qualitative terms and a single construct on a sliding scale can usually describe two distinctive disorders. For example, we all have varying levels of self-esteem. Anyone with a normative level of self-esteem will not seem extraordinary from observations of others. Both clinical psychologists and marketers are not interested in the normative measurements. However, if someone has extremely high or low self-esteem, it will become significant and the individual will be classified as meeting criteria for inclusion as having a disorder. Extremely high self-esteem in this case will present itself as narcissism while extremely low self-esteem will present as depression. This is a form of segmentation that clinical psychologists use. Psychologists will develop prototypes that fit criteria on these sliding scales and study other individuals that report similar features as the initial prototype. In marketing, these prototypes are driven by an individuals needs instead of their psychological status and the prototypical individual defines a segment, or collection of individuals, that share the same needs.

The qualitative descriptions of a segment in addition to the quantitative measurement are what really illustrates a more complete and complex description of customers and grants marketers greater utility. (CR)
2.4 Market Segmentation Analysis

Two factors guide our data collection and segment creation – that people are uniquely different, and that the preferences and sensitivities of an individual are inherently relative. Our approach to the task of understanding the needs and motivating conditions of individuals is to pursue analysis based on Pick any/J data, where we attempt to minimize assumptions about the relative importance of variables. Pick any/J items are either important to an individual or they are not, and our goal in market segmentation analysis is to identify the unique and important motivating conditions associated with a focal group of respondents. Our goal at this stage of analysis is not to provide definitive information on the degree of importance of need-states, but to identify candidate variables for further study.

Figure 2.4 displays the proportion of respondents who indicate the presence of a need, or reason for choosing to purchase ice cream. The responses are for question Q9 in the ice cream survey contained in Appendix A. Responses are displayed according to brand usage as reported in question S8. The figure shows large variation across needs, but relatively small variation within needs. That is, less than 20% of the respondents indicated that need 7 (I’m not picky, ice cream is ice cream) or 8 (I am amazed at how ice cream tastes are created) motivates them to consume ice cream, while about 60% of the respondents responded that need 12 (everyone loves ice cream) applied to them. However, for a specific need state (e.g., need 1 “Ice cream helps me relax and enjoy life”) the users of different brands report about the same incidence of the need. With a few exceptions, (e.g., need 5 “Ice cream helps make an event special,”) brand usage is not well distinguished by computing the incidence of the need for the entire sample.

Figure 2.5 shows a similar pattern in the Florida Vacation data where the aggregate incidence of the needs do not point to relative brand preference. Respondents with needs 1 (“I go on vacation to relax and enjoy life”) and 3 (“I go on vacations to have fun with family and/or friends”) are found in more than 80% of the respondents, while need 7 (“I don’t have strong feelings about where to go”) is found in only 10% of respondents. Responses to the these questions do not depend much on which theme park was recently visited.

The results displayed in figures 2.4 and 2.5 are disappointing. The theme parks and ice cream brands investigated do not appear to be differentiated by any of the individual needs states investigated, except possibly for a few in the ice cream category. It is important to remember, however, that the proportions displayed in these figures are estimates of the mean response probability for all respondents, and do not reflect important aspects of the entire distribution of responses. In particular, the proportions do not provide information about tail behavior of the distributions. It is possible that
the need variables are differentially related to brands for specific subsets or segments of people, while in aggregate there appears to be no differences.

A useful tool for analyzing the tail behavior of a distribution is cluster analysis. Cluster analysis assumes the existence of naturally occurring groups, or clusters of respondents. When applied to data reflecting the presence of need states, it produces groups of individuals who share similar concerns and interests as they engage in a focal activity, such as eating ice cream or going on Florida vacation. Analysis can then be conducted at the group level as opposed to the entire sample of data, thus allowing us to quantify aspects of the distribution other than its grand mean. That is, it allows us to potentially investigate the tail-behavior, or extremes of a distribution of responses.
Cluster Analysis

Cluster analysis is a multivariate technique used to find groupings of individuals. The technique uses an algorithm to maximize both the heterogeneity between clusters and the homogeneity within clusters. The analysis assumes that a natural structure of clusters exists. Since there is little within marketing to support a strict interpretation of this assumption, prior research and market-specific experience serve an important function in helping determine plausible groupings. For our analysis, we are interested in finding groups of respondents with approximately the same set of needs in a way that is not necessarily related to current brand use. This means that we will not be using brand usage information to help form the clusters.
If you’ve ever played with LEGO blocks, or had children who did, then you’ve probably done cluster analysis. You’ve got a box, or a huge mound of parts, and you know it will take twenty minutes of sifting through bricks, brackets, and plates to find just the right one unless you get organized. So you sort all those parts into a manageable number of groups so that you can find parts more easily. What’s the best way to do it? It doesn’t matter much when there actually are distinct groups of similar parts. (MB)

The algorithm uses a measure of the distance between respondents to generate clusters. The most common metric used in optimizing cluster differentiation is the Euclidean distance. Also known as the straight-line distance, it is simply the sum of the squared differences between respondents calculated on the variables under study. Other distance measures are used, such as the absolute distance that replaces the squared differences with the sum of the absolute differences, but the Euclidean distance is often preferable.

Whatever nearness metric is used, it relies entirely on the basis variables chosen for the technique. This makes the selection of the basis variables critical to the outcome of the analysis and adds another layer of subjectivity in the development and selection of a solution. To cluster on needs independent of brand use, we use the Pick any/J data that maps to the motivating conditions as described in Figure 2.3. To ensure optimal differentiation between groups, we need to avoid using highly correlated variables within our battery of inputs. If highly correlated variables are retained, they serve to over-weight certain responses and can skew the clustering results.

To check for correlation and reduce the battery of basis variables into its own component parts, we run a principal component analysis on the variables using the model for Pick any/J described in Chapter 1. We then take the variables that are representative of each component to include as our basis variables. If each variable exists as a separate component, the basis variables are sufficiently differentiated and no data reduction is required before proceeding with the cluster analysis. But, if some of the basis variables are highly correlated, it makes more sense to use just one of the variables in the analysis as being representative of the others. Many people use a technique known as factor analysis to provide an initial screening of the variables for inclusion as basis variables. Factor analysis is similar to principal component analysis, and produces identical results for the analysis reported below.

The end result of cluster analysis is a set of respondent groupings differentiated by their needs as defined in our basis variables, but the algorithms used to get there are as varied as the potential solutions. Three algorithms are of particular use, both separately and together: hierarchical clustering, K-means clustering and Ward’s method.
2.4. MARKET SEGMENTATION ANALYSIS

- Hierarchical clustering uses an agglomerating process to generate clusters. Each individual starts the process in his or her own cluster, and at each stage in the algorithm similar clusters, as determined by the distance measured between their respective centroids that start as individual respondents, are combined to form a new cluster. This continues until the specified number of clusters is reached. A possible hierarchical algorithm is as follows:

  1. Compute the cluster centers by taking the mean of each basis variable for each respondent included in the cluster. To begin the algorithm we assume each respondent is their own cluster.
  2. Compute the distance between each cluster center and all other cluster centers.
  3. Combine the clusters that have the smallest distance.
  4. Repeat until the desired number of clusters is reached.

- K-means clustering starts with a number of seed points generated at random or specified by the user. A common algorithm for K-means clustering is:

  1. Specify the number of clusters desired and select a starting point for each cluster. As a general rule these seed points should be well spread out and reside within the mass of the data. These seed points will be considered the starting cluster centers.
  2. Calculate the distance between each cluster center and each point.
  3. Assign each point to the cluster with the smallest distance.
  4. Compute a new cluster center by taking the mean of the points included in the cluster.
  5. Repeat the above steps until either points don’t change clusters or an arbitrary stopping point is reached.

Since the final cluster solutions can depend on the starting cluster centers it is often desirable to re-run the algorithm a number of time with different starting points to be sure that the clusters are stable.

- Ward’s method is similar to the above methods except for the criteria used to form new clusters. Ward’s method evaluates potential mergers of clusters in terms of the variance of responses within clusters. This differs from the distance-based measure used in the above methods in that the criteria is a squared measure instead of linear one. Ward’s method is often criticized in that it produces clusters that are “striped,” where the resulting clusters include too many respondents.
Moving day is exhausting, but unpacking your belongings can be a nightmare, especially in the kitchen. Who knew that just one room would house such a heterogeneous group of items: pots, pans, plates, cups, bowls, toaster, coffee-maker, baking utensils, blender, cleaning supplies, and silverware (just to name a few). You may not know it at the time, but as you are pulling all of these items out of the many boxes on your new kitchen floor, you are performing a cluster analysis. You find yourself putting the plates, bowls and cups in the same cabinet; the pots, pans and baking trays in another cabinet; and the silverware and baking utensils in a single drawer. By distinctly grouping your kitchen items into more homogeneous groups, you are making it much easier to find exactly what you need, when you need it. Similarly, by organizing a heterogeneous population of people into more homogeneous clusters, you make it much easier to target exactly what they need. And if you want to keep it really simple: pack your kitchen items based on clusters rather than just throwing things in a box with leftover space!

Each clustering method has its advantages and disadvantages that are not exclusive to each. One common technique is to use hierarchical clustering to define the cluster starting points for a K-means clustering run. This has the advantage of making the starting seed points less arbitrary, and can lead to faster convergence of the K-means clustering algorithm. This approach is often called a two-stage approach since it involves first using hierarchical clustering and then K-Means clustering.

The interactive decision tool (IDT) for the class contains clustering results using two-stage, K-means and Ward’s methods for solutions of four to seven clusters. It is important to remember that determining a good clustering solution is not a precise science, and it is beneficial to compare multiple solutions before settling on a final alternative.

Ice Cream

The principal component analysis for the Ice Cream study produced four representative variables from the original set (Q9) that correspond to the following needs:

1. Ice cream helps me relax and enjoy life.
2. Ice cream helps make an event special.
3. I’m not picky – ice cream is ice cream.
9. I enjoy trying new flavors of ice cream.

While only these four are included as basis variables, we profile based on the entire battery of needs questions.

Figure 2.6 reports the results of clustering the Pick any/J need variables using our two solutions. Reported is the proportion of respondents indicating that the need applies to them or others in their household. The proportional size of each cluster is indicated at the bottom of the figure.

With these results it is easier to look for striping, where we have a single cluster that is either all high or all low across profiling variables. Striping indicates either a problem with how the basis variables were selected, or simply the absence of clear clusters in the data. Striping does not appear to be an issue in the solutions reported in the figure. We find that cluster 2 in the Solution A results includes nearly half of all respondents, while in Solution B the cluster sizes are more evenly distributed. Additional information about striping and how to deal with it is provided in the IDT tutorial for Chapter 2.

The means of the profiling variable within each cluster provides a snapshot of how each cluster is differentiated. The variables that most differentiate the groupings from our two methods are provided in Table 2.1.
Table 2.1: Ice Cream Differentiation Variables

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Solution A Differentiation</th>
<th>Solution B Differentiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>None of the above –</td>
<td>5. Makes event special.</td>
</tr>
<tr>
<td>3</td>
<td>5. Makes event special.</td>
<td>1. Helps me relax.</td>
</tr>
<tr>
<td>4</td>
<td>9. Enjoy trying new flavors.</td>
<td>None of the above –</td>
</tr>
<tr>
<td>5</td>
<td>7. Ice cream is ice cream.</td>
<td>7. Ice cream is ice cream.</td>
</tr>
</tbody>
</table>

Florida Vacations

The principal component analysis for the Florida Vacation study also produced four representative variables:

4. Vacations provide relief from regular life.

6. Vacations remind me of my childhood.

7. When I travel on vacation, I don’t have strong feelings about where to go.

11. Florida vacations help me enjoy my family and friends more.

It is interesting to note that the basis variables for both the Ice Cream and Florida Vacations datasets do not have duplicate items from the same motivational class (see Figure 2.3). It appears that respondents do see differences among the market segmentation basis variables across classes, and that differences within a class are less pronounced.

Figure 2.7 shows that we have issues with striping in cluster 1, where the means are high for every variable. We also have uneven cluster sizes in Solution A, but Solution B provides clusters of more equal sizes. This indicates that Cluster Solution A is not a good solution for identifying a set of unique needs to drives a particular group of respondents.

Table 2.2 shows that the Solution A procedure has created two groupings that are both differentiated by the same profiling variable. The None of the above cluster exists in both solutions, and is the reason for the striping in the two-stage method. Again we see that the hierarchical process has informed the K-means to create clusters that are each differentiated by different variables, which correspond to our basis variables.

The preceding analysis illustrates that results obtained from a cluster analysis need to be thoughtfully considered. The clusters allow exploration of the tail-behavior the
distribution of responses, and there does not exists an automatic method for their exploration. In contrast to standard statistical analysis involving parameters and coefficients, there do not exists a natural, model-based metric for the many possible extremes of a high-dimensional distribution. It is therefore necessary to explore multiple cluster solutions to identify the extremes most meaningful and useful to a firm.

Table 2.2: Florida Vacations Differentiation Variables

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Solution A Differentiation</th>
<th>Solution B Differentiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6. Reminds me of my childhood</td>
<td>6. Reminds me of my childhood</td>
</tr>
<tr>
<td>2</td>
<td>4. Relief from life.</td>
<td>4. Relief from life.</td>
</tr>
<tr>
<td>3</td>
<td>11. Helps me enjoy family more.</td>
<td>11. Helps me enjoy family more.</td>
</tr>
<tr>
<td>4</td>
<td>11. Helps me enjoy family more.</td>
<td>7. No strong preferences.</td>
</tr>
<tr>
<td>5</td>
<td>– None of the above –</td>
<td>– None of the above –</td>
</tr>
</tbody>
</table>
Clusters are clusters ... or are they? There could be a considerable difference among clusters depending on how they are formed if there really aren’t a small number of distinct groups to begin with. Confused? Think of it like a magnet. Where you place a magnet affects how metal shavings or objects get pulled or grouped together. If the metal objects are evenly scattered, you can place a number of magnets in different parts of the field and collect different clusters of objects.

If the metal objects were not evenly scattered on your field, they might gather differently. If your objective was to gather the most objects using the least number of magnets, you would look for the spots with the highest concentration of metal and place your magnets in the center of each, gathering up as much as you could with each placement. That is a lot like a two-stage method of clustering. Using a K-means clustering system with metal shavings could look less like the magnet below on the left and a lot more like changing the placement of hair on our friend Harriet. (EW)
2.5 Summary

The goal of market segmentation analysis is to understand the needs and attitudes that drive people to the marketplace in search of help to affect change in their lives. It attempts to provide insight into “where people are coming from” by relating measures of needs to measures of brand preference. It does this in an exploratory way, leaving the definitive measurement of the effects of needs on preferences and utility for later study in Chapters 4 and 5. Just as market definition analysis in Chapter 1 defined competition in terms of brands, consumption contexts, channels and media, market segmentation analysis refines analysis by identifying the needs associated with brand usage.

We explore the use of cluster analysis for the identification of naturally occurring groups of respondents that have relatively homogeneous needs. The assumption of respondent homogeneity does not have strict empirical support in marketing, and it is typically made so that conversations can take place with concrete representations in mind. The assumption of discrete groups also facilitates discussions about intended actions and whom they affect, and the exploration of sub-groups, or clusters of individuals who are relatively similar to each other. These individuals can be thought of as populating the tails of the distribution of heterogeneity. Since there is no natural, model-based approach for identifying extreme clusters, researchers need to explore and assess multiple potential solutions before settling on a segmentation scheme.

Market segmentation is a task involving the selection of respondents and variables for further study. Once a group or groups of promising respondents, i.e., prospects, are identified, the next step in an analysis is to develop aspects of an offering for them, such as product attributes and communication themes. Selection criteria such as a segment being of sufficient size to be profitable, and stable enough to be sustainable, are sometimes offered. These steps are examined in more detail in the following chapters.
What’s in a name? As it turns out, Shakespeare might not have been entirely correct in his famous quote about a rose. While it is true that a name does not necessarily define an object, names can help us clarify, classify and relate to an object. This can be very useful when we describe market segments to colleagues, but names can also lead to confusion. Take for example this Two-Stage 4-cluster solution for Florida vacations:

| Q8.1 | I go on vacation to relax and enjoy life. | 0.86 | 0.82 | 0.88 | 0.78 |
| Q8.2 | I want a wholesome vacation experience. | 0.54 | 0.34 | 0.47 | 0.52 |
| Q8.3 | I go on vacations to have fun with family and/or friends. | 0.90 | 0.73 | 0.94 | 0.79 |
| Q8.4 | Vacations provide relief from regular life. | 0.80 | 0.65 | 0.64 | 0.76 |
| Q8.5 | Vacations are where lasting memories are made. | 0.87 | 0.56 | 0.68 | 0.61 |
| Q8.6 | Vacations remind me of my childhood. | 1.00 | 0.00 | 0.00 | 0.33 |
| Q8.7 | When I travel on vacation, I don’t have strong feelings about where to go. | 0.00 | 0.00 | 0.00 | 1.00 |
| Q8.8 | I am amazed at how the vacation attractions work. | 0.32 | 0.13 | 0.16 | 0.25 |
| Q8.9 | I enjoy trying new attractions when I’m on vacation. | 0.73 | 0.51 | 0.64 | 0.64 |
| Q8.10 | I love having a variety of attractions when I’m on vacation. | 0.69 | 0.51 | 0.63 | 0.55 |
| Q8.11 | Florida vacations help me enjoy my family and friends more. | 0.41 | 0.00 | 1.00 | 0.33 |
| Q8.12 | Everyone loves going on a Florida vacation. | 0.40 | 0.19 | 0.63 | 0.31 |

Cluster 3 could be defined as someone who goes on vacation to relax and enjoy life, and to be with family and friends. Based on this, we might give Cluster 3 the name of “Fun-time Fran” – a perfectly fitting name. Unfortunately, names elicit certain descriptors and stereotypes that exist solely in the mind of the beholder. If you heard the name “Fun-time Fran,” but did not have access to the data, which of these images would come to mind?

As we have discussed in this chapter, no two people are alike. This is true, not only for where they come from, but also for their interpretation of names. One person may interpret “Fun-time Fran” as someone who wants to enjoy the Florida weather and attractions with family and friends; another person may think Fun-time Fran is a party girl who wants to stay out late and hit the clubs. Its easy to see that these interpretations can lead to serious confusion about potential target segments and, more worrisome, positioning and marketing message misses.

To name or not to name? Go ahead and name away! It helps everyone visualize the segment. Just be sure to keep the conversation focused on what the data says and not just what the names might otherwise imply. (MP and KT)
2.6 Homework

1. What demographic variables are most relevant to your brand? What context, or situational variables are most relevant?

2. Propose an additional basis variable (i.e., need) for your brand using Figure 2.3 as a guide.

3. For the demographic, situational and need variables identified in Q1 and Q2, comment on their ability to describe relative brand preference. To which category do these variables belong? (see Figure 2.2):
   a) Domain (i.e., broad area) versus context (specific location) of an activity.
   b) Resource allocation (i.e., amount) versus deployment (to bring into action).
   c) Market definition (i.e., product category) versus market segmentation (brand).

4. Study at least one K-means and one 2-Stage cluster solutions in the IDT for your brand, and be prepared to explain why one solution is preferred to the other. Provide names for each of the clusters in your preferred solution, and discuss how you arrived at the names you selected.

5. Identify one or more of the segments as target segments that you believe hold promise for repositioning your brand (e.g., by changing product features, price or advertising) and increasing sales.

6. Continue your journal for the class by recording your thoughts about the following issues:
   a) What did you learn from this chapter?
   b) Where are you confused?
   c) What do you believe to be true from the class material?
   d) What do you question or not totally believe?
   e) What learning can you apply to your brand?
   f) What next piece of analysis would you conduct?

7. What vignette topic is needed to better explain, illustrate or apply the chapter material? Where would it be located? What would you say?
Chapter 3

Customer Satisfaction

Identifying potential drivers of brand value.

Driver analysis seeks to identify attributes and benefits of the brand that lead to satisfaction and repeated brand purchase. In this chapter, we examine alternative methods of analyzing customer satisfaction data and the underlying assumptions that distinguish the methods. We find that data collected on standard 7-point scales is not very informative when analysis is conducted on all respondents. Analysis improves dramatically, however, when it is restricted to a particular segment of respondents. This chapter sets the stage for new analysis and methods that are discussed in Chapters 4 and 5 where drivers of brand value are successfully identified at the individual, as opposed to the segment-level.

3.1 Introduction

Branded offerings embody a large collection of attributes and benefits that result from their consumption. Some are related to the physical formulation of the offering, such as the flavor of an ice cream or the rides in an amusement park, and some are related to psychological attributes created through package design, price levels, advertising messages and channel choice. The number of attributes a brand has is typically very large, and preliminary analysis is needed to identify which of the many attributes and benefits may influence, or drive, customer satisfaction. The goal of customer satisfaction analysis is to provide an initial screening of the attributes and benefits for further study. A more
formal quantification of effects of attributes and benefits on brand preference is developed in Chapter 4 in our discussion of Product Analysis.

Assessing the level of customer satisfaction with an offering and the likelihood of customers choosing to continue to do business with a firm requires three pieces of information:

1. Beliefs about the performance of a brand.
2. The importance of these beliefs.
3. A measure of behavior, or intended behavior, that indicates how these beliefs lead to action.

If a brand provides benefits to users on important dimensions, then the likelihood of continued business is large. However, if a customer feels that a brand performs poorly on important dimensions, then the likelihood of continued business is small. Our analysis of customer satisfaction data allows us to make inferences about which beliefs are important by relating behavioral measures to brand beliefs. This requires us to first consider how to collect brand belief information beyond the simple Pick any/J data format used in previous chapters. In doing so, however, we must first examine what information can be obtained in more complicated scale formats.

Some people may believe that electric cars are better for the environment than gas-powered cars, but if they don’t care about the environment then they may not be satisfied with an electric car. Other people may care deeply about the environment, but if they don’t believe that electric cars are better for the environment, then they, too, would not be satisfied with an electric car. In order to be fully satisfied with a product, people must believe that the product provides a specific benefit and that benefit must be important to them. (RL)

3.2 Measurement Scales

Brand beliefs and behavioral measures are often obtained with surveys that ask respondents to evaluate aspects of an offering, and to indicate their expected, future behavior. Behavioral measures include the likelihood of continued purchase, the likelihood of recommending the product to a friend or possibly an overall measure of satisfaction with
3.2. MEASUREMENT SCALES

the brand. Brand beliefs pertain to specific benefits provided by the branded offerings. Offerings are viewed as instruments embodying collections of benefits that consumers can purchase for the right to use. Toothpastes, for example, embody a variety of medical benefits that help prevent cavities, promote good oral hygiene, provide breath freshening, remove stains from teeth and, if properly formulated, can be sensitive to the taste preferences of children.

Experience has shown that directly questioning respondents about the importance of benefits is unreliable, and that it is often better to infer the importance of benefits through the use of statistical models. When directly asked, respondents have a tendency to claim that everything is important, and to not accurately reflect their behavior in the marketplace. Respondents may indicate that price is important to them and yet have a tendency to ignore temporary price discounts. Or, they may claim that a specific benefit is important and yet purchase brands in which that attribute is absent. Thus in many instances it is necessary to estimate the importance of brand characteristics using a statistical model.

The simplest statistical model for customer satisfaction and loyalty analysis is the regression model:

\[ y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_k x_{k,i} + \epsilon_i \]

where “\( y \)” is the overall measure, “\( x_k \)” is the brand belief about attribute \( k \) of the brand, and “\( \beta_k \)” is the importance of the \( k^{th} \) attribute. The index “\( i \)” indexes the respondent, with data from \( i = 1, \cdots, N \) respondents collected in the survey. The estimated regression coefficients (\( \beta \)) indicate the combination of beliefs that best explain the overall satisfaction or loyalty measure that is focal to the study. A coefficient estimated to be zero indicates that the attribute plays no role in explaining variation in the overall measure (\( y \)). Non-zero coefficients indicate the expected change in the overall measure for changes in the brand belief (\( x \)). When the brand belief data are collected on the same scale (e.g., a 7-point rating scale), the \( \beta \) coefficients indicate the importance of the brand belief in “driving” the overall measure. For this reason, analysis using the regression model above is often referred to as “driver analysis.”

Using the regression model for driver analysis requires many assumptions that need to be verified. If any of these assumptions are not true then regression analysis is not appropriate and may lead to incorrect inference. The assumptions are:

1. The data provided by the respondents (\( i \)) all follow the relationship expressed by the regression model. If the regression model expresses a linear relationship between brand beliefs and overall satisfaction, then this relationship is valid for all respondents.
2. The brand belief data \((x)\) are commonly scaled across respondents \((i)\) - that is, values of \(x\) have common meaning.

3. The brand belief data \((x)\) arise from a process that is unrelated to the parameters in the regression model - i.e., they are independently determined. They reflect a process of overall satisfaction formation, not reflection as is present in “haloed” responses. The presence of haloed responses requires more sophisticated models than those discussed in this book.

4. The brand belief data \((x)\) have at least interval scale properties so that multiplication by the regression coefficient has meaning.

5. The dependent variable \((y_i)\) is constructed by adding together the component measures.

6. The error term \((\epsilon_i)\) describing the variation of the observed summary measure from the predicted value follows a normal distribution.

The first three assumptions pertain to a property of models known as “exchange-ability.” Exchangeability means that the respondent index \((i)\) does not, in itself, provide information that would help predict the overall measure. This assumption is often violated, for example, when conducting multi-national customer satisfaction surveys where respondents in some parts of the world report their beliefs in a conservative manner, where in other parts of the world responses tend to be more optimistic, or possibly more polarized and more extreme. Here, the respondent index \((i)\) is related to likely values of \(x\) and \(y\). Another example occurs when poor performing units (e.g., service providers, branch offices, sales territories) are characterized by a different relationship \((\beta)\) than those that perform well. In this case, the index \((i)\) is related to the expected values of \(y\) given the brand beliefs. A final example is when respondents report improved (higher valued) brand beliefs for brands that they currently own, possibly because they are more familiar with these brands. In this case, the overall measure of performance drives the brand beliefs, not the other way around.

Assumptions 4-6 pertain to the nature of the observed data, which are either ratio, interval, ordinal or nominally scaled. Nominally scaled data indicate the presence or absence of an event, and do not indicate any ordering among the levels of a variable. The numbers of football jerseys, for example, are nominally scaled and their values cannot be used to indicate anything other than the identity of a player and possibly the position they play in the game (e.g., low number for a quarterback, high number for a lineman in American football).
3.2. MEASUREMENT SCALES

Ordinal data indicate a rank order with ranks that do not necessarily indicate a uniform progression of importance, relevance or belief. Ordinal scales are often used to collect marketing data in terms of a fixed number of categories as shown in Figure 3.1.

A violation of assumptions occurs when survey responses are “haloed.” Let’s say you are a raving fan of the brand Apple. You think the design of Apple’s products is far superior to any of its competitors in the market. When you’re asked to take a survey about different brands of tablets, you tend to rate Apple much higher than the other brands on attributes like price, performance, quality of applications and ease of use, without really separating these attributes in your mind. Every time you’re asked to evaluate Apple’s iPad, it’s almost an automatic response to rate highly.

This presents a problem for driver analysis. Based on your responses, there is no way to analyze which of the different beliefs about Apple’s iPad are truly driving your satisfaction with the brand or your decision to purchase: your belief about Apple’s product design has haloed your other responses. When this occurs, regression analysis should not be used to conduct driver analysis because it is not a good representation of how the data arise. (TT)

Interval-scaled data, like temperature measured on a Fahrenheit scale, has the property of being able to identify ranks in addition to the property that movement along the scale has uniform meaning. Thus, the movement from 10 degrees to 20 degrees is the same physical difference as the movement from 70 degrees to 80 degrees. With an interval scale, we not only know that an 80 degree temperature is greater than a 70 degree temperature, but we also know how much hotter it is.

Finally, ratio-scaled data is characterized as having a natural zero-point. Time and money are examples of ratio-scaled data, where t=0 might mean “now” and m=0 indicates that the balance of funds is zero. With ratio-scaled data, the t=20 is twice as long as t=10, and m=20 indicates twice the wealth of the outcome m=10. Thus, with ratio-scaled data, ratios of the variable values have meaning.
[BASE: ALL RESPONDENTS]

Q11. From what you have heard or experienced, please indicate how well each statement describes [INSERT BRAND FROM S7].

[SHOW ONE BRAND AT A TIME FROM THOSE SELECTED IN S6– RANDOMIZE STATEMENTS FOR EACH NEW SURVEY BUT KEEP IN SAME ORDER FOR EACH RESPONDENT.]

<table>
<thead>
<tr>
<th></th>
<th>Does not Describe at all</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Relaxing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Wholesome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Fun</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Exciting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Premium – uses better quality ingredients</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Memorable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Buy as a special treat but not regularly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Good enough to keep on hand for regular consumption</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Interesting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Tastes better than most brands</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Offers a wide variety of flavors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Enjoyable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Best value for the price</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Natural/organic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Low calorie</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Great for the whole family</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Great for guests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Nominal, ordinal and interval scales reveal an increasing amount of information about the object being measured. The nominal scale is the simplest, using frequencies to summarize its information. A survey of food items might ask the question:

| Which of the following items do you buy for your family at least once a week? |
|-----------------------------|-----------------------------|-----------------------------|
| Milk                       | Snack Bars                  | Frozen Meals                |
| Cereal                     | Drink Boxes                 | Fresh Fruit                 |

Ordinal scales reveal rankings of attributes and brands:

<table>
<thead>
<tr>
<th>Please indicate your order of preference:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order Preference</td>
</tr>
<tr>
<td>1 Frozen Dinners</td>
</tr>
<tr>
<td>2 Frozed Sides</td>
</tr>
<tr>
<td>3 Frozen Breakfasts</td>
</tr>
<tr>
<td>4 Frozen Snacks</td>
</tr>
<tr>
<td>5 Frozed Deserts</td>
</tr>
</tbody>
</table>

Interval scales assume equal units of measurement along a common scale:

<table>
<thead>
<tr>
<th>Please indicate your views on Frozen Meals using the following scale. (1=Strongly Disagree; 5=Strongly Agree)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convenient</td>
</tr>
<tr>
<td>Cost Effective</td>
</tr>
<tr>
<td>Healthy</td>
</tr>
<tr>
<td>Delicious</td>
</tr>
<tr>
<td>Well Balanced</td>
</tr>
<tr>
<td>1 2 3 4 5</td>
</tr>
</tbody>
</table>

Of course, there is no guarantee that respondents understand or use scales properly. A challenge in analyzing data collected on ratings scales is verifying that moving from a “1” and “2” represents the same improvement in one’s view as moving from “4” to “5.” This is often more of a hope than a reality. (LG)

Having at least interval-scale properties is needed for the algebraic product of brand belief data and coefficients to have meaning. This does not occur when brand beliefs are expressed ordinally - that is, when brand belief data provide only rank information about
the attribute. When data are ordinal and lack the property of constancy across levels, then the multiplication and addition of terms will not lead to a meaningful predictive summary measure, \( \hat{y} \). For example, if all that is revealed is the fact that \( x_i > x_j \), then multiplication by \( (\beta > 0) \) results in the same set of ranks and conveys no additional information.

Additional model structure is needed for nominal and ordinal data to be meaningfully analyzed with coefficients that identify strong and weak relationships among the data. An example is our hierarchical model for analyzing the nominal Pick any/J data in Chapters 1 and 2, where principal components and discriminant analysis were applied to the latent data \( (z) \), not the observed data \( (x) \). One form of analysis is to use nominally scaled (i.e., dummy) variables \( (x = \{0, 1\}) \) to represent the presence or absence of an effect or condition. Another approach assumes that the observed rank data arise from latent interval-scaled variables associated with cutoffs that result in censored observations that are ordinally scaled. This latter form of analysis is particularly useful for conducting analysis on fixed-point rating scales, and is referred to as “cut-point models.” Table 3.1 displays different forms of analysis associated with the assumptions about scale and exchangeability of the data.

<table>
<thead>
<tr>
<th>Table 3.1: Models for Customer Satisfaction and Loyalty Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio/Interval Scaled Data</td>
</tr>
<tr>
<td>Data are Exchangeable</td>
</tr>
<tr>
<td>Data are not Exchangeable</td>
</tr>
</tbody>
</table>

The importance of correcting for scale-use tendencies of consumers can be seen in Figure 3.2, where the responses for two questions are plotted. If some individuals tend to use the upper end of the scale and others tend to use the lower end of the scale, then analysis across respondents will show strong positive covariance due to the effects of scale usage. Interpreting the strong covariance as a strong driver influence among the variables would then be incorrect because some of the covariance is due to scale-use tendencies. The extent to which scale usage confounds driver analysis can be determined by comparing results from the cut-point model to that obtained from simpler models.
Making decisions based on marketing data can be very risky – like walking on a tightrope. You want to take all the precautions you can to ensure that the rope you are about to step onto is securely mounted, that the fibers are uncompromised and that the rope is taut. The rope may look perfectly safe and secure but you aren’t going to take that first step until you have personally checked to make sure everything is in order.

Yet when we make decisions based on survey data we are often all too willing to step onto a tightrope that we haven’t inspected. This tightrope may be anchored on inaccurate data – data that has not been corrected for scale bias. The ends of the rope may be anchored by the fact that some people tend to use the low end of the scale, some people tend to use the high end of the scale and some people may even tend to straight-line at the high or low end of the scale (i.e. select all 1’s or all 7’s).

If we do not take the time to inspect our data tightrope, we might get lucky. The decisions we make might not be a complete failure and they might actually work. But, there’s also the chance that they will fail - that the scale bias within the data is just enough to make the tightrope fall. We can’t blindly trust that the data we are given is infallible - we might just be working without a net. (MP)

Evidence of scale-use patterns are seen in both the Ice Cream and Florida Vacations data. Figure 3.3 plots the median and range of responses to questions Q1 through Q3 in the Ice Cream survey that relate to overall customer satisfaction. Question Q1 is an overall satisfaction measure, Q2 measures repeat purchase intention and Q3 measures the respondent’s likelihood of recommendation of a brand to a friend. All three questions are asked for each of the seven brands in the analysis. Each respondent is displayed as a point in the plot, with their median response plotted against the range of their responses over all three questions and seven brands per question. The points in the plot are slightly “jittered” so that the relative number of respondents associated with each plot value can be determined.

Respondents plotted in the lower right portion of the plot have a median response of seven, and a range of zero. This means that all of their responses are a seven – the highest level of satisfaction for each of the questions. The cluster of points to the left of this corresponds to respondents who give a score of six to every question. Respondents who have a range of zero are said to straight-line their responses, which provides no information about relative brand preferences. Straight-lining is a common problem
encountered in the analysis of customer satisfaction data, representing an extreme form of scale-use bias.

We find that respondents report they are generally satisfied with the ice cream brands, with few responses to the overall satisfaction questions below the average scale value of four on the 7-point scale. It is doubtful however, that these responses truly reflect the preferences of consumers who demonstrate brand switching in the marketplace. In addition, most respondents are seen to use a small portion of the scale, with nearly everyone using less than half of the seven points available to them. One of the reasons for the popularity of discrete choice experiments studied in Chapter 4 is that it forces greater discrimination in responses than those exhibited in the plot.

A plot of data from the Florida Vacations survey (Q1-Q3) shows a similar pattern of responses in Figure 3.4. The clustering of responses at the upper end of the scale
suggests that a simple analysis of the data using regression analysis might be relatively uninformative about the satisfaction drivers. We conclude from these plots that we must pay careful attention to how customer satisfaction data is modeled.

My family is Italian. Deciding on a restaurant on a Saturday night requires the same amount of careful planning, deliberation, and delicate mix of coercion and bribery as Napoleon required to map out his battle strategies. Victoria loves Carlos’s Trattoria, but Uncle Russ hates it. Marissa loathes Mama Guzzardi’s, but its Dad’s all-time favorite restaurant.

In short, there is no middle ground. There are no 5’s - only 1’s and 10’s. Choosing any restaurant means that someone - inevitably - will spend the evening in sullen misery. Oftentimes, surveys that do not have any responses at or near the middle of the scale, and these responses may be a result of a person’s background (culture, region, religion, etc.). This kind of cultural bias is just another challenge to analyzing data collected on a fixed-point rating scale. (DB)
3.3 Analysis of Customer Satisfaction Data

Regression Analysis

Regression analysis is a statistical tool for estimating relationships among variables collected on ratio and interval scales. Regression coefficients (β) measure the expected change in the dependent variable (y) for a unit change in one of the independent variables (x), assuming that the other independent variables are fixed at their current values. In addition to providing estimates of the regression coefficients, regression analysis provides a measure of uncertainty in the estimated coefficients, which allows for the testing of hypotheses that an attribute does not affect the dependent variables, i.e., β = 0. Regression analysis also provides a measure of overall fit of the regression model, $R^2$. $R^2$ measures the proportion of variation in the dependent variable (y) that is accounted for, or explained, by the independent variables (x). Figure 3.5 displays data and the fitted regression line for the relationship:

$$y = 2x + \varepsilon \quad \varepsilon \sim Normal(0, \sigma^2 = 25)$$
Coefficient estimates based on the simulated data displayed in Figure 3.5 are provided in Table 3.2.

Table 3.2: Regression Analysis of Simulated Data

| Coefficient   | True Value | Estimate | Std. Error | t value | Pr(>|t|)     |
|---------------|------------|----------|------------|---------|-------------|
| Intercept ($\beta_0$) | 0          | -0.755   | 1.169      | -0.646  | 0.520       |
| Slope ($\beta_1$) | 2          | 2.278    | 0.201      | 11.341  | <2e-16 ***  |

$R^2 = 0.567$, Adj $R^2 = 0.563$, $n = 100$
Significance levels: ***=0.001; **=0.01; *=0.05; .=0.1

An estimate is a “best guess” of a coefficient’s value, and the standard error provides a measure of uncertainty of the estimate. In general, the mean $\pm 2$ standard errors forms a confidence interval for being 95% certain of likely values of the true coefficients. We take a Bayesian interpretation of the confidence interval and not a frequentist interval that involves estimates from multiple and hypothetical datasets. In doing this, we note that the p-values reported in the rightmost column, $Pr(>|t|)$, are calculated as the probability
of observing the test statistic “t,” or something larger in value, under the assumption that
the null hypothesis ($\beta_i = 0$) is true. P-values are a frequentist concept that involve the
likelihood of a test statistic in hypothetical datasets that were not actually observed in
our analysis. As a result, we loosely interpret the p-values as indicators of the probability
that coefficients are truly non-zero.

The above mentioned “Bayesian interpretation” refers to the work of Thomas
Bayes, a pioneer in statistical modeling. Bayes was a minister, philosopher, and
mathematician in the 1700’s whose work led to a modern branch of statistics.
His theories, published after his death, are used to determine the probability of
an event by using data to build models that allow inferences about unobserved
variables. His methods were way before their time, and helped lay the foundation
for modern marketing research. For more information on Bayesian Analysis, see
appendix C of the book. (DM)

The estimates differ from the true values of the coefficients because of the error term,
$\varepsilon$, that introduces randomness into the data. That is, knowing only the values of $y$ and
$x$, it is not possible to exactly recover the true value of $\beta$. The “t value” in the table
is equal to the Estimate divided by the Standard Error, and is the test statistic for the
hypothesis that $\beta = 0$. When the hypothesis is true, the explanatory variable ($x$) is
not associated with the dependent variable ($y$). The probability of observing the test
statistic, or something bigger in magnitude assuming the hypothesis is true, is provided
in the right-most columns of the table. Small values correspond to significant coefficients.
In Table 3.2, the value of 2e-16 implies that there is virtually no chance that the coefficient
is really equal to zero. Significance levels less than $\alpha = 0.05$ are generally viewed as being
statistically significant.

“Slim to none” is exactly the response you’re looking for when trying to determine
if something is statistically significant. As you review the probability calculations
in the right-most column of Table 3.2, you’re trying to identify which are most
certainly not equal to zero. The question you are asking yourself is “What are the
chances (i.e., is the probability) that the response is really different from zero?”
If your answer is “slim to none,” you’re in luck! The closer that value is to zero,
the more you can rely on an association with the dependent variable. (EW)
3.3. ANALYSIS OF CUSTOMER SATISFACTION DATA

One benefit of employing a regression model is that it allows an assessment of the importance of the drivers. This is done by considering the tradeoff among drivers that would lead to the same overall level of satisfaction. For a given value of the overall level \( y \), we can consider changing just two of the drivers, and for them to produce the same value of \( y \) it must be the case that:

\[
x_i \beta_i = -x_j \beta_j
\]

implying that the tradeoff among attributes that produces the same value of \( y \) is a line with a slope of \(-\beta_i/\beta_j\). For \( \beta_1 = 2 \) and \( \beta_2 = 4 \) the indifference curves for different values of \( y \) are shown below in Figure 3.6.

The estimated coefficients in a multiple regression analysis provide a conversion ratio for translating from one attribute to another in terms of their effect on the overall measure. More generally, since a regression equation relates all of the attributes to the same dependent variable \( y \), the same equation can be used to relate the attributes to each other.

Consumers face trade-offs in their purchase decisions because their income is limited and they have many options from which to choose. The idea behind indifference curves is that if a consumer prefers two things (i.e. two different products, such as ice cream and donuts, or two different product features, such as delicious and healthy), then there are many combinations of the two that can provide them with equal satisfaction.

Take Florida Vacations for instance. Lets say two important vacation features for “Fun-time Fran” from Chapter 2 are “relaxing” and “fun,” with fun being twice as important to her as relaxing. Think about the hours spent relaxing represented on the horizontal \( (x_1) \) axis above and the hours spent having fun represented on the vertical \( (x_2) \) axis. Any point along an indifference curve is the combination of fun and relaxing that equates to the same level of happiness for Fran. She would be willing to tradeoff less fun \( (x_2) \) for more relaxation \( (x_1) \), and would be equally happy so long as she gets two more hours of relaxation for every hour of fun she gives up. Each line in Figure 3.6 represents a combination of fun hours and relaxation hours that gives rise to a specific level of happiness, with the lines further from the origin giving greater happiness. (JA)
Running a regression analysis on a set of data provides estimates of the relationship among the variables under study. Remember, that the estimates are based on the best fit for a particular set of observed data, and have some possibility of being wrong. This is where the regression table helps.

The first thing the observer should pay attention to is the $R^2$ value of the output. The higher this number, the more reliable the overall model is in fitting the data. For example, an $R^2 = 0.60$ says that the equation resulting from the reported coefficients explains 60% of the original variability of the dependent variable ($y$).

Next, we look for drivers (X’s) that are statistically significant. If a particular driver is not significant, there is no need to consider it any further because it is plausible that it’s regression coefficient could, in fact, be zero. The level of significance is dependent on the tolerance of the analyst, with greater tolerance leading to a greater willingness to declare significance and lower tolerance being more sensitive to making a mistake in declaring a significant relationship. A “p-value” of 0.10 is generally considered to be tolerant, while a p-value of 0.01 or less indicates intolerance. The p-value is calculated as the probability of observing test statistic (t value), or something greater in magnitude if the regression coefficient were truly zero.

Once the variables are short-listed for their statistical significance, the analyst should then consider the value of the estimated coefficient and its algebraic sign. For a linear multiple regression model, a coefficient of -0.70 implies that a one unit increase in $x$ is associated with a decrease in $y$ of 0.7, whereas a coefficient of +0.8 indicates that $y$ is expected to increase by 0.8.

Note of caution: Regression is simply a best fit of data. It should not be confused with “causality.” Just because the stock market goes up during the winter months (in regressing stock returns onto temperatures) doesn’t mean that the cold weather is causing brokers to bid up stock prices to keep warm! (JC)
3.3. ANALYSIS OF CUSTOMER SATISFACTION DATA

Figure 3.6: Indifference curves for the linear model.

Ice Cream

We begin our analysis of the data by conducting a regression analysis treating overall
brand satisfaction (Q1) as the dependent variable $y$ and the brand beliefs (Q11) as
the independent variables $X$. Figure 3.7 displays the covariance matrix of brand belief
responses for Breyers Ice Cream. We calculate the covariances among all possible pairs of
question items, resulting in a $18 \times 18$ matrix with diagonal entries corresponding to the
variance of each item. Covariances are reported above the diagonal which is shaded to
make the table easier to read, and correlations are reported in *italics* below the diagonal
entries. The correlation coefficient is bound to the interval (-1, 1) and is defined as

$$
\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}}
$$

where a value of +1 indicates a perfect positive association, 0 indicates no association
and -1 indicates a perfect negative association.

The covariance calculations ignore the discreteness of the underlying data collected on
a 7-point scale, and is displayed for completeness and comparison to more formal analysis.
The off-diagonal elements of the covariance matrix are almost uniformly positive and moderate in size, consistent with the presence of scale usage effects.

**Figure 3.7: Covariance / Correlation Matrix for Breyers Ice Cream**

Covariances \((\sigma_{i,j})\) are reported above the diagonal, Correlations \((\rho_{i,j})\) below the diagonal.

The first row of numeric entries in the table are the covariances between the brand beliefs and the overall measure of satisfaction. Many of the covariances are large, especially compared to variance of overall satisfaction. In a simple regression model:

\[
y = \beta_0 + x_i \hat{\beta}_i + \varepsilon
\]

the value of the estimated regression coefficient can be shown to be \(\hat{\beta}_i \approx \sigma_{y|x_i}/\sigma_{x_i|x_i}\), where \(\sigma_{y|x_i}\) is the entry in the first row, and \(\sigma_{x_i|x_i}\) is the diagonal entry (i.e., the variance) of the \(i^{th}\) covariate. The simple regression coefficients are as large as 0.34 for “17. Great for Guests,” which suggests that a multiple regression analysis of all seventeen items might be highly predictive of overall satisfaction.

A multiple regression analysis of overall satisfaction regressed on the responses to the Q11 items results in the coefficients displayed in Table 3.3. The results are generally disappointing in that few of the estimated coefficients are significant, even at the level of \(\alpha = 0.10\). In fact, the adjusted \(R^2\) statistic is the same as for the simple regression of “17. Great for Guests.” Adding additional explanatory variables does result in improved predictive performance because of the high correlation among the covariates reported in Figure 3.7.
3.3. ANALYSIS OF CUSTOMER SATISFACTION DATA

Table 3.3: Breyers Ice Cream Regression Analysis – Raw Data

| Coefficient                        | Estimate | Std. Error | t value | Pr(>|t|) |
|------------------------------------|----------|------------|---------|----------|
| Intercept                          | 4.044    | 0.385      | 10.502  | <2e-16   *** |
| 1. Relaxing                        | -0.012   | 0.062      | -0.190  | 0.849    |
| 2. Wholesome                       | 0.112    | 0.060      | 1.843   | 0.068    . |
| 3. Fun                             | -0.068   | 0.078      | -0.865  | 0.389    |
| 4. Exciting                        | 0.028    | 0.077      | 0.368   | 0.714    |
| 5. Premium                         | -0.069   | 0.091      | -0.763  | 0.447    |
| 6. Memorable                       | 0.129    | 0.071      | 1.812   | 0.072    . |
| 7. Special Treat                   | -0.002   | 0.037      | -0.052  | 0.959    |
| 8. Regular Consumption             | -0.041   | 0.072      | -0.566  | 0.573    |
| 9. Interesting                     | -0.080   | 0.074      | -1.074  | 0.285    |
| 10. Tastes Better                  | 0.142    | 0.086      | 1.646   | 0.102    |
| 11. Variety of Flavors             | 0.054    | 0.076      | 0.714   | 0.476    |
| 12. Enjoyable                      | 0.041    | 0.069      | 0.592   | 0.555    |
| 13. Best Value                     | 0.004    | 0.063      | 0.069   | 0.945    |
| 14. Natural/Organic                | -0.005   | 0.046      | -0.109  | 0.913    |
| 15. Low Calorie                    | -0.063   | 0.049      | -1.301  | 0.195    |
| 16. Great for Whole Family         | 0.057    | 0.089      | 0.643   | 0.522    |
| 17. Great for Guests               | 0.157    | 0.099      | 1.580   | 0.117    |

$R^2 = 0.332$, $\text{Adj } R^2 = 0.247$, $n = 152$

Significance levels: ***=0.001; **=0.01; *=0.05; .=0.1

Tradeoff analysis and other forms of inference are normally conducted only on coefficients that are statistically significant. However, for illustration only, let’s conduct tradeoff analysis as if the coefficients above were significantly different from zero. The coefficient for “16. Great for the Whole Family” is 0.057 and the coefficient for “17. Great for Guests” is 0.157, or about three times larger. This implies that movement along the 7-point rating scale in Q11 of the questionnaire results in $3 \times$ the effect for item 17 than for item 16.
When analyzing customer satisfaction data it is important to keep in mind the goal of the analysis. If your goal is to predict a person’s satisfaction, then you can simply find the model with the highest $R^2$. However, if your goal is to make inferences about drivers of satisfaction with the hope of making an improvement, then you must constantly question the data you are working with by asking whether the underlying story being told makes sense. For example table 3.3 indicates that a one unit increase in “fun” is associated with a decrease in overall satisfaction by -0.068. This is essentially indicating that the more you enjoy the ice cream the less satisfied you are. This does not make much sense and should not be used as a basis for decreasing “fun” to increase satisfaction. (JP)

Florida Vacations

Covariances for the Florida Vacation survey between the overall measure of satisfaction (Q1) and the brand belief data (Q10) are displayed in Figure 3.8. The pattern of covariances are similar to that for the Ice Cream data with one exception: item 7 “Is a one-time trip only.” The covariances associated with this item are uniformly negative, implying that as respondents are more in agreement with Disney being a one-time only trip, they tend to be in less agreement with all of the other satisfaction drivers. Item 7 is an example of a question with a negative orientation – i.e., the higher the rating the less satisfied the customer. Apart from item 7, the remaining items exhibit evidence of scale-use covariance inflation in which the remaining off-diagonal elements of the covariance matrix are uniform positive and moderate in magnitude.

Regression analysis of the Florida Vacation satisfaction data uses Q10 as the independent variables ($X$) and Q1 as the dependent variable ($y$). The regression results suggest that three of the satisfaction items are significant at $\alpha = 0.05$, and one is significant at $\alpha = 0.10$. Item 1, “Relaxing” is positively associated with overall satisfaction, with a one point improvement on the 7-point rating scale associated with an expected improvement of 0.20 in overall satisfaction. Item 9, “Interesting,” is also positively associated with overall satisfaction, with a one point increase on the 7-point scale associated with an expected improvement of 0.363 in overall satisfaction. While the estimates indicate that improvements in “Interesting” would lead to greater changes in overall satisfaction than improvements in “Relaxing,” differences in the standard error of these estimates imply that “Relaxing” is actually a better bet for producing positive change. In other words, there is a greater chance that the “Interesting” attribute is not truly associated with overall satisfaction.
Covariances ($\sigma_{i,j}$) are reported above the diagonal, Correlations ($\rho_{i,j}$) below the diagonal.

As a general rule, analysis should proceed by first identifying coefficients that are significant, and then attempting to interpret the meaning of the coefficients as it relates to the problem at hand. If a coefficient has a high probability of truly being zero (e.g., “13. Enjoyable” in Table 3.4), then it makes little sense to attempt to draw inference about its meaning.

Two of the items in Table 3.4 have negative coefficients that are significantly different from zero. The first, item 7 “Is a one-time only trip” expresses disappointment with the vacation experience. A higher rating on the 7-point scale indicates less satisfaction, not greater satisfaction with the vacation experience. Item 11 “Offers a wide variety of entertainment options” also has a significant negative coefficient. The coefficient value for item 11 is large, with a one point increase in agreement with this statement associated with an expected decrease of 0.462 in overall satisfaction. This result is counter-intuitive in that a positive coefficient is expected. After all, the covariance between item 11 and overall satisfaction is reported to be positive 0.77 in Figure 3.8, not negative 0.77.

The reason for the counter-intuitive result for the item 11 coefficient is in how multiple regression analysis differs from simple regression analysis. The estimates in Table 3.4 are from a multiple regression analysis, and the regression coefficients are interpreted as the expected change in the dependent variable ($y$) for a one unit change in the explanatory variable ($x_i$), holding fixed the other explanatory variables. In simple regression analysis there are no other explanatory variables, and as a result they cannot be held constant.
Table 3.4: Magic Kingdom Regression Analysis – Raw Data

| Coefficient                  | Estimate | Std. Error | t value | Pr(>|t|) |
|------------------------------|----------|------------|---------|----------|
| (Intercept)                  | 3.224    | 0.667      | 4.815   | 5.76e-06 *** |
| 1. Relaxing                  | 0.195    | 0.097      | 2.018   | 0.046    *  |
| 2. Wholesome                 | -0.033   | 0.177      | -0.186  | 0.853    |
| 3. Fun                       | 0.231    | 0.266      | 0.866   | 0.390    |
| 4. Exciting                  | 0.016    | 0.213      | 0.076   | 0.940    |
| 5. Premium                   | -0.066   | 0.158      | -0.421  | 0.675    |
| 6. Memorable                 | 0.101    | 0.133      | 0.760   | 0.449    |
| 7. One-time Trip Only        | -0.157   | 0.070      | -2.255  | 0.026    *  |
| 8. Good Enough to Go Back    | -0.234   | 0.183      | -1.296  | 0.198    |
| 9. Interesting               | 0.363    | 0.211      | 1.721   | 0.089    .  |
| 10. More Fun                 | -0.003   | 0.161      | -0.017  | 0.987    |
| 11. Variety of Options       | -0.462   | 0.199      | -2.320  | 0.022    *  |
| 12. Enjoyable                | -0.023   | 0.265      | -0.085  | 0.932    |
| 13. Best Value               | 0.007    | 0.120      | 0.061   | 0.952    |
| 14. Authentic                | 0.078    | 0.119      | 0.650   | 0.517    |
| 15. Safe                     | 0.210    | 0.152      | 1.381   | 0.170    |
| 16. Great for Whole Family   | 0.101    | 0.163      | 0.622   | 0.536    |
| 17. Great for Adults         | 0.096    | 0.169      | 0.570   | 0.570    |

$R^2 = 0.367$, Adj $R^2 = 0.250$, n = 110

Significance levels: ***=0.001; **=0.01; *=0.05; .=0.1

The presence of non-zero covariation in the explanatory variables (Figure 3.8) is what accounts for the difference in the estimated coefficients. A unit increase in item 11 is associated with a positive increase in the other explanatory variables, except for an expected decrease in item 7, and these covariations in the explanatory variables are simultaneously considered when deriving the regression estimates that provide the best explanation of the dependent variable $y$. As a result, it is important to interpret regression coefficients carefully, giving consideration to the meaning of “holding fixed” the other explanatory variables and understanding the difference between the marginal effects reported in a multiple regression model versus the gross effect in simple regression analysis.

The analysis reported thus far is disappointing for understanding the drivers of overall satisfaction. The satisfaction measures Q1-Q3 do not show much variation across the choice alternatives, with many respondents straight-lining their responses. There is also evidence of scale-usage effects that tend to inflate covariations among the items in the
3.3. ANALYSIS OF CUSTOMER SATISFACTION DATA

analysis, resulting in multiple regression estimates that are not much more predictive than estimates from simple regression analysis. The degree of covariance inflation is larger in the Florida Vacation data than in the Ice Cream data (i.e., larger covariances), producing larger standard errors in the regression estimates and more negative and counter-intuitive estimates. However, as demonstrated in the homework exercises, more meaningful results are obtained with analysis is conducted at the segment-level. In segment-level analysis, we find many significant and plausible relationships that point to important drivers of analysis. This finding highlights the importance of not assuming everyone has the same needs, wants and sensitivities. While this point is intuitively easy to accept, it often is not adhered to in marketing analysis, especially analysis done at the aggregate level.

When choosing which restaurant to eat at, moms-on-the-go would likely place more value on the number of kids’ activities available, quick speed of service, and wholesome food options. Empty nesters, however, would likely place more value on food quality, attentive service, and atmosphere. Although both of these segments could have high overall satisfaction for the same restaurant, each has different needs, wants, and sensitivities which would result in unique regression equations. For this reason, it is important to conduct regression analysis at the segment-level. (TW)
Regression Analysis with Standardized Variables

One possibility for the disappointing results obtained above is that the data are not exchangeable. Evidence of non-exchangeable data are the large covariances in Figures 3.7 and 3.8, which might be produced by scale usage effects among respondents (see Figure 3.2). Some adjustment or standardization of the data is needed before analysis can take place when respondents use different portions of a scale to express the same sentiments. Such adjustments make the data exchangeable, or equally informative, about the parameters of a model used in their analysis.

It is commonly known in the analysis of survey data that individuals respond to scales in different ways. Some people are optimistic in their responses and tend to use the upper part of the rating scale. Others are pessimistic and tend to use the lower end of the scale. Some individuals use the entire range of the scale, while others confine their responses to a narrow range. An extreme form of scale usage are individuals who straight-line their responses by giving the same response to each question. It is difficult to interpret the responses of these individuals in terms of the regression model used in driver analysis, and straight-liners are usually removed from the data prior to analysis.

One transformation of the data often used in customer satisfaction research is standardization. Standardization involves mean-centering the data so that its variance is one:

$$z_{k,i} = \frac{x_{k,i} - \bar{x}_i}{s_i}$$

where $i$ indexes the respondent, $k$ indexes the question, and the standard deviation, $s$, is calculated as:

$$s_i = \sqrt{\frac{\sum_{k=1}^{n} (x_{k,i} - \bar{x}_i)^2}{n - 1}}$$
3.3. ANALYSIS OF CUSTOMER SATISFACTION DATA

Tutu was a young woman who lived in a small island nation off the coast of Asia yet had great global ambition. Tutu had invented a new refreshing drink, Tsychi, which tasted like carbonated ketchup. The drink was a sensation within her small island nation and Tutu was looking at market expansion. She had limited funds and needed to identify a new geography for her market, but she believed that consumers would either love or hate this uniquely flavored drink. Expanding into a new market on a large enough scale was important in order to gain a wider audience, but she would have to invest her entire capital on this next phase. This meant if Tsychi did not take off in the new market, she would have to potentially close her business venture.

Unable to afford a marketing consultant in order to identify a new market, she decided to do her own analysis. Based on a lot of qualitative research amongst her multinational friends, she identified two very different Asian markets: Bangalore, India and Tokyo, Japan. Bangalore is very populous and hot and therefore ripe for high volume sales of a cooling drink. Tokyo’s consumers had high disposable income and were always on the hunt for something unique and edgy. What she didn’t know about either market was the flavor preferences of the local consumers, but thankfully she had friends in both locations that would help her conduct some consumer research.

Tutu’s friends randomly selected names from the phone book and visited those homes to collect data on a scale of 1 to 7. When Tutu reviewed the results, the winner was clear. Bangalore consumers loved Tsychi, since the average rating was a 6, whereas their Japanese counterparts were much more lukewarm and on average rated Tsychi a 4 with not much variation.

Very excited, Tutu decided to expand into the greater Southern India region on a very large scale. Unfortunately, within the first 3 months the overall sales were minimal, and she eventually was forced to close the business. Ten years later, as an MBA student in Hong Kong, she noticed her fellow Japanese classmates were drinking a canned soda, called Tokato. She tasted it and found that it was carbonated ketchup drink! Her friends told her that everyone in Japan couldn’t get enough of this drink. Next semester in her market research class, she realized her error in interpreting her satisfaction survey. It is well know in multinational surveys that consumers in some parts of the world tend to be overly optimistic in their ratings, while the opposite is prevalent in other regions. Tutu committed a very common mistake of not norming her data to historic patterns. Multinational surveys often suffer from a lack of “exchangeability.” (DG)
When standardizing the data it is important to compute the mean $\bar{x}_i$ and standard deviation $s_i$ across multiple questions, not just the items within a particular question, so that a strict linear relationship is set up among the responses $\{x_i\}$ in which the $z_{k,i}$ sum to zero. The presence of a relationship like this will result in a regression $R^2 = 1.00$ and coefficient values $\beta_i = -1.0$ that are artificially induced by the transformation. For standardized data, a “unit” change in the explanatory variable becomes a “standard deviation” change, which is respondent-specific.

Standardization of the data is justified when it is believed that driver analysis should be based on relative measures of judgment and evaluation, not absolute measures. The argument is that human judgement is inherently relative, and all that a person can reasonably be expected to report is how much more they like one aspect of a product versus another. The equations above adjust responses to a respondent’s average response: dividing by the standard deviation controls for scale effects by reflecting the range of scale usage. Analysis then proceeds using the transformed data ($z_i$) instead of the original data ($x_i$).

Patients visiting a doctor are often asked to rate their pain on a scale of 0 to 10, where 0 indicates no pain, 5 indicates moderate pain and 10 indicates the worst pain imaginable. This scale helps people describe their level of pain as it is difficult to describe it in words. The pain scale is attempting to standardize responses with one universal scale. (RW)

**Ice Cream**

We re-analyze the Breyers ice cream data with a regression model using standardized explanatory variables while retaining the original scale for the dependent variable ($y$). We do this to facilitate comparison to the unstandardized analysis reported above, leaving analysis of fully standardized data to the cut-point model discussed below. Table 3.5 displays the standardized regression estimates.

The overall fit of the standardized regression model is worse than the model using the raw data – the adjusted $R^2$ statistic declines from 0.247 to 0.126. This is expected because standardization removes the scale usage effects that artificially inflate covariances and measured relationships among the data (see Figures 3.2, 3.7 and 3.8). Covariance inflation makes it difficult to identify which of the regressors have large marginal effects, and we find standardization results in many of the regression coefficients becoming larger and more significant. The regression estimate for item 2 “Wholesome” increases from
3.3. ANALYSIS OF CUSTOMER SATISFACTION DATA

0.112 to 0.215, item 6 “Memorable” increases from 0.129 to 0.218 and item 17 “Great for Guests” increases from 0.157 to 0.318. Levels of significance for these coefficients also increase.

Table 3.5: Breyers Ice Cream Regression Analysis – Standardized Data

| Coefficient                  | Estimate | Std. Error | t value | Pr(>|t|)  |
|-----------------------------|----------|------------|---------|-----------|
| (Intercept)                 | 5.913    | 0.141      | 41.643  | <2e-16 ***|
| 1. Relaxing                 | 0.085    | 0.094      | 0.899   | 0.370     |
| 2. Wholesome                | 0.215    | 0.093      | 2.295   | 0.023 *   |
| 3. Fun                      | -0.020   | 0.118      | -0.168  | 0.867     |
| 4. Exciting                 | 0.025    | 0.111      | 0.229   | 0.819     |
| 5. Premium                  | -0.109   | 0.137      | -0.792  | 0.429     |
| 6. Memorable                | 0.218    | 0.113      | 1.927   | 0.056 .   |
| 7. Special Treat            | -0.075   | 0.048      | -1.542  | 0.125     |
| 8. Regular Consumption      | -0.134   | 0.107      | -1.245  | 0.215     |
| 9. Interesting              | -0.049   | 0.108      | -0.454  | 0.650     |
| 10. Tastes Better           | 0.157    | 0.117      | 1.331   | 0.185     |
| 11. Variety of Flavors      | -0.027   | 0.094      | -0.285  | 0.776     |
| 12. Enjoyable               | -0.007   | 0.099      | -0.067  | 0.946     |
| 13. Best Value              | -0.044   | 0.090      | -0.479  | 0.632     |
| 14. Natural/Organic         | 0.002    | 0.067      | 0.026   | 0.979     |
| 15. Low Calorie             | -0.092   | 0.067      | -1.356  | 0.177     |
| 16. Great for Whole Family  | 0.065    | 0.113      | 0.570   | 0.569     |
| 17. Great for Guests        | 0.318    | 0.144      | 2.207   | 0.029 *   |

$R^2 = 0.226$, Adj $R^2 = 0.126$, $n = 149$

Significance levels: ***=0.001; **=0.01; *=0.05; .=.01

Florida Vacations

Standardized regression results for Magic Kingdom are displayed in Table 3.6. The results are similar to those found with the Ice Cream data – the adjusted $R^2$ measure decreases from 0.250 to 0.178 while the magnitude of most of the coefficients increase. The coefficient for item 7 “Is a one-time trip only” increases from -0.157 to -0.308 and becomes highly significant. Item 11 “Offers a wide variety of entertainment options” remains approximately the same in magnitude with a value of -0.464, implying that
the relative importance of item 7 has doubled in the standardized regression analysis as compared to the analysis with the raw data.

Differences between the standardized regression results and results based on the raw data highlight the importance of coming to an analysis with a modeling perspective in mind. Without a perspective, one is forced to choose from among two sets of analysis without substantive criteria, and the choice will be driven by non-substantive issues such as picking the analysis with the better $R^2$ measure. Substantive factors that should weigh into the decision include i) whether respondents are capable of providing absolute, as opposed to relative, measures of their beliefs; and ii) whether respondents engage in selective use of scales (e.g., yeasayers and naysayers). An affirmative response to either criteria points to using analysis based on the standardized data.

Table 3.6: Magic Kingdom Regression Analysis – Standardized Data

| Coefficient | Estimate | Std. Error | t value | Pr(>|t|) |
|-------------|----------|------------|---------|---------|
| (Intercept) | 5.164    | 0.267      | 19.317  | <2e-16  *** |
| 1. Relaxing | 0.195    | 0.142      | 1.374   | 0.172   |
| 2. Wholesome| -0.039   | 0.217      | -0.179  | 0.858   |
| 3. Fun      | 0.358    | 0.349      | 1.027   | 0.307   |
| 4. Exciting | -0.071   | 0.258      | -0.276  | 0.783   |
| 5. Premium  | -0.088   | 0.213      | -0.412  | 0.681   |
| 6. Memorable| 0.028    | 0.201      | 0.139   | 0.889   |
| 7. One-time Trip Only | -0.308 | 0.080 | -3.842 | 0.000 *** |
| 8. Good Enough to Go Back | -0.357 | 0.243 | -1.468 | 0.145   |
| 9. Interesting | 0.400 | 0.257 | 1.558 | 0.122   |
| 10. More Fun | 0.031 | 0.188 | 0.168 | 0.866   |
| 11. Variety of Options | -0.464 | 0.240 | -1.934 | 0.056 . |
| 12. Enjoyable | 0.399 | 0.332 | 1.201 | 0.232   |
| 13. Best Value | 0.065 | 0.165 | 0.392 | 0.696   |
| 14. Authentic | 0.189 | 0.180 | 1.053 | 0.295   |
| 15. Safe      | 0.267    | 0.218      | 1.223   | 0.224   |
| 16. Great for Whole Family | -0.039 | 0.234 | -0.168 | 0.866   |
| 17. Great for Adults | 0.143 | 0.223 | 0.643 | 0.522   |

$R^2 = 0.313$, Adj $R^2 = 0.178$, $n = 101$
Significance levels: ***=0.001; **=0.01; *=0.05; .=0.1
3.3. ANALYSIS OF CUSTOMER SATISFACTION DATA

Dummy Variables and Logistic Regression

Suppose it is believed that the explanatory data \( x \) provides ordinal-scaled information about each of the drivers. Ordinal data can be represented by a series of nominal dummy (0-1) variables in a regression model, each associated with a particular item’s response value. For the 7-point scale displayed in Figure 3.1, we would re-code the observed data by producing a series of dummy variable terms \( (D_1 \beta_1 + D_2 \beta_2 + \cdots + D_7 \beta_7) \) for each of the original terms \( (x_i \beta_i) \) in the regression equation. The dummy variables avoid making the assumption that the data are interval-scaled by associating a unique regression coefficient with every possible data value.

Table 3.7: Dummy Variable Coding for 7-Point Scaled Responses

<table>
<thead>
<tr>
<th>Scale Value</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
<th>D6</th>
<th>D7</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x = 1 )</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( x = 2 )</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( x = 3 )</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( x = 4 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( x = 5 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( x = 6 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( x = 7 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3.8: Effects of Scale Values

<table>
<thead>
<tr>
<th>A scale value of ( \cdots ) ( \cdots ) is represented as:</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \beta_3 )</th>
<th>( \beta_4 )</th>
<th>( \beta_5 )</th>
<th>( \beta_6 )</th>
<th>( \beta_7 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x = 1 )</td>
<td>( \beta_1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x = 2 )</td>
<td>( \beta_2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x = 3 )</td>
<td>( \beta_3 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x = 4 )</td>
<td>( \beta_4 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x = 5 )</td>
<td>( \beta_5 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x = 6 )</td>
<td>( \beta_6 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x = 7 )</td>
<td>( \beta_7 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The effect on the dependent variable, \( y \), of moving from a scale value of \((j-1)\) to \((j)\) is represented by the difference of coefficients \( (\beta_j - \beta_{j-1}) \). Since the original data are assumed to be nominal, there are no restrictions placed on the coefficients to ensure
a progressive ordering of values. The nominal coding scheme is often encountered in conjoint studies, which will be discussed in Chapter 4.

It is often the case that the regression model is specified with an intercept, requiring additional assumptions to identify models with dummy variables. It is standard in models to “zero out” the lowest level of the scale value by imposing the restriction \( \beta_1 = 0 \) in the model. When this occurs, the model intercept, \( \beta_0 \), is interpreted as the expected value of the dependent variable when all scale values are at their lowest level, and the remaining \( \beta \) coefficients measure the expected change in the dependent variable for changes in one of the independent variables.

The use of dummy variables in a regression model is plausible when there are relatively few variables and few values for which a dummy representation is needed. If we were to substitute the seventeen brand belief items in Q11 of the Ice Cream survey, or Q10 of the Florida Vacations survey, there would be \( 6 \times 17 = 102 \) dummy variables introduced into the analysis. This would overwhelm the analysis as there would then be about as many coefficients as observations. Regression analysis requires at least 4-5 observations per coefficient to produce reliable and stable results that are potentially free from over-fitting the data.

We therefore do not pursue analysis involving a dummy variable representation of the explanatory variables, instead advocating the use of cut-point models to deal with ordinally scaled variables. A parsimonious alternative is to use just one dummy variable for each item to reflect a specific scale value. We might argue that the 7-point scale is only informative about extreme positive views, i.e., a response of a seven. If we hold this belief, or if we were to entertain this hypothesis, then we could re-code the data to indicate “top-box” responses:

\[
x^*_i = \begin{cases} 
1 & \text{if } x_i = K \\
0 & \text{if } x_i < K 
\end{cases}
\]

where “K” is the response associated with the highest rating, or top box. In a top box regression model, the coefficient \( \beta \) is the expected change in the dependent variable \( y \) for a driver having a score of \( K \) (e.g., 7) versus having a score less than \( K \).

**Logistic Regression**

Ordered and nominal dependent variables require models that explicitly account for the discreteness of the dependent variable. In regression models, the dependent variable is assumed to be some function of the independent variables, or attributes, plus a continuous error term. A normal distribution is assumed for the regression model, admitting the possibility of any real-valued realization of the dependent variable \( y \). The regression
3.3. ANALYSIS OF CUSTOMER SATISFACTION DATA

model should be viewed as an approximation because the dependent variable is discrete and not continuous.

A simple approach to dealing with a discrete dependent variable is to assume that the observed rating on the fixed-point scale is a censored realization of a latent, unobserved variable. In an effort to drive satisfaction levels to the highest possible value, analysis can focus on whether a response is in the top-box or not:

\[ y_i^* = \begin{cases} 1 & \text{if } y_i = K \\ 0 & \text{if } y_i < K \end{cases} \]

The dependent variable is treated as if it were dichotomous, similar to a dummy variable for the independent variables. This specification is known as a binary probit model when the error term is distributed according to a normal distribution.

Have you ever wondered what the chances are that a customer will repurchase your product? Remain a customer? Respond to a change in your product? Then you will find logistic regression useful. It may seem complicated, but basically it is a tool used to model the probability of success. When properly applied, the analysis yields insights into what variables are more or less likely to predict a discrete outcome. (RB)

An alternative model for top-box analysis is obtained when the error term is assumed to be distributed according to an extreme value distribution, resulting in what is known as a “logit” model. It is remarkable that for extreme value errors the probability of observing \( y_i = 1 \) can be expressed as:

\[ \Pr(y_i^* = 1) = \frac{\exp[V_i]}{1 + \exp[V_i]} \]

where \( V_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_k x_{k,i} \). Figure 3.9 provides a graph of the relationship between the value of \( V_i \) and the probability.

**Ice Cream**

Logistic regression is inherently less informative than regular regression because the dependent variable \((y_i^*)\) collapses across responses of the original variable \((y_i)\). This loss of information results in larger standard errors in Table 3.9, and as a result fewer coefficients that are viewed as significant. Item 2 “Wholesome” is the only variable found to be marginally significant in the model, with a coefficient estimate of 0.343.
Interpreting coefficients in a logistic regression model is different from a regular regression model. Regression coefficients for the raw data in Table 3.3 are interpreted as the expected change in the dependent variable on the 7-point scale for a one point change in the explanatory variable. This interpretation is due to all variables being measured on the same 7-point scale. The regression coefficients for the standardized analysis reported in Table 3.5 are the expected change in the dependent variable on the 7-point scale for a one standard deviation increase in the explanatory variable. The difference is due to the “unit” measurement of the explanatory variable. For logistic regression, the regression coefficients are interpreted as the change in the logarithm of the odds of observing $y_i^* = 1$ versus $y_i^* = 0$. This is because the odds are defined as:

$$\frac{\text{Pr} (y_i^* = 1)}{\text{Pr} (y_i^* = 0)} = \frac{\text{Pr} (y_i^* = 1)}{1 - \text{Pr} (y_i^* = 1)} = \frac{\exp[V_i]}{1 + \exp[V_i]} = \exp[V_i]$$

and taking the logarithm of the odds results in the linear regression expression in the explanatory variables:
3.3. ANALYSIS OF CUSTOMER SATISFACTION DATA

\[
\ln \left( \frac{\Pr (y_i^* = 1)}{\Pr (y_i^* = 0)} \right) = \ln (\exp [V_i]) = V_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_k x_{k,i}
\]

The coefficient for item 2 “Wholesome” is equal to 0.343, and implies that a one unit increase along the 7-point rating scale is associated with an expected increase of 0.343 in the log-odds of observing a top box score for \( y \). This translates to an increase in the odds of \( \exp[0.343] = 1.409 \) or a 40.9% increase. Coefficients greater than zero indicate an increase in the odds, and negative coefficients indicate a decrease in the odds. For item 2, we see that the size of the effect on the top-box score is large, but barely significant. Other items that were previously found to be significant (item 6 “Memorable” and item 17 “Great for Guests”) continue to have large coefficients but lack statistical significance. The results are therefore suggestive of relationships, but not definitive.

Table 3.9: Breyers Ice Cream Logistic Regression Analysis

| Coefficient                | Estimate | Std. Error | t value | \( \Pr(> |t|) \)    |
|----------------------------|----------|------------|---------|----------------|
| (Intercept)                | -7.260   | 1.551      | -4.680  | 2.87e-06 ***   |
| 1. Relaxing                | 0.011    | 0.212      | 0.052   | 0.958          |
| 2. Wholesome               | 0.343    | 0.204      | 1.681   | 0.092 .       |
| 3. Fun                     | -0.227   | 0.262      | -0.868  | 0.385          |
| 4. Exciting                | 0.071    | 0.249      | 0.287   | 0.774          |
| 5. Premium                 | -0.289   | 0.328      | -0.882  | 0.377          |
| 6. Memorable               | 0.307    | 0.234      | 1.311   | 0.189          |
| 7. Special Treat           | 0.013    | 0.123      | 0.108   | 0.914          |
| 8. Regular Consumption     | -0.275   | 0.263      | -1.046  | 0.295          |
| 9. Interesting             | -0.356   | 0.255      | -1.394  | 0.163          |
| 10. Tastes Better          | 0.300    | 0.290      | 1.034   | 0.301          |
| 11. Variety of Flavors     | 0.287    | 0.256      | 1.120   | 0.262          |
| 12. Enjoyable              | 0.254    | 0.226      | 1.123   | 0.261          |
| 13. Best Value             | 0.123    | 0.218      | 0.568   | 0.569          |
| 14. Natural/Organic        | -0.029   | 0.151      | -0.198  | 0.843          |
| 15. Low Calorie            | -0.016   | 0.161      | -0.104  | 0.917          |
| 16. Great for Whole Family | 0.258    | 0.303      | 0.852   | 0.394          |
| 17. Great for Guests       | 0.491    | 0.332      | 1.475   | 0.140          |

Significance levels: ***=0.001; **=0.01; *=0.05; .=0.1
Florida Vacations

Table 3.10 displays logistic regression results for Magic Kingdom. Similar to the Ice Cream analysis, the standard errors of the coefficients are larger due to the collapsing of the dependent variable responses. In contrast, many of the coefficients are more significant than in previous analysis. We find that items 6 “Memorable” and 11 “Offers a wide variety of entertainment options” continue to be significant. In addition, items 7 “Is a one-time only trip,” 8 “Good enough to go back for another vacation,” 9 “Interesting” and 12 “Enjoyable” are now marginally significant with coefficients that are large in magnitude.

Table 3.10: Magic Kingdom Logistic Regression Analysis

| Coefficient             | Estimate | Std. Error | t value | Pr(>|t|) |
|-------------------------|----------|------------|---------|----------|
| (Intercept)             | -8.535   | 2.262      | -3.773  | 0.001 ***|
| 1. Relaxing             | 0.036    | 0.201      | 0.180   | 0.857    |
| 2. Wholesome            | -0.520   | 0.412      | -1.262  | 0.206    |
| 3. Fun                  | 0.442    | 0.547      | 0.808   | 0.419    |
| 4. Exciting             | -0.356   | 0.461      | -0.773  | 0.439    |
| 5. Premium              | -0.224   | 0.350      | -0.641  | 0.521    |
| 6. Memorable            | 0.811    | 0.305      | 2.652   | 0.007 ** |
| 7. One-time Trip Only   | -0.272   | 0.154      | -1.771  | 0.076 .  |
| 8. Good Enough to Go Back | -0.794 | 0.411      | -1.933  | 0.053 .  |
| 9. Interesting          | 0.764    | 0.462      | 1.652   | 0.098 .  |
| 10. More Fun            | 0.077    | 0.332      | 0.233   | 0.816    |
| 11. Variety of Options  | -1.156   | 0.489      | -2.361  | 0.018 *  |
| 12. Enjoyable           | 1.173    | 0.608      | 1.928   | 0.053 .  |
| 13. Best Value          | 0.175    | 0.241      | 0.727   | 0.467    |
| 14. Authentic           | 0.154    | 0.248      | 0.619   | 0.536    |
| 15. Safe                | 0.360    | 0.358      | 1.006   | 0.314    |
| 16. Great for Whole Family | 0.030 | 0.373      | 0.082   | 0.934    |
| 17. Great for Adults    | 0.561    | 0.345      | 1.625   | 0.104    |

Significance levels: ***=0.001; **=0.01; *=0.05; .=0.1

It is difficult to fully make sense of the algebraic signs of the coefficients in Table 3.10. Positive coefficients imply that the odds of observing a top box score increases as the rating of the explanatory variable increases, and negative coefficients imply that the odds decrease. It is not clear, for example, why item 11 in the table should have a negative
3.3. ANALYSIS OF CUSTOMER SATISFACTION DATA

coefficient – i.e., why greater agreement with the statement “Offers a wide variety of entertainment options” would lead to decreased odds of Magic Kingdom receiving a top box score for overall satisfaction.

There are many potential reasons for this, none of which can be substantiated without more study and data. For example, it may be that individuals giving high scores for item 11 feel that the entertainment options, while varied, are not of high quality. Or they may feel that the options are many, but not relevant to them. Additional analysis is needed to clarify the meaning of this and other coefficients. Research in marketing, as in all scientific inquiry, raises many questions as it attempts to find answers, often pointing to the need for additional research.

Cut-Point Models

The most flexible model for analyzing customer satisfaction data is the cut-point model that allows for heterogeneity in how respondents use the fix-point scale (Rossi et al. (2001), Buschken et al. (2012)). Cut-point models assume the data are ordinally scaled and are non-exchangeable, so transformation of the data is needed prior to analysis. The cut-point model assumes that the observed data are a censored realization of continuous variables that are distributed multivariate normal:

\[ x_{i,j} = k \text{ if } c_{k-1} \leq z_{i,j} \leq c_k \]

and

\[ z_i \sim \text{Normal} \left( \mu + \tau_i \iota, \sigma_i^2 \Sigma \right) \]

where “i” denotes the respondent, “j” indexes the question, “k” is the response and \( \iota \) is a vector of ones. In this model, there is no distinction between the dependent (y) and independent (x) as in a regression model. Instead, all observations (x) are treated as draws from a latent multivariate normal distribution where the mean and covariance matrix that are individual-specific.

The cut-point model relates the probability of observing a response of “k” on the rating scale to normally distributed normal variable, z. The probability is equal to the mass associated with the area underneath the latent normal distribution within the cut points, as shown in Figure 3.10.

The cut-point model allows for the possibility that respondents are heterogeneous in the use of the rating scale, with some respondents using the upper end of the scale, others tending to use the lower end of the scale and some using the entire scale. Yeasayers who use the upper end of the scale would have large values of \( \tau_i \) that would increase the
probability that larger response values would be observed. Naysayers would have negative values ($\tau_i < 0$) that would push the latent density down toward the bottom of the scale. Large values of the $\sigma_i^2$ parameter in equation cut-point model spreads out the latent density, making the use of the entire scale more probable.

The introduction of the model parameters for scale-use heterogeneity, i.e., $\tau_i$ and $\sigma_i^2$, allow us to make adjustments to the responses of an individual to make them exchangeable. If respondents have a natural tendency to use different portions of the scale for cultural or other reasons, then analysis must adjust for these tendencies before information is combined across respondents. Alternatively, one can view the information arising from fixed-point scales as inherently relative, with respondents providing information on a scale that is relative to their other responses. Thus, an “outstanding” rating for one person is not the same as another person’s, and adjustments need to be made to transform the data onto a common scale.

The latent variable of the cut-point model can be alternatively expressed as:

$$z_i^* = \frac{z_i - \tau_i - \mu}{\sigma_i} + \mu \sim \text{Normal} (\mu, \Sigma)$$
3.3. **ANALYSIS OF CUSTOMER SATISFACTION DATA**

This alternative form of the model indicates that data can be put on a common scale by subtracting the latent \( z \) values by a location parameter that centers the data, and then dividing by a scale parameter. This is similar to the previously described standardization that involves subtraction of the mean and division by the standard deviation, except that simple standardization of the data does not account for integer, ordinal responses on a discrete rating scale. The implication of the above equation is that the common parameters \( \mu \) and \( \Sigma \) can be used to make across-respondent inferences corresponding to the standardized, or exchangeable, data \( z^*_i \).

Mary Jones went to the hospital under the encouragement of friends and family after experiencing pain in her abdomen for over 24 hours. When asked at the hospital to rate her pain on a scale of 1-10, she gave it a 6. The doctors and nurses spent many hours conducting tests only to find out she had severe appendicitis that was on the verge of bursting. She was rushed into surgery just in time to save her life. After the surgery, the hospital staff pondered why Mary had only rated her pain a 6. If it had been a 10, they would have looked into the possibility of a ruptured appendix much sooner. Why did Mary feel that her pain was not worthy of a 10 when others would have?

Without knowing Mary’s cultural background, personality or past physical experiences (perhaps she experienced pain quite frequently for other reasons and this was not as relatively high compared to her normal state than for others) it is difficult for others to assess the response given as high or low. For some, a 6 may mean slight or moderate pain and not the sign of a life threatening issue as it was for Mary. This is an example of responses that are not exchangeable.

In order to adjust for the relative nature of responses in non-exchangeable ordinal scaled data, the cut-point model can be used to adjust data for scale-use tendencies. However, because the underlying raw data is relative and not exchangeable, analysis and interpretation of results must be done relative to the past levels of pain expressed by each individual. (BB)

A property of the multivariate normal distribution is that all conditional distributions of its elements are also normal and resemble a regression model. For any partitioning of the vector \( z^* \) into components we have for similarly partitioned parameters:

\[
p (z_1^* | z_2^*) = \text{Normal} \left( \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (z_2^* - \mu_2), \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \right)
\]
where the mean of the conditional distribution of $z_1^*$ depends on the values of $z_2^*$. For $z_1^*$ the overall measure of satisfaction and $z_2^*$ the drivers of satisfaction, the expression $\Sigma_{12}\Sigma_{22}^{-1}$ is equivalent to the expression for regression coefficients:

$$
\beta' = \left( (X'X)^{-1} X'y \right)^t
= y'X(X'X)^{-1}
= \Sigma_{12}\Sigma_{22}^{-1}
$$

Thus the assumption that the latent, scale-adjusted responses $z_i^*$ follow a multivariate Normal distribution allows us to conduct regression analysis with the estimates of the covariance matrix $\Sigma$. The difference between driver analysis conducted using the cut-point model versus the regression model is that the data have been adjusted for scale-use tendencies.

**Ice Cream**

Regression estimates for Breyers from the cut-point model are reported in Table 3.11. Previously we found that three items, Q2 “Wholesome,” Q6 “Memorable” and Q17 “Great for guests,” showed signs of being explanatory of overall satisfaction. Unfortunately, these items are found to have neither large nor significant coefficients in the cut-point model. Instead, item 7 “Buy as a special treat but not regularly” and item 15 “Low calorie” are found to have marginally significant negative coefficients, implying that satisfaction is higher among respondents who evaluate these items lower on the 7-point scale.

It must be remembered that there exists many possible models to apply to these data, with the cut-point model being a simple example of more elaborate extensions. Cut-point model enhancements could include more flexible specifications for the cut-points, or possible alternative models for the latent distribution of $z_i^*$. An example of an alternative structure is one in which the latent distribution reflects the “haloing” of responses where respondents retrieve from memory an overall evaluation from which their response to all scale items originate. If the respondent had an excellent experience, then all their responses reflect this excellence without discrimination. If the respondent had a bad experience, then their responses exhibit the same degree of negativity. Responses generated in this way should not be used to conduct driver analysis that assumes that the individual items “drive” the overall satisfaction rating. With haloed responses, the items are instead being driven by the overall evaluation. The fact that results across the models differ may point to the need for more realistic and elaborate explanations of the
3.3. **ANALYSIS OF CUSTOMER SATISFACTION DATA**

process by which the data are generated by respondents. It is important that marketing research results tell a plausible story of consumer behavior, and not simply rely on a model because it has the best fit.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Posterior Mean</th>
<th>Posterior Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Relaxing</td>
<td>-0.017</td>
<td>0.139</td>
</tr>
<tr>
<td>2. Wholesome</td>
<td>0.050</td>
<td>0.155</td>
</tr>
<tr>
<td>3. Fun</td>
<td>-0.092</td>
<td>0.164</td>
</tr>
<tr>
<td>4. Exciting</td>
<td>0.088</td>
<td>0.160</td>
</tr>
<tr>
<td>5. Premium</td>
<td>-0.147</td>
<td>0.177</td>
</tr>
<tr>
<td>6. Memorable</td>
<td>0.038</td>
<td>0.186</td>
</tr>
<tr>
<td>7. Special Treat</td>
<td>-0.220</td>
<td>0.129</td>
</tr>
<tr>
<td>8. Regular Consumption</td>
<td>-0.157</td>
<td>0.158</td>
</tr>
<tr>
<td>9. Interesting</td>
<td>-0.198</td>
<td>0.179</td>
</tr>
<tr>
<td>10. Tastes Better</td>
<td>0.080</td>
<td>0.156</td>
</tr>
<tr>
<td>11. Variety of Flavors</td>
<td>-0.024</td>
<td>0.156</td>
</tr>
<tr>
<td>12. Enjoyable</td>
<td>0.068</td>
<td>0.144</td>
</tr>
<tr>
<td>13. Best Value</td>
<td>-0.084</td>
<td>0.147</td>
</tr>
<tr>
<td>14. Natural/Organic</td>
<td>-0.062</td>
<td>0.116</td>
</tr>
<tr>
<td>15. Low Calorie</td>
<td>-0.197</td>
<td>0.118</td>
</tr>
<tr>
<td>16. Great for Whole Family</td>
<td>0.024</td>
<td>0.157</td>
</tr>
<tr>
<td>17. Great for Guests</td>
<td>0.062</td>
<td>0.171</td>
</tr>
</tbody>
</table>

Significance levels: ***=0.001; **=0.01; *=0.05; .=0.1

**Florida Vacations**

Regression estimates for Magic Kingdom for the cut-point model are reported in Table 3.12. Item 7 “Is a one-time only trip” is found to be significantly related to the measure of overall satisfaction. This item was previously identified as being a potentially important driver of satisfaction, and should be retained for further analysis along with other variables that show evidence of relationship to the overall measure. As with the Breyers ice cream results, we find that the cut-point model produces results that are different from those generated by other models.
Table 3.12: Magic Kingdom Cut-Point Model Analysis ($\Sigma_{12}^{-1}\Sigma_{22}$)

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Posterior Mean</th>
<th>Posterior Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Relaxing</td>
<td>0.076</td>
<td>0.160</td>
</tr>
<tr>
<td>2. Wholesome</td>
<td>0.017</td>
<td>0.204</td>
</tr>
<tr>
<td>3. Fun</td>
<td>0.027</td>
<td>0.231</td>
</tr>
<tr>
<td>4. Exciting</td>
<td>-0.057</td>
<td>0.189</td>
</tr>
<tr>
<td>5. Premium</td>
<td>-0.019</td>
<td>0.212</td>
</tr>
<tr>
<td>6. Memorable</td>
<td>0.074</td>
<td>0.182</td>
</tr>
<tr>
<td>7. One-time Trip Only</td>
<td>-0.422</td>
<td>0.147 *</td>
</tr>
<tr>
<td>8. Good Enough to Go Back</td>
<td>-0.089</td>
<td>0.180</td>
</tr>
<tr>
<td>9. Interesting</td>
<td>0.118</td>
<td>0.217</td>
</tr>
<tr>
<td>10. More Fun</td>
<td>-0.182</td>
<td>0.192</td>
</tr>
<tr>
<td>11. Variety of Options</td>
<td>-0.290</td>
<td>0.203</td>
</tr>
<tr>
<td>12. Enjoyable</td>
<td>0.148</td>
<td>0.256</td>
</tr>
<tr>
<td>13. Best Value</td>
<td>-0.113</td>
<td>0.175</td>
</tr>
<tr>
<td>14. Authentic</td>
<td>0.000</td>
<td>0.181</td>
</tr>
<tr>
<td>15. Safe</td>
<td>-0.128</td>
<td>0.204</td>
</tr>
<tr>
<td>16. Great for Whole Family</td>
<td>-0.090</td>
<td>0.193</td>
</tr>
<tr>
<td>17. Great for Adults</td>
<td>0.101</td>
<td>0.193</td>
</tr>
</tbody>
</table>

Significance levels: ***=0.001; **=0.01; *=0.05; .=0.1

3.4 Summary

There are many challenges in analyzing customer satisfaction data. We found in our analysis that most people are quite satisfied with offerings in both the Ice Cream and Florida Vacations product categories (Figures 3.3 and 3.4). This frequently occurs when analyzing healthy brands, and results in little available information for identifying the drivers of satisfaction. In the extreme case, if all respondents give a brand a top-box evaluation in overall satisfaction, then it would not be possible to make inferences about any of the potential drivers. It is only through variation in responses and drivers that statistical relationships can be identified.

Lack of variation is often found in survey data when respondents are not forced to choose among alternatives, or otherwise probe the depths of their thoughts. It is often easier to report the first thing that comes to mind, and this economizing results in a general lack of discrimination among the items being considered. This is one reason why
3.4. SUMMARY

we often favor collecting data using a Pick any/J format – it gets at the basic issue of whether the variable plays a role in a process without assuming that there exists finer degrees of support for a statement, intention or need. Choice models, introduced in the next chapter, are one way of inducing greater discrimination among items and have been shown to lead to better predictions of marketplace behavior.

A second and more important challenge encountered with satisfaction data is the assumption that the data are exchangeable. Our view is that the preferences and sensitivities of an individual are inherently relative. This uniqueness makes it difficult to measure the importance of one person’s needs relative to another person. Consider, for example, any of the motivations displayed in Figure 3.1, and consider the possible ways that these items could be studied, either through the use of rating scales, ranking scales or choice exercises. These methods can only reveal an item’s importance on a scale relative to the same respondent’s other responses.

To illustrate this point, reflect on your answer to the question “How much does your wife (husband, significant other) love you?” Suppose you were given a 7-point rating scale for your answer ranging from “not at all” to “a lot.” Responses to this question have little meaning unless respondents agree on the meaning of descriptors such as “not at all” or “a lot” and their responses can be related to some variable that has absolute meaning.

We compare ourselves with other people all the time. Would you be happy if you got a score of 25 on your test? You can’t possibly tell. What if you found out that the average was 10? Would you feel differently in the average was 35?

The truth is that we value things relatively. Instead of using an absolute scale, we continuously make comparisons. Do you like brand A more than brand B? Are you happier today than you were last week? Really, how much does your significant other loves you? Can you put a number to that? At least, one should hope, she loves you more than she loves the dog! (PM)

We believe that an individual’s responses to questions about needs, attitudes and opinions are inherently relative, and that it is not realistic to assume that people can agree on a common meaning of scales. One person’s rating of a 6 on the 7-point scale is not necessarily the same as another person’s rating of a 6, particularly when considering cultural norms that exist in response styles. Fundamentally, the importance, pleasure and pain of one person cannot be directly measured relative to another’s without the presence of additional qualifications and assumptions.
Our analysis in this section is a bit disappointing in that none of the approaches examined produced satisfactory results when applied to all the data in the sample. However, we find in the homework exercises that more meaningful results are obtained when analysis is restricted to respondents within a particular market segment. We will see in Chapter 4 that conducting disaggregate analysis is further enhanced by including variables such as prices that serve to tie together across-respondent analysis. In models of economic behavior, observed choices are related to the marginal utility (i.e., the change in utility), which allows the actual level of utility to be different across respondents. Analysis across respondents is conducted using measures such as a person’s willingness to pay for the right to acquire and use the benefits of an offering. Here price plays the role of a common denominator by translating the marginal utility into a monetary equivalent that is comparable across people. Furthermore, we will show that analysis based on Pick any/J data analyzed within the context of a choice model yields results that are interesting and interpretable. The advantage of using Pick any/J data is that it minimizes the unwanted influence of scale effects.

When shopping for new eyeglasses, the opthalmologist gives you a variety of options to choose from. Based on your unique needs and anatomical limitations, certain lenses may fit better than others. As you narrow your consideration set by eliminating options in a compensatory and non-compensatory manner, you finally arrive at your final choice – your newest pair of glasses.

The metaphor of choosing eyeglasses represents a similar choice task performed in marketing research. Regression analysis is the first set of lenses we use to analyze any problem. This leaves the resulting image a bit blurry. The next sets of lenses correspond to the use of standardized variables. As we gain greater resolution, scale issues become the new concern. We then introduce dummy variables, logistic regression and the cut-point model to decipher our best fit and control for nominal scaling.

Data interpretation viewed from one lens can lead market researchers to develop a dangerous tunnel vision for the future. The challenge remains how to integrate new methods to analyze the same data that may give wildly different results. New tools help market researchers develop a whole new line of sight for the future. Marketing is all about developing peripheral vision by fitting new lenses to the same set of eyes. (BK and JL)
3.4. SUMMARY

We saw in this chapter that the presence of scale-usage effects artificially inflates covariances among the variables measured (see Figures 3.2, 3.7 and 3.8). This results in greater (but artificially induced) fits in regression models (i.e., higher $R^2$ measures) when using the raw data without some form of correction. Care must therefore be exercised in assessing the fits of different models because a purely statistical approach to model selection would lead to models in which scale-use effects are most prominent. It is important that marketing research results tell a plausible story of consumer behavior, and that a model isn’t relied upon simply because it has the best fit.

Finally, we stress that the analysis of customer satisfaction data is suggestive, but not definitive about the drivers of brand preference. Analysis in marketing is often iterative and exploratory, and it is useful to seek convergent results across multiple methods to bolster confidence that the results are truly present in the data and not an artifact of one methodology. Our view is that the analysis conducted in this chapter will lead to better-informed decisions about aspects of the choice model we develop and implement next in Chapter 4.

Progressing through statistical analysis is very similar to the process of becoming a better artist. In early artistic instruction, classes require students to paint by numbers and stay within the lines of a pre-set picture. The template and the outcome are predetermined and intended to be easy to achieve. As the artist progresses, guidance falls away and primary blocks of color give way to shading and blurring of edges. A wider range of tools and methods are employed to add nuance and detail to the artwork that emerges.

Eventually the artist is adept at producing realist works as well as more abstract and thought provoking images. In the same way, beginning statisticians are given fully formed datasets and software packages that dictate one certain outcome. As more skills are developed, error-free manufactured datasets give way to real-life data collection. Rote software regression outputs evolve into more thoughtful analyses and different applications of statistical tools. In both fields, the concrete and predetermined gives way to abstraction and freedom of thought which makes for better results. This ambiguity no doubt produces frustration along the way, but is a necessary fact of working through difficult problems in the real world. (CM)
CHAPTER 3. CUSTOMER SATISFACTION

3.5 Homework

1. Select a dependent variable – Q1 (overall brand satisfaction), Q2 (repeat purchase intent) or Q3 (likelihood to recommend) – that you believe is a good summary measure of value for your brand. Conduct a regression analysis of the relationship between this dependent variable and the brand belief variables using Tab 3 of the IDT:

   a) Start with an analysis that includes just one variable that makes sense to include. Note the $R^2$ measure of model fit associated with this initial model, and retain this variable if the regression coefficient is significant.

   b) Add another candidate variables and note the difference in model fit and associated t statistic. Retain this variable if it is significant and the sign of the coefficient makes sense to you.

   c) Continue adding variables until you are satisfied with the model and its interpretation. It is acceptable to have just a few variables in your final model.

2. Repeat your analysis using the other three models:

   a) Regression analysis with standardized variables.

   b) Logistic regression.

   c) Cut-point model.

3. Repeat your analysis for the regression, standardized regression and logistic regression models using one or more of the target segments identified in your Tab 2 analysis.

4. Which result to you believe to be true? Come to class ready to discuss your assumptions, analyses and conclusions.

5. Continue your journal for the class by recording your thoughts about the following issues:

   a) What did you learn from this chapter?

   b) Where are you confused?

   c) What do you believe to be true from the class material?

   d) What do you question or not totally believe?

   e) What learning can you apply to your brand?
3.5. *HOMEWORK*

f) What next piece of analysis would you conduct?

6. What vignette topic is needed to better explain, illustrate or apply the chapter material? Where would it be located? What would you say?
Chapter 4

Product Analysis

Relating needs to wants...

Product decisions involve two important constructs that are closely related to each other - choice and heterogeneity. Choice involves how specific individuals make specific choices, what they want and how they process information about products to arrive at a preference ordering. Heterogeneity involves characterizing why different respondents want different things, i.e., how their needs are fulfilled. Product analysis combines the study of choice with the study of heterogeneity to determine how needs translate into wants. In this chapter we introduce a heterogeneous multinomial logit model for measuring the relationship between needs and wants. In contrast to the somewhat disappointing results in Chapter 3, we show that the use of the multinomial logit model leads to abundant evidence on the drivers of brand value.

4.1 Introduction

There are, in general, two ways that firms can make money. The first is to be a low-cost provider and attempt to generate profits by making an offering that appeals to the largest possible portion of the market. The second is to attempt to create a local monopoly and cater to a market niche. The first approach makes investments in production and operations competencies that provide a strategic cost advantage. The second approach makes investments in marketing to better understand where prospects are coming from so they will want the right to acquire and use a product. The goal of either approach is to provide greater value to the consumer, either by lowering price or by increasing the
utility of the offering. In this chapter, we will use the term “utility” to indicate the value, or preference for an offering. A more formal definition and connection to elements of our analysis is discussed in Chapter 5.

Choice is an extensively studied area in marketing, psychology and economics, and many approaches have been used to describe the choice process. A choice model begins with a process describing the manner in which product benefits combine to form an overall measure of value:

\[ y_i = \beta_{0,i} + a_{1,i}\beta_1 + a_{2,i}\beta_2 + \cdots + a_{k,i}\beta_k + \varepsilon_i \]

It helps to think of a specific person to better understand the value creation equation for a specific choice alternative \((i)\). So, let’s consider one respondent to our Florida vacations survey named Leah: (DO)

| Overall value: Given everything that Leah is considering about Magic Kingdom, this is what she thinks it is worth. |
| Attributes: These are the optional attributes that you’re studying and offering to Leah at various levels for your conjoint analysis (e.g., speed ticket, mobile app). |
| Intercept: This represents the essential attributes of the product that Leah expects. Without them, it just wouldn’t be Magic Kingdom (e.g., the rides she can enjoy and the Disney characters she can meet) |
| Importance: For each optional attribute, there is a level of importance that Leah has mentally assigned to it, and which you can estimate based upon her responses. |
| Error term: This represents everything about Leah’s decision process that you cannot observe from the data (e.g., her best friend lives in Northern Florida and has three children who love to go to Magic Kingdom whenever she visits). |

The subscript “\(i\)” indexes different product offerings, the vector \(a_i = (a_{1,i}, a_{2,i}, \cdots , a_{k,i})\)
represents the product attributes of the \(i^{th}\) product offering, and the vector \(\beta' = (\beta_1, \cdots, \beta_k)\) are weights associated with the attributes that reflect their importance to the decision maker. If a weight \(\beta_j\) is large, then the associated attribute \(a_{j,i}\) is given large weight in the decision process, and if the weight \(\beta_j\) is zero then the attribute does not play any role in arriving at the overall measure of value \(y_i\). The intercept \(\beta_{0,i}\) reflects the value ascribed to the brand that is not measured by the specified attributes \(a_{1,i} \cdots a_{k,i}\).

A simple choice model combines the model of value creation with a second equation describing choice, e.g.:

Choose \(i\) if \(y_i = \max \{y_1, y_2, \cdots, y_N, z\}\)

where \(z\) is a “none” choice option where the respondent declines to make a selection at this time.

Imagine yourself shopping for a new cell phone, confused as you consider all the possible features to choose from. Does it have to be an iPhone, or are you a Windows addict? Would you be willing to buy any brand, but you must have a large screen for watching videos and a keypad for texting? Will you only consider phones that are supported by your current cellular carrier? A conjoint experiment simulates the shopping experience by letting you compare different options and choosing the one you like best, or none at all. Your responses allow us to figure out which attributes are important to you, and how important they are relative to each other; from this information we can build a model of your preferences that allows us to predict how likely you are to choose a cell phone with a specific set of attributes, such as a touch screen Droid with a tiny screen. (CF-B)

Table 4.1 displays the most preferred Florida Vacation destination choices for the first respondent in the survey. We see that this respondent never selected Magic Kingdom, Epcot Center or Islands of Adventure as their most preferred choice in the 12 choice tasks presented to them. Variation present for the other theme parks is because of changing prices and product attributes. It is from these data, plus knowledge of the description of each choice alternative, that we will derive estimates for the coefficients.

The above equations are an example of a simple choice model, and alternative formulations have been proposed in various literatures. Modifications to the first equation might include moving away from an additive linear structure that implies a compensatory process where weak performance on one attribute can be compensated by strong performance on another. Modification to the second equation might include the use of
Table 4.1: Florida Vacation Data – First Respondent’s Most Preferred Choice

<table>
<thead>
<tr>
<th>Choice</th>
<th>Magic Kingdom</th>
<th>Epcot Center</th>
<th>Animal Kingdom</th>
<th>Hollywood Studios</th>
<th>Universal Studios</th>
<th>Islands of Adventure</th>
<th>Busch Gardens</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

budget constraints and other approaches that limit the set of offerings considered. We discuss extensions to the simple model in Chapter 5.

### 4.2 A Model Describing Discrete Choice

A popular choice model known as the multinomial logit model arises from the above equations when the error terms ($\zeta_i$) are assumed to be distributed according to an extreme value distribution (see Appendix C). This assumption results in an expression for the choice probability that takes on a relatively simple form:

$$
Pr_i = \frac{\exp[\beta_0 + a_1 \beta_1 + a_2 \beta_2 + \cdots + a_k \beta_k]}{1 + \sum_{j=1}^{N} \exp[\beta_0 + a_1 \beta_1 + a_2 \beta_2 + \cdots + a_k \beta_k]}
$$

$$
= \frac{\exp[V_i]}{1 + \sum_{j=1}^{N} \exp[V_j]}
$$
where \( V_i \) denotes the measured value of consuming one unit of an alternative and the “1” in the denominator of the fraction is the term associated with the outside good \( z \) (i.e., \( V_z = 0 \) and \( \exp[0] = 1 \)). Thus, the higher the measured value \( V_i \), the higher the probability of choice, \( \Pr(i) \). This expression is a probability because the error term \( (\varepsilon_i) \) in the expression \( y_i = V_i + \varepsilon_i \) represents information known by the decision maker and not observed by the analyst, i.e., unmeasured value. The formal connection between the decision model and the statistical model will be explored more fully in Chapter 5.

The table below illustrates calculations for computing choice probabilities with the multinomial logit model. I am selecting a bottle of wine for my friend Tracy who prefers a dry California wine. Columns C, D and E display the dummy variable coding for the product attributes. A score of one (1) indicates the presence of an attribute, while a score of zero indicates its absence. The first choice alternative is a 750 ml bottle of dry Italian wine, and the second is a 1.5 L Magnum sweet California. The right-most column is for the outside good. I decided to purchase the California wine, as you would have predicted with the model. (RW)

<table>
<thead>
<tr>
<th>A</th>
<th>Attributes</th>
<th>B</th>
<th>Betas</th>
<th>C</th>
<th>First Choice Alternative (a1)</th>
<th>D</th>
<th>Second Choice Alternative (a2)</th>
<th>E</th>
<th>Outside Good (z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>750 ml bottle</td>
<td>2</td>
<td></td>
<td>1</td>
<td></td>
<td>0</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.5 L Magnum</td>
<td>1</td>
<td></td>
<td>0</td>
<td></td>
<td>1</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Sweet taste</td>
<td>0</td>
<td></td>
<td>0</td>
<td></td>
<td>1</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Dry taste</td>
<td>2</td>
<td></td>
<td>1</td>
<td></td>
<td>0</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>California Wine</td>
<td>1</td>
<td></td>
<td>0</td>
<td></td>
<td>1</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Italian Wine</td>
<td>0</td>
<td></td>
<td>1</td>
<td></td>
<td>0</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Value Levels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Values</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Formulas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Probabilities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Values</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Formulas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Our investigation assumes that products can be defined in terms of a collection of product attributes and benefits. Figure 4.1 displays the attributes and levels for Ice Cream, and Figure 4.2 provides the list for Florida Vacations. Size and price are combined in the Ice Cream study for now to yield a price-per-serving statistic, allowing us to compare results to that reported later in Chapter 5. The speed ticket for the Florida
Vacations study allows you to save a place in line while you enjoy the rest of the theme park and return during a specific time window and enjoy minimal waiting. The mobile application allows you to see times and activities from wherever you may be in the park (e.g., wait times or character meetings).

Figure 4.1: Ice Cream Attributes and Levels.

<table>
<thead>
<tr>
<th>Brand</th>
<th>Flavor</th>
<th>Size</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dreyer’s</td>
<td>Vanilla</td>
<td>Pint (4 servings)</td>
<td>1.99</td>
</tr>
<tr>
<td>Blue Bunny</td>
<td>Vanilla Bean</td>
<td>Quart (8 servings)</td>
<td>2.49</td>
</tr>
<tr>
<td>Blue Bell</td>
<td>Chocolate</td>
<td>Half-gallon (16 servings)</td>
<td>2.99</td>
</tr>
<tr>
<td>Bigger</td>
<td>Cookies and Cream</td>
<td>3.49</td>
<td></td>
</tr>
<tr>
<td>Häagen-Dazs</td>
<td>Neapolitan</td>
<td></td>
<td>3.99</td>
</tr>
<tr>
<td>Ben &amp; Jerry’s</td>
<td>Vanilla Fudge Ripple</td>
<td>4.49</td>
<td></td>
</tr>
<tr>
<td>Storebrand</td>
<td>Oreo</td>
<td></td>
<td>4.99</td>
</tr>
<tr>
<td></td>
<td>Rocky Road</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chocolate Chip</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chocolate Chip Cookie Dough</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Conjoint analysis is used to estimate the coefficient vector $\beta$. A conjoint study involves providing a series of choices to respondents involving hypothetical product offerings and asking them to indicate their preferred alternative. As the product descriptions vary over the choice sets, variation in the product features ($a_i$) is generated so that the preference coefficients ($\beta$) can be measured through the observed choices. Dummy variables are used to describe the presence and absence of attribute levels.

An example of a conjoint choice for the Florida Vacations exercise is provided in Figure 4.3. In a typical conjoint study, respondents are asked to respond to at least ten choice exercises, and in this study respondents are asked to provide responses to 12 choice scenarios. Responses are provided in a dual format, where respondents are asked to indicate their preferred choice, and also asked to provide a volumetric response that we will analyze later in Chapter 5. For now, we will consider only responses to
4.2. A MODEL DESCRIBING DISCRETE CHOICE

Figure 4.2: Florida Vacations Attributes and Levels.

<table>
<thead>
<tr>
<th>Brand</th>
<th>Speed Ticket</th>
<th>Mobile App</th>
<th>Extended Hours</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magic Kingdom</td>
<td>No</td>
<td>No</td>
<td>Regular</td>
<td>35</td>
</tr>
<tr>
<td>Epcot</td>
<td>Yes</td>
<td>Yes</td>
<td>2 Hours Early</td>
<td>40</td>
</tr>
<tr>
<td>Animal Kingdom</td>
<td></td>
<td></td>
<td>2 Hours Late</td>
<td>45</td>
</tr>
<tr>
<td>Hollywood Studios</td>
<td></td>
<td></td>
<td></td>
<td>50</td>
</tr>
<tr>
<td>Universal Studios</td>
<td></td>
<td></td>
<td></td>
<td>55</td>
</tr>
<tr>
<td>Islands of Adventure</td>
<td></td>
<td></td>
<td></td>
<td>60</td>
</tr>
<tr>
<td>Busch Gardens</td>
<td></td>
<td></td>
<td></td>
<td>65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70</td>
</tr>
</tbody>
</table>

the question “Which single option do you most prefer?” To the right of the figure is a choice option allowing the respondent to indicate that none of the displayed options are attractive to them, and that they would rather not purchase any. The “non-choice” option corresponds to choosing the outside good, \( z \), which we assign a value of zero, i.e., \( V_z = 0 \). This normalization is necessary so that we can derive unique coefficient estimates \( \hat{\beta} \) for the attribute-levels.

Conjoint models allow us to quantify consumer preferences for brands and features. In doing so, there has to be a base value, or benchmark, that can be used to find the relative value of the choice alternatives. The “outside good” is used for this purpose. Think of the outside good as the sea level, which is assigned a value of zero elevation. Everything else is measured relative to this. Mt. Everest, the world’s highest peak, is 29,000 feet above sea level and Death Valley, located in Eastern California, is the world’s lowest place at 282 feet below sea level. Sea level gives us a way to measure the relative elevation of a place, just as the outside good allows us to measure relative preference. (AM)
### Park Vacation Package Assumption

- 7 day vacation
- 2 adults
- 2 kids

### Scenario 1 of 12

<table>
<thead>
<tr>
<th>Theme Park</th>
<th>Premium speed ticket at theme park</th>
<th>Mobile app for wait times, character meetings, etc.</th>
<th>Special extended hours</th>
<th>Theme park daily entrance fee (per person)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disney</td>
<td>✓</td>
<td>✓</td>
<td>2 hours early</td>
<td>$40</td>
</tr>
<tr>
<td>Universal Studios</td>
<td>✓</td>
<td>✓</td>
<td>Regular</td>
<td>$50</td>
</tr>
<tr>
<td>Epcot</td>
<td>✓</td>
<td></td>
<td>1 hour early or 1 hour late</td>
<td>$35</td>
</tr>
<tr>
<td>Islands of Adventure</td>
<td>✓</td>
<td></td>
<td>2 hours late</td>
<td>$55</td>
</tr>
<tr>
<td>Kingdoms</td>
<td>✓</td>
<td></td>
<td>Regular</td>
<td>$40</td>
</tr>
<tr>
<td>Animal Kingdom</td>
<td>✓</td>
<td></td>
<td>2 hours late</td>
<td>$45</td>
</tr>
</tbody>
</table>

Would not go to any of these parks

**Which single option do you most prefer?**

**How many days will you attend each park during your stay?**

| Days | 0 | Days | 0 | Days | 0 | Days | 0 | Days | 0 | Days | 0 |

**[CALCULATE] Total Price**

**Cost(s) Summary**

Total cost of theme parks (2 adults & 2 kids) $ ___ ___
Tradeoff Analysis

Product analysis involves identifying the combination of attributes and benefits that drives value for a particular market target. It does this by combining knowledge about how people make choices with what we know about preference heterogeneity, described below, to arrive at an optimal product design. In product design, it is the attributes and benefits that are being manipulated, while in price and advertising analysis other variables are being manipulated to maximize sales. We define attributes and benefits as the physical and psychological features of an offering that makes it a useful instrument for effecting change in a person’s life. The regression coefficients $\beta$ indicate the value of these product features.

The value of relating product attributes to a single measure of preference is that it provides a way of relating the attributes to each other. Tradeoff analysis is a process where the level of value ($V_i$) is held fixed, and the analysis considers ways in which the attribute levels can be substituted to arrive at the same fixed level. Consider two products that generate the exact same level of value by modifying the levels of just two of the attributes. That is, the other attributes for these two products are identical:

$$y_i = \beta_0 + a_{1,i}\beta_1 + a_{2,i}\beta_2 + \cdots + a_{2,i}^*\beta_3 + \cdots + \varepsilon_i$$

which implies that

$$a_{1,i}\beta_1 + a_{2,i}\beta_2 = a_{1,i}^*\beta_1 + a_{2,i}^*\beta_2$$

or

$$\left(a_{1,i} - a_{1,i}^*\right) = -\left(a_{2,i} - a_{2,i}^*\right)\beta_2/\beta_1$$

This last equation provides an explicit formula for measuring the impact of changing the level of one attribute in terms of another attribute. When attribute levels are coded using dummy variables, the lowest level of the attribute is often coded as the null level, and attribute enhancements are coded with a “1”. In this case, the ratio $-\beta_2/\beta_1$ provides a direct measure of the value of moving from the null level to a higher level of one attribute in terms of a second attribute.

Warning: Compensatory models are not always valid. If your child is allergic to nuts, for example, then no matter how tasty, organic, or low calorie a snack is, if it contains nuts, you won’t buy it. In this case, strong performance on another attribute cannot compensate for the attribute “Contains Nuts.” (RL)
In many conjoint studies, price is treated as if it were another attribute and trade-off analysis is used to compute the monetary equivalent of changes in the levels of an attribute. The practice of treating price as a product attribute should be viewed as an approximation to a treatment based on first-order conditions, which we develop in more detail in Chapter 5 on pricing analysis. In general, this practice will work when choices are discrete and involve just one unit (i.e., the quantity of consumption is not of particular interest). When consumers purchase multiple quantities over time, as is the case in product categories such as ice cream and Florida vacation purchases, price needs to be more formally entered into the model.

Consider this situation: two fifth graders are sitting on one end of a teeter totter. Your goal is to perfectly balance the teeter totter using the rest of the students on the playground. Perhaps you discover that two third graders and a fourth grader provide perfect balance. But you could also use three first graders and two second graders. Or maybe a fifth grader, a third grader, and a first grader would work. There are many combinations of children that could balance the two fifth graders.

This is tradeoff analysis. Just as fifth graders have a larger effect than first graders, the most important attributes of a brand have a larger effect than those of lesser importance. Examining this balancing act and the possible combinations helps us see the levels of tradeoffs between each attribute, giving us an understanding of what the consumer values most. But keep in mind, while having strong performance on the most important attributes has a large effect, there are also numerous combinations involving all attributes that will have the exact same positive effect.

4.3 Heterogeneity

If the attributes \(a\) and importance weights \(\beta\) of a brand are known, then they can be combined to obtain an estimate of the value \(V_i\) and estimates of choice probabilities. A challenge in doing this is in obtaining reliable estimates of the importance weights \(\beta\) because it is rare to have more than a few dozen observations (data points) from a specific respondent when doing analysis in marketing. After a dozen or so choice tasks a respondent becomes fatigued and ceases to provide very reliable information about their preferences in a survey. Similarly, it is rare to have more than a few dozen observations...
of an individual making purchases in a product category before there is a new entrant in the market that makes preference estimates change. The limited amount of respondent information, coupled with the presence of heterogeneity in both needs and wants, presents a challenge to marketing analysis that has only recently been addressed with the advent of modern statistical techniques (see Appendix C).

Sophia wants a sofa for her apartment, but she is overwhelmed by the options: leather or cloth; do-it-yourself or pre-assembled; patterned or solid; Scotch-guarded or unprotected; a sectional or separate pieces. Ideally, an online recommendation system would help her consider the attributes of the furniture that Sophia values and determine importance of each feature to her so that it could help navigate through the options. It would also be important to know why Sophia has the preferences she maintains. Does she want Scotchgard because her toddler niece visits her often?

Heterogeneity is important for market segmentation. Because people are different, they have different preferences that bring them to the market. In our model of value creation, the value $y$ of a sofa is dependent on the attributes $a$ of the sofa and the importance Sophia attaches to the attributes $\beta$. To maximize benefit to Sophia, the recommendation system needs to know both the product attributes $a$ and importances $\beta$ in order to make good recommendations. (KW)

Heterogeneity was discussed previously in the chapter on Market Segmentation. People are heterogeneous in the motivations that bring them to a market, and are also heterogeneous in the value they place on the attributes and benefits of product features that comprise an offering. Every product offering is composed of one or more product benefits, and the fact that consumers differ in their desire for them is what leads to the creation of market targets and market niches. In terms of the standard regression model we add the subscript “$h$” to refer to regression coefficients that are “household” or respondent-specific. For our additive model of value creation we have:

$$y_{i,h} = \beta_{0,i,h} + \beta_{1,i}a_{1,i} + \beta_{2,i}a_{2,i} + \cdots + \beta_{k,i}a_{k,i} + \varepsilon_{i,h}$$

We now add a second equation that describes the variation of $\beta$ coefficients among the respondents. This second equation describes across-household variation in the logit model coefficients $\beta_{h}$, whereas the first equation describes within-household variation in choices. A simple model of heterogeneity is the random-effects model:
\[
\beta_h \sim \text{Normal}(\mu, \Sigma)
\]

where \(\beta_h\) is a vector of coefficients, \(\mu\) is the mean of the random-effect distribution and \(\Sigma = [\sigma_{ij}]\) is the covariance matrix of the random effects. The diagonal elements of the covariance matrix \(\Sigma\), \(\{\sigma_{ii}\}\), describe the amount of heterogeneity, or variance, of each element of the \(\beta_h\) vector among respondents. Since the diagonal elements are estimates of the variance, the square root of each element is the standard deviation, and \(\pm 2\sqrt{\sigma_{ii}}\) provides a 95\% credible interval of the respondent-level coefficients. The off-diagonal elements of \(\Sigma\), \(\{\sigma_{ij}\}\), measure the covariation of pair-wise elements of the \(\beta\) vector. Positive values of covariation indicate the co-occurrence of elements that are above their mean, and negative covariation indicates that when one element is above average the other is below average. The correlation coefficient is defined as:

\[
\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}} \sqrt{\sigma_{jj}}}
\]

and is reported in *italics* below the diagonal entries.

The heterogeneity of consumers in the marketplace has huge implications on product development. Instead of developing products tailored to meet the needs of the mean of all consumers, good marketers focus on the extremes of consumers and tailor products to meet their specific needs. Consider Adam, fresh off two long years of business school. Adam decides that he is going use his new MBA to open a small candle shop in his small hometown. To help guide what types of candles he will sell, he conducts a survey of 1,500 people in his hometown and finds that the average consumer purchases 14.5 ounce wax candles in the scents of vanilla, lavender, and cinnamon. “Perfect,” thinks Adam, “to appeal to as many consumers as possible, Ill produce candles that are favored by the average consumer: 14.5 ounces, wax, in the scents of vanilla, lavender, and cinnamon.”

So, Adam makes his first batch of candles, prices them comparable to other wax candles in the market, and opens his new store. After a few months, however, his sales are well below his predictions. So, what happened? Adam forgot about the heterogeneity of consumers in the marketplace and instead developed a product that appeals to no one in particular. Instead, he should have instead focused on the extremes, tailoring his products to meet the needs of one or more segments of the market. (AW)
4.3. HETEROGENEITY

Figure 4.4 displays estimates of the mean ($\mu$) and covariance matrix ($\Sigma$) of the random effects logit model estimated with the Ice Cream data. The first column of numbers are estimates of the mean of the distribution of $\beta_h$ among respondents, and the remaining columns are for the covariance matrix of $\beta_h$ among respondents. The diagonal elements of the covariance matrix are shaded to improve readability. The first shaded entry of the covariance matrix is the variance of random effects for Dreyer’s ice cream. The variance is equal to 8.50, implying that the standard deviation of random-effects is $\sqrt{8.50} = 2.92$, or that a 95% credible interval of individual-level coefficients ranges from

\[
(\mu - 2\sigma, \mu + 2\sigma) = (-0.52 - 2\sqrt{8.50}, -0.52 + 2\sqrt{8.50})
\]
or
\[
(-5.89, 5.32)
\]
indicating large respondent heterogeneity for the Dreyer’s ice cream brand. Positive values of the brand coefficients indicate preference relative to the outside good $z$, the non-choice option. Likewise, positive values of the attributes indicate preference relative to the default flavor that is selected to be vanilla.

Figure 4.4: Ice Cream Coefficients ($\mu, \Sigma$): Random-Effects Logit Model

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Dry</th>
<th>BB</th>
<th>BBun</th>
<th>HD</th>
<th>Bry</th>
<th>B&amp;J</th>
<th>SB</th>
<th>VB</th>
<th>Choc</th>
<th>C&amp;C</th>
<th>Neo</th>
<th>VFR</th>
<th>Oreo</th>
<th>RR</th>
<th>CC</th>
<th>CCCD</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dreyer’s</td>
<td>-0.52</td>
<td>8.5</td>
<td>6.05</td>
<td>6.15</td>
<td>5.32</td>
<td>6.16</td>
<td>4.55</td>
<td>5.96</td>
<td>1.19</td>
<td>-1.76</td>
<td>-0.41</td>
<td>-0.29</td>
<td>0.37</td>
<td>-1.14</td>
<td>-0.63</td>
<td>-1.62</td>
<td>-2.19</td>
<td>-1.21</td>
</tr>
<tr>
<td>Blue Bell</td>
<td>-1.05</td>
<td>0.58</td>
<td>32.61</td>
<td>7.85</td>
<td>4.15</td>
<td>4.76</td>
<td>3.46</td>
<td>7.29</td>
<td>1.08</td>
<td>-1.73</td>
<td>-1.07</td>
<td>-0.02</td>
<td>0.27</td>
<td>-1.56</td>
<td>-2.79</td>
<td>-2.68</td>
<td>-2.72</td>
<td>-1.72</td>
</tr>
<tr>
<td>Blue Bunny</td>
<td>-0.78</td>
<td>0.66</td>
<td>0.69</td>
<td>10.31</td>
<td>4.15</td>
<td>4.99</td>
<td>3.77</td>
<td>6.56</td>
<td>1.42</td>
<td>-0.92</td>
<td>-0.75</td>
<td>0.39</td>
<td>0.59</td>
<td>-0.91</td>
<td>-1.18</td>
<td>-1.48</td>
<td>-1.92</td>
<td>-1.58</td>
</tr>
<tr>
<td>Haagen-Dazs</td>
<td>-0.02</td>
<td>0.66</td>
<td>0.42</td>
<td>0.47</td>
<td>7.65</td>
<td>4.67</td>
<td>6.02</td>
<td>3.58</td>
<td>1.47</td>
<td>-0.93</td>
<td>-0.01</td>
<td>-0.11</td>
<td>0.69</td>
<td>-0.05</td>
<td>-1.25</td>
<td>-0.45</td>
<td>-1.19</td>
<td>-0.82</td>
</tr>
<tr>
<td>Breyers</td>
<td>0.41</td>
<td>0.73</td>
<td>0.46</td>
<td>0.53</td>
<td>0.58</td>
<td>8.45</td>
<td>4.21</td>
<td>5.6</td>
<td>1.66</td>
<td>-1.55</td>
<td>-0.54</td>
<td>0.12</td>
<td>0.2</td>
<td>-0.88</td>
<td>-1.35</td>
<td>-0.99</td>
<td>-2.11</td>
<td>-0.98</td>
</tr>
<tr>
<td>Ben &amp; Jerry’s</td>
<td>0.27</td>
<td>0.59</td>
<td>0.37</td>
<td>0.45</td>
<td>0.83</td>
<td>0.55</td>
<td>6.95</td>
<td>2.98</td>
<td>0.94</td>
<td>-0.99</td>
<td>0.35</td>
<td>0.16</td>
<td>0.99</td>
<td>0.44</td>
<td>-0.28</td>
<td>-0.05</td>
<td>0.00</td>
<td>-0.65</td>
</tr>
<tr>
<td>Store Brand</td>
<td>-0.66</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
</tr>
<tr>
<td>Vanilla Bean</td>
<td>-0.23</td>
<td>0.27</td>
<td>0.2</td>
<td>0.29</td>
<td>0.35</td>
<td>0.37</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>Chocolate</td>
<td>-0.67</td>
<td>-0.23</td>
<td>-0.19</td>
<td>-0.11</td>
<td>-0.13</td>
<td>-0.21</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.14</td>
</tr>
<tr>
<td>Cookies and Cream</td>
<td>-0.6</td>
<td>-0.05</td>
<td>-0.11</td>
<td>-0.08</td>
<td>0</td>
<td>-0.06</td>
<td>0.05</td>
<td>-0.12</td>
<td>-0.38</td>
<td>0.11</td>
<td>3.23</td>
<td>1.54</td>
<td>1.03</td>
<td>7.1</td>
<td>3.99</td>
<td>3.07</td>
<td>6.96</td>
<td>-0.61</td>
</tr>
<tr>
<td>Neapolitan</td>
<td>-1.53</td>
<td>-0.04</td>
<td>0</td>
<td>0.05</td>
<td>-0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>0.05</td>
<td>-0.23</td>
<td>0.45</td>
<td>0.2</td>
<td>7.03</td>
<td>1.31</td>
<td>1.89</td>
<td>3.41</td>
<td>1.27</td>
<td>-0.29</td>
<td></td>
</tr>
<tr>
<td>Vanilla Fudge Ripple</td>
<td>-0.48</td>
<td>0.05</td>
<td>0.03</td>
<td>0.07</td>
<td>0.1</td>
<td>0.03</td>
<td>0.15</td>
<td>-0.05</td>
<td>-0.21</td>
<td>0.44</td>
<td>0.14</td>
<td>0.19</td>
<td>6.65</td>
<td>1.16</td>
<td>5.01</td>
<td>3.25</td>
<td>2.51</td>
<td>0.03</td>
</tr>
<tr>
<td>Oreo</td>
<td>-0.98</td>
<td>-0.14</td>
<td>-0.16</td>
<td>-0.1</td>
<td>-0.01</td>
<td>-0.11</td>
<td>0.06</td>
<td>-0.14</td>
<td>-0.41</td>
<td>0.24</td>
<td>0.24</td>
<td>0.2</td>
<td>7.03</td>
<td>1.31</td>
<td>1.89</td>
<td>3.41</td>
<td>1.27</td>
<td>-0.29</td>
</tr>
<tr>
<td>Rocky Road</td>
<td>-0.86</td>
<td>-0.06</td>
<td>-0.21</td>
<td>-0.1</td>
<td>-0.12</td>
<td>-0.13</td>
<td>-0.03</td>
<td>-0.16</td>
<td>-0.26</td>
<td>0.35</td>
<td>0.37</td>
<td>0.35</td>
<td>0.52</td>
<td>0.36</td>
<td>13.76</td>
<td>5.07</td>
<td>4.95</td>
<td>0.05</td>
</tr>
<tr>
<td>Chocolate Chip</td>
<td>-0.44</td>
<td>-0.22</td>
<td>-0.3</td>
<td>-0.18</td>
<td>-0.06</td>
<td>-0.13</td>
<td>-0.04</td>
<td>-0.22</td>
<td>-0.34</td>
<td>0.56</td>
<td>0.42</td>
<td>0.19</td>
<td>0.5</td>
<td>0.44</td>
<td>0.54</td>
<td>1.45</td>
<td>0.43</td>
<td>-0.01</td>
</tr>
<tr>
<td>Chocolate Chip Cookie Dough</td>
<td>-0.56</td>
<td>-0.22</td>
<td>-0.23</td>
<td>-0.18</td>
<td>-0.13</td>
<td>-0.22</td>
<td>0</td>
<td>-0.29</td>
<td>-0.44</td>
<td>0.13</td>
<td>0.72</td>
<td>0.2</td>
<td>0.29</td>
<td>0.63</td>
<td>0.4</td>
<td>0.53</td>
<td>11.33</td>
<td>-0.18</td>
</tr>
<tr>
<td>Price</td>
<td>-0.92</td>
<td>-0.34</td>
<td>-0.39</td>
<td>-0.4</td>
<td>-0.24</td>
<td>-0.27</td>
<td>-0.2</td>
<td>-0.41</td>
<td>-0.18</td>
<td>0.06</td>
<td>-0.17</td>
<td>-0.09</td>
<td>0.01</td>
<td>-0.1</td>
<td>0.01</td>
<td>-0.04</td>
<td>1.51</td>
<td></td>
</tr>
</tbody>
</table>

Covariances ($\sigma_{i,j}$) are reported above the diagonal, Correlations ($\rho_{i,j}$) below the diagonal.

The table is divided into four regions – two for the mean ($\mu$) and two for the covariance matrix ($\Sigma$). The upper regions for the elements of $\beta_h$ are related to the brand intercepts, and the lower portion of the table corresponds to the attributes. That is,
In general, we find large diagonal entries for the covariance matrix, indicating that there is a large amount of heterogeneity of respondent preferences for the ice cream brands and flavors. Brand coefficients are, on average, positive for Breyers and Ben & Jerry’s, indicating the presence of some positive brand equity relative to the outside good. The brand preference estimates are different from the Pick any/J data in Chapter 1 (see Table 1.7) in that the preference estimates control for the effects of price and flavors. We find that the flavor coefficients are all negative, indicating that the default flavor (vanilla) is preferred, on average, above all other flavors. Finally, we find moderate positive covariances among all the brands, implying that as a respondent tends to like one of the brands more than average, they also tend to like the other brands too. Thus we can infer the existence of preference patterns similar to that encountered in Chapter 1 (Figure 1.3). For example, the covariance between the brands Haagen-Dazs and Ben & Jerry’s is 6.02, corresponding to a correlation of:

\[ \rho_{4,6} = \frac{\sigma_{4,6}}{\sqrt{\sigma_{4,4}\sigma_{6,6}}} = \frac{6.02}{\sqrt{7.65}\sqrt{6.95}} = 0.83 \]

which is large. We also found that these brands mapped close to each other when conducting analysis in Chapter 1.

A tradeoff analysis conducted on the mean coefficients in Figure 4.4 helps to interpret the results. We find that the average brand coefficient for Ben & Jerry’s is 0.27 while the average brand coefficient for Haagen-Dazs is -0.02, on average. We also find that the price coefficient is -0.92. This implies that the monetary equivalent for these brands relative to the outside good is:

Ben & Jerry’s: \[ \frac{-0.27}{-0.92} = $0.29 \] and Haagen-Dazs: \[ \frac{-0.02}{-0.92} = -$0.02 \]

with the difference in brand coefficients equal to $0.31. That is, consumers are willing to pay an average of 31 cents more per serving for the Ben & Jerry’s brand than the Haagen-Dazs brand for a similar flavor.

Figure 4.5 displays estimates of the mean \((\mu)\) and covariance matrix \((\Sigma)\) of the random-effects model estimated with the Florida Vacations data. Similar to the ice cream results, we find large diagonal elements of the covariance matrix, indicating large heterogeneity in the brand preferences. Covariances are smaller than those found earlier.
with ice cream, with the implied correlation coefficients small as well. For example, the correlation coefficient between the Magic Kingdom and Epcot brand coefficients is:

$$\rho_{1,2} = \frac{3.48}{\sqrt{18.63} \sqrt{12.58}} = 0.28$$

which is much smaller than the correlations measured in Chapter 1. This difference can be attributed to many factors, such as past attendance over a five year period (Question S9 in the Florida Vacations survey) reflecting an accumulation of choices and the coefficients in Figure 4.5 pertaining to a specific choice.

Figure 4.5: Florida Vacations Coefficients ($\mu, \Sigma$): Random-Effects Logit Model

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Mean</th>
<th>MK</th>
<th>Ep</th>
<th>AK</th>
<th>HS</th>
<th>US</th>
<th>IA</th>
<th>BG</th>
<th>Speed</th>
<th>Mobile</th>
<th>Early</th>
<th>Late</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magic Kingdom</td>
<td>7.25</td>
<td>18.63</td>
<td>3.48</td>
<td>2.54</td>
<td>-0.88</td>
<td>-4.1</td>
<td>-3.61</td>
<td>0.04</td>
<td>0.13</td>
<td>0.27</td>
<td>0.41</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Epcot</td>
<td>6.69</td>
<td>0.23</td>
<td>12.58</td>
<td>3.58</td>
<td>1.83</td>
<td>-0.72</td>
<td>-6.63</td>
<td>-0.34</td>
<td>-0.09</td>
<td>-0.91</td>
<td>-0.29</td>
<td>-0.64</td>
<td>-0.01</td>
</tr>
<tr>
<td>Animal Kingdom</td>
<td>5.24</td>
<td>0.26</td>
<td>0.28</td>
<td>12.6</td>
<td>3.77</td>
<td>0.03</td>
<td>-1.41</td>
<td>3.11</td>
<td>0.61</td>
<td>0.15</td>
<td>-0.19</td>
<td>-0.03</td>
<td></td>
</tr>
<tr>
<td>Hollywood Studios</td>
<td>4.89</td>
<td>0.24</td>
<td>0.21</td>
<td>0.43</td>
<td>6.1</td>
<td>3.04</td>
<td>-1.56</td>
<td>2.08</td>
<td>-0.11</td>
<td>0.15</td>
<td>0.21</td>
<td>-0.02</td>
<td></td>
</tr>
<tr>
<td>Universal Studios</td>
<td>5.72</td>
<td>-0.08</td>
<td>-0.08</td>
<td>0.04</td>
<td>6.34</td>
<td>-0.84</td>
<td>2.07</td>
<td>-0.2</td>
<td>0.29</td>
<td>0.24</td>
<td>-0.03</td>
<td>-0.02</td>
<td></td>
</tr>
<tr>
<td>Islands of Adventure</td>
<td>4.5</td>
<td>-0.25</td>
<td>-0.5</td>
<td>-0.11</td>
<td>-0.17</td>
<td>-0.09</td>
<td>14.13</td>
<td>-2.31</td>
<td>0.83</td>
<td>0.01</td>
<td>0.44</td>
<td>1.07</td>
<td>0.01</td>
</tr>
<tr>
<td>Busch Gardens</td>
<td>4.54</td>
<td>-0.23</td>
<td>-0.03</td>
<td>0.24</td>
<td>0.24</td>
<td>0.23</td>
<td>-0.17</td>
<td>12.83</td>
<td>0.25</td>
<td>0.02</td>
<td>-0.09</td>
<td>-0.04</td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td>0.47</td>
<td>0.14</td>
<td>-0.03</td>
<td>0.2</td>
<td>-0.05</td>
<td>-0.09</td>
<td>0.25</td>
<td>0.08</td>
<td>0.77</td>
<td>0.2</td>
<td>0</td>
<td>0.07</td>
<td>-0.01</td>
</tr>
<tr>
<td>Mobile</td>
<td>0.41</td>
<td>0.04</td>
<td>0.04</td>
<td>0.06</td>
<td>0.16</td>
<td>0</td>
<td>0.01</td>
<td>0.32</td>
<td>0.5</td>
<td>0.02</td>
<td>0.03</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Early</td>
<td>0.2</td>
<td>0.08</td>
<td>-0.11</td>
<td>0.06</td>
<td>0.08</td>
<td>0.13</td>
<td>0.15</td>
<td>0.15</td>
<td>0</td>
<td>0.04</td>
<td>0.58</td>
<td>0.14</td>
<td>0</td>
</tr>
<tr>
<td>Late</td>
<td>0.21</td>
<td>0.12</td>
<td>-0.23</td>
<td>-0.07</td>
<td>0.11</td>
<td>-0.02</td>
<td>0.36</td>
<td>-0.03</td>
<td>0.1</td>
<td>0.05</td>
<td>0.24</td>
<td>0.61</td>
<td>0</td>
</tr>
<tr>
<td>Price</td>
<td>0.04</td>
<td>0.01</td>
<td>0.02</td>
<td>-0.05</td>
<td>-0.05</td>
<td>-0.05</td>
<td>0.02</td>
<td>-0.06</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Covariances ($\sigma_{i,j}$) are reported above the diagonal, Correlations ($\rho_{i,j}$) below the diagonal.

The brand coefficients are an order of magnitude larger than the attribute coefficients in the table, indicating heightened preference relative to the outside good. That is, the outside good is rarely chosen in the data. The attributes for the premium speed ticket and the mobile app for wait times have coefficients of 0.47 and 0.41, indicating that they can influence choices among the theme parks because their magnitude is similar to the differences among many of the parks. The price coefficient is, on average, equal to -0.04. This implies that these attributes have a dollar equivalent of:

Speed Ticket: $$\frac{0.47}{0.04} = \$11.75$$  and  Mobile App: $$\frac{0.41}{0.04} = \$10.25$$
and that early and late admission into the park is worth approximately $5.00. The prices considered in the conjoint task ranged from $35 to $70, and multiplying this difference by the price coefficient results in price affecting the utility by a maximum amount of $35 \times -0.04 = 1.4$. This value is enough to affect the choice of which theme park to visit (as reflected in the differences among brand intercepts), but not enough to price someone out of the market so that the outside good ($z$) is chosen instead.

**Needs**

There are often many beliefs people have about a brand, and many reasons for wanting a brand. Some of these reasons pertain to physical attributes of a brand, others are related to the psychological formulation developed through advertising. Most of the psychological attributes are difficult to study in a choice model because people cannot conceive of particular brands with different psychological attributes. For example, Heineken beer is from Holland and it positions itself as a premium Dutch beer. It is difficult to conceive of a Heineken that is not from Holland. It is also difficult to conceive of it in a bottle that is anything other than the color green. Country-of-origin and other distinguishing features of a brand are not realistically changed in a conjoint choice experiment. As a result, it is better to measure the influence of needs on the brand intercepts using cross-section analysis. Moreover, since the number of different beliefs held by consumers about a brand is often large, an initial screening for important drivers (see Chapter 3) is recommended prior to specifying the random-effect model.

| Brands are built over time and it is important to recognize which attributes belong in a conjoint choice experiment. Psychological attributes can lead to strong associations, and sometimes it is difficult to separate the attribute from the brand. Case in point: brand extension blunders. Brand-extensions-gone-bad make for some of the most entertaining articles found online today and leave most of us scratching our heads wondering “What were they thinking?” Cheetos Lip Balm, Smith & Wesson mountain bikes and Colgate Kitchen Entrees are all great examples of how strong brands can fail. As a product manager, it is important for you to know the attributes of your brand and distinguish those that are psychological from the ones that are physical. You want to make sure that you only include those that can be changed when running your conjoint choice experiment. Otherwise, you'll be just as confused with your survey results as the respondents in your survey. (EW) |
The challenge in moving from a “coming from” orientation in market segmentation to a “going to” orientation in product optimization is that people are heterogeneous in their beliefs about the product features that are expected to be responsive to their needs. A high-horsepower engine may be seen as necessary for freeway driving by young drivers while being seen as wasteful by older drivers. Breath freshening may need to be long-lasting for people with heartburn or indigestion, but may need to provide a short burst of flavor for young people on a date.

The words “need” and “want” have different meanings depending on the context in which they are used. You may be tempted to think of a need in the Maslovian sense, as in, what is absolutely necessary to survive. In some cases, we might think of a need as something we feel we just cannot live without, a very intense want. A need can sometimes mean something evaluative, as in “I know what that person really needs.”

For purposes of this book, and especially this chapter, needs and wants have a different meaning. A need is a condition that motivates a person to search for a solution in the marketplace. For example, a mother may be looking for ways to spend quality time with her children. This is a need for her, and it describes where she is coming from. The want associated with this need might be a membership to the local zoo, buying a new Nintendo Wii, or a night out at an Italian restaurant.

In principle, both the efficacy of product attributes ($\beta_h$) and their association to the needs ($n_h$) of an individual are heterogeneous. However, we can expand our model of heterogeneity to include needs as covariates, allowing for deviations from the average relationship using the random-effects error term:

$$\beta_h = \Gamma' n_h + \zeta_h \quad ; \quad \zeta_h \sim Normal (0, \Sigma)$$

where $\Gamma$ is a matrix of coefficients. The relationship between the vector of coefficients $\beta_h$ and the vector of need-state covariates $n_h$ is captured by the coefficient matrix $\Gamma$, where the $(j,k)$ element of the matrix can be interpreted as the expected increase in the $k^{th}$ element of $\beta_h$ for a unit increase in the $j^{th}$ element of $n_h$. That is, the coefficient matrix $\Gamma$ provides the map from needs to wants.
Bruce Banner was a brilliant nuclear physicist who, like his father, was committed
to using his scientific acumen to help the greater good. One day, Bruce was hit by
powerful Gamma rays while trying to rescue an innocent teenager who had wan-
dered into a nuclear testing facility. Suddenly, do-gooder Bruce Banner became
The Incredible Hulk – strong, powerful fighter of evil and injustice everywhere.

What does Bruce Banner have to do with this book? Gamma (Γ)! As marketers,
we’re pretty good at identifying needs that might potentially impact behavior.
But, with the introduction of Gamma, we can be much better – we can become
marketing superheroes! Gamma allows us to power through multiple combinations
of need states (2^12 or 4,096 to be precise) as easily as the Hulk bursts through a
brick wall. Gamma provides the mapping from these needs to wants and indicates
the drivers of brand and attribute preference, something we didn’t really do before
becoming marketing superheroes. We can use the magnitude and uniqueness of
Gamma to determine key attributes to use to accurately and expertly position our
brand. We were good before, but now we have Gamma which allows us to do so
much more than we previously could - just like Bruce Banner and The Incredible
Hulk. (MP)

Figure 4.6 is the estimated coefficient matrix for the Ice Cream data using Q9 in
the Ice Cream survey for \( z_h \). Along the top of the matrix are listed the elements of
the \( \beta_h \) vector, and along the side are the 12 needs collected in the Pick any/J format
and represented as dummy variables. Coefficients are highlighted in blue if they are
“significant” in the sense that the estimate is more than two standard deviations from
zero. Thus the coefficient 0.85 in the first column of the table is the expected increase in
the value of the brand coefficient for Dreyer’s ice cream for individuals who responded
“yes” to the motivation “Ice cream is a wholesome treat” relative to those who responded
“no.”

Many of the coefficients in Table 4.6 are significant and large in magnitude, indicating
that the matrix \( \Gamma \) effectively translates how needs and wants interact, on average. Positive
entries in the figure increase the \( \beta_h \) coefficient by the indicated amount for respondents
reporting that they purchase ice cream for the indicated reason, and negative entries
decrease the \( \beta_h \) coefficient. The \( \Gamma \) matrix therefore indicates conditions in which specific
brands and specific flavors are expected to be highly valued. For example, Dreyer’s ice
cream is more valued by individuals looking for a wholesome treat as well as individuals
who feel that ice cream is good for entertaining (Need 11). Recalling that the average
price coefficient in Figure 4.4 is -0.92, we can re-interpet the magnitude of the “wholesome
4.3. HETEROGENEITY

Figure 4.6: Ice Cream Regression Coefficients (Γ): Needs (Q9)

<table>
<thead>
<tr>
<th>Q9</th>
<th>Dry</th>
<th>BB</th>
<th>BBun</th>
<th>HD</th>
<th>Bry</th>
<th>B&amp;J</th>
<th>SB</th>
<th>VB</th>
<th>Choc</th>
<th>C&amp;C</th>
<th>Neo</th>
<th>VF</th>
<th>Oreo</th>
<th>RR</th>
<th>CC</th>
<th>CCCD</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ice cream helps me relax and enjoy life.</td>
<td>0.54</td>
<td>-0.03</td>
<td>-0.14</td>
<td>0.33</td>
<td>0.08</td>
<td>0.45</td>
<td>-0.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Ice cream is a wholesome treat.</td>
<td>0.85</td>
<td>0.32</td>
<td>1.3</td>
<td>0.12</td>
<td>0.85</td>
<td>0.2</td>
<td>0.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Ice cream gives me something fun to do.</td>
<td>0.69</td>
<td>0.88</td>
<td>0.65</td>
<td>0.82</td>
<td>0.63</td>
<td>0.35</td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Ice cream provides relief from regular life.</td>
<td>-0.25</td>
<td>0.03</td>
<td>-0.22</td>
<td>-0.05</td>
<td>-0.16</td>
<td>0.06</td>
<td>-0.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Ice cream helps make an event special.</td>
<td>0.41</td>
<td>0.37</td>
<td>0.45</td>
<td>0.31</td>
<td>0.4</td>
<td>0.28</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Ice cream reminds me of my childhood.</td>
<td>-0.09</td>
<td>0.27</td>
<td>-0.26</td>
<td>0.73</td>
<td>0.49</td>
<td>0.69</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Ice cream is ice cream.</td>
<td>0.94</td>
<td>0.42</td>
<td>1.41</td>
<td>0.27</td>
<td>1.1</td>
<td>0</td>
<td>1.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. I am amazed at how tastes created.</td>
<td>-0.35</td>
<td>-0.26</td>
<td>-0.31</td>
<td>-0.28</td>
<td>0.34</td>
<td>-0.1</td>
<td>-0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. I enjoy trying new flavors.</td>
<td>-0.47</td>
<td>-0.33</td>
<td>-0.55</td>
<td>-0.53</td>
<td>-0.85</td>
<td>-0.09</td>
<td>-0.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. I love having lots of flavors to choose.</td>
<td>-0.71</td>
<td>-0.58</td>
<td>-0.69</td>
<td>-0.42</td>
<td>-0.67</td>
<td>-0.51</td>
<td>-0.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Ice cream is good for entertaining.</td>
<td>0.78</td>
<td>0.17</td>
<td>0.27</td>
<td>0.39</td>
<td>0.13</td>
<td>0.19</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Everyone loves ice cream.</td>
<td>0</td>
<td>0.06</td>
<td>0.16</td>
<td>0.29</td>
<td>0.02</td>
<td>0.19</td>
<td>-0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Coefficients in blue are statistically significant.

treat” effect of 0.82 monetarily as $0.89 (= 0.82/0.92). That is, respondents who believe ice cream is a wholesome treat are willing to pay $0.89 more per serving for Dreyer’s ice cream than respondents who do not believe it is a wholesome treat. This is a large value.

Coefficients on the left side of Figure 4.6 provide information useful for understanding the conditions in which one’s brand is preferred and avoided. Blue Bell brand, for example, is favorably viewed by respondents who look to the ice cream category for help in having fun. The desire for wholesomeness (item 2) points to Dreyer’s and Breyers brands, and individuals who have minimal interest in the category (item 7) are seen to prefer Blue Bunny and the Store Brand. The coefficient matrix Γ indicates how the brands are perceived by consumers, providing an alternative to driver analysis discussed in Chapter 3.

Coefficients on the right side of Figure 4.6 indicate the conditions in which specific flavors are desired relative to the flavor Vanilla. Vanilla Bean is preferred when respondents are seeking fun (item 3) and trying to make an event special (item 5), but avoided when seeking relief from regular life (item 4). Cookies & Cream and many other flavors are preferred among respondents who report enjoying new flavors, while neopolitan flavor is avoided by respondents indicating that ice cream is used to help them relax and enjoy life. This information is useful in effectively communicating with consumers.
Further examination of Figure 4.6 raises some interesting questions. It can be seen that there are many coefficients in the table that are significant and that can be used to understand the influence of need states on brand preference. But what if there are no significant coefficients for any of the need states, as is the case for Ben & Jerry’s ice cream? It is important to remember that the lack of significance of coefficients for a brand does not imply that the needs are not important, just that we have failed to detect a significant effect (i.e., failing to reject the null hypothesis does not imply that the null is true). The coefficients in Figure 4.6 express the average, or expected increase in the conjoint part-worth for the presence (versus absence) of the need state. They do not reflect the expected increase for specific segments of our sample. It is always important to conduct segment-level analysis using the IDT to determine the presence or absence of an effect for extremes of the heterogeneity distribution. (IF)

Figure 4.7: Florida Vacations Regression Coefficients (Γ): Needs (Q8)

<table>
<thead>
<tr>
<th>Q8</th>
<th>MK</th>
<th>Ep</th>
<th>AK</th>
<th>HS</th>
<th>US</th>
<th>IA</th>
<th>BG</th>
<th>Speed</th>
<th>Mobile</th>
<th>Early</th>
<th>Late</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I go on vacation to relax and enjoy life.</td>
<td>-1.02</td>
<td>0.62</td>
<td>-0.19</td>
<td>-0.03</td>
<td>-0.27</td>
<td>-1.13</td>
<td>0.26</td>
<td>0.1</td>
<td>-0.19</td>
<td>-0.2</td>
<td>-0.09</td>
<td>0</td>
</tr>
<tr>
<td>2. I want a wholesome vacation experience.</td>
<td>0.3</td>
<td>-0.06</td>
<td>0</td>
<td>-0.27</td>
<td>-0.27</td>
<td>-1.24</td>
<td>-0.3</td>
<td>0.05</td>
<td>0.13</td>
<td>-0.14</td>
<td>-0.02</td>
<td>0</td>
</tr>
<tr>
<td>3. I go on vacation to have fun with family/friends.</td>
<td>1.74</td>
<td>0.54</td>
<td>1.17</td>
<td>0.59</td>
<td>0.59</td>
<td>1.57</td>
<td>0.9</td>
<td>-0.3</td>
<td>0.14</td>
<td>0.12</td>
<td>0.15</td>
<td>0</td>
</tr>
<tr>
<td>4. Vacations provide relief from regular life.</td>
<td>-0.09</td>
<td>-0.16</td>
<td>0.6</td>
<td>0.14</td>
<td>-0.35</td>
<td>1.16</td>
<td>0.33</td>
<td>0.24</td>
<td>0.24</td>
<td>-0.02</td>
<td>-0.09</td>
<td>0</td>
</tr>
<tr>
<td>5. Vacations are where lasting memories are made.</td>
<td>1.59</td>
<td>1.39</td>
<td>1.12</td>
<td>0.6</td>
<td>0.63</td>
<td>-0.28</td>
<td>-0.38</td>
<td>-0.11</td>
<td>-0.26</td>
<td>0.12</td>
<td>0.27</td>
<td>0</td>
</tr>
<tr>
<td>6. Vacations remind me of my childhood.</td>
<td>1.59</td>
<td>0.24</td>
<td>0.36</td>
<td>0.43</td>
<td>-0.15</td>
<td>1.07</td>
<td>-1.42</td>
<td>0.4</td>
<td>-0.01</td>
<td>0.22</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>7. I don’t have strong feelings about where to go.</td>
<td>0.76</td>
<td>1.22</td>
<td>1.05</td>
<td>0.66</td>
<td>0.92</td>
<td>-0.23</td>
<td>1.39</td>
<td>-0.25</td>
<td>-0.14</td>
<td>0.08</td>
<td>-0.28</td>
<td>-0.01</td>
</tr>
<tr>
<td>8. I am amazed at how vacation attractions work.</td>
<td>-0.66</td>
<td>-0.53</td>
<td>-0.84</td>
<td>-0.35</td>
<td>-0.05</td>
<td>0</td>
<td>-0.02</td>
<td>-0.2</td>
<td>-0.02</td>
<td>-0.08</td>
<td>-0.2</td>
<td>0.01</td>
</tr>
<tr>
<td>9. I enjoy trying new attractions.</td>
<td>-0.23</td>
<td>0.6</td>
<td>1.21</td>
<td>0.77</td>
<td>0.51</td>
<td>2.06</td>
<td>1.35</td>
<td>0.08</td>
<td>-0.11</td>
<td>-0.17</td>
<td>-0.07</td>
<td>-0.02</td>
</tr>
<tr>
<td>10. I love having a variety of attractions.</td>
<td>2.1</td>
<td>1.17</td>
<td>0.45</td>
<td>1.36</td>
<td>1.98</td>
<td>1.52</td>
<td>0.58</td>
<td>-0.23</td>
<td>0.05</td>
<td>0.09</td>
<td>0.13</td>
<td>0</td>
</tr>
<tr>
<td>11. Vacations help me enjoy my family/friends.</td>
<td>-0.55</td>
<td>-0.3</td>
<td>-0.09</td>
<td>-0.63</td>
<td>-0.48</td>
<td>-0.32</td>
<td>0.23</td>
<td>-0.08</td>
<td>0.01</td>
<td>0.05</td>
<td>0.01</td>
<td>0</td>
</tr>
<tr>
<td>12. Everyone loves going on a Florida vacation.</td>
<td>-0.44</td>
<td>-0.8</td>
<td>-0.8</td>
<td>-0.82</td>
<td>-0.63</td>
<td>-0.24</td>
<td>-1.09</td>
<td>0.1</td>
<td>0.03</td>
<td>0.22</td>
<td>0.01</td>
<td>0</td>
</tr>
</tbody>
</table>

Coefficients in blue are statistically significant.

Figure 4.7 displays the regression coefficients for the Florida Vacation data. The magnitude of many of the coefficients is large, and many of the coefficients have a significant amount of their mass away from zero – i.e., are statistically significant. It is interesting to contrast Magic Kingdom and Busch Gardens. Preference for Magic Kingdom is heightened when respondents want to be with family and friends (item 3), to make lasting memories (item 5), to be reminded of childhood (item 6) and to have access to a variety of attractions (item 10). Preference for Busch Gardens is heightened when
people don’t have a strong feeling of where to go (item 7) and when they want to try something new (item 9), while preference is diminished when respondents indicate a need for being reminded of their childhood (item 6) and believing that everyone loves going on a Florida vacation (item 12). The analysis portrays respondents differently who are drawn to Magic Kingdom than those drawn to Busch Gardens.

The relationship between the needs and product attributes is less pronounced than the relationship between needs and brands, with only three significant relationships detected. This is not surprising because the need variables were selected independent of the product attributes included in the analysis. More significant coefficients would occur if the need-state variables were specifically selected to understand the conditions for which specific attributes were desired. The speed ticket is preferred among respondents who report that vacations remind them of their childhood (item 6), and the mobile app is desired among people who want relief from their regular life (item 4) but not those who believe that vacations are where lasting memories are made (item 5). This information is useful for crafting media messages that “speak” to specific individuals.

By looking at the attributes that rank as statistically significant (in blue), Magic Kingdom (MK) marketers see that their consumers: 1) go to MK to have fun with family and friends, 2) make lasting memories there, 3) believe MK reminds them of their childhood, and 4) go to MK because of the large selection of attractions. These findings suggest that MK marketers should create advertising promoting these aspects of a MK vacation. Advertising that fails to relate to these items are less likely to be heard and remembered by consumers who might favor MK. Likewise, ads that focus on being amazed at how vacation attraction work (number 8 on the list) are not likely to identify with MK or any of the theme parks, and are not likely to speak to any specific market target. (CC)

Demographics

A commonly used set of explanatory variables for conducting analysis in marketing is demographic variables. Their effect on the logit model coefficients $\beta_h$ are reported in Figures 4.8 and 4.9. The heterogeneity model changes only in terms of the covariates used:

$$\beta_h = \Gamma d_h + \zeta_h \quad \zeta_h \sim Normal (0, \Sigma)$$

where $d_h$ is a vector of demographic variables (see Allenby and Ginter (1995)).
The variable “Female” (D1) is a dummy variable equal to one if the respondent is female and equal to zero if male. The variable “Age” (D2) is measured in terms of the age grouping in the questionnaire where a unit increase in the variable is equivalent to an increase of 10 years. The variable “No Children Under 18” (D4) is equal to one if no children are living in the household. The variable “Number of People in the House” (D5) is a simple count of the size of the household, and the variable “Education” (D6) is an ordinal variable where higher values correspond to higher levels of education. For this later variable it might be better to represent it as a series of dummy variables, but we opt for the simple representation to obtain a general feel for the variable. Last, the variable “Income” (D9) is measured in $25,000 increments in the survey.

Figure 4.8: Ice Cream Regression Coefficients (\( \Gamma \)): Demographic Variables

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Dry</th>
<th>BB</th>
<th>BBun</th>
<th>HD</th>
<th>Bry</th>
<th>B&amp;J</th>
<th>SB</th>
<th>VB</th>
<th>Choc</th>
<th>C&amp;C</th>
<th>Neo</th>
<th>VF</th>
<th>Oreo</th>
<th>RR</th>
<th>CC</th>
<th>CCCD</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>0.47</td>
<td>0.62</td>
<td>0.77</td>
<td>0.38</td>
<td>0.57</td>
<td>0.29</td>
<td>0.81</td>
<td>0.21</td>
<td>0.17</td>
<td>0.03</td>
<td>0.18</td>
<td>-0.33</td>
<td>0.19</td>
<td>0.15</td>
<td>-0.33</td>
<td>-0.07</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.16</td>
<td>0.37</td>
<td>0.39</td>
<td>0.13</td>
<td>0.33</td>
<td>0.04</td>
<td>0.56</td>
<td>0.01</td>
<td>0.03</td>
<td>-0.63</td>
<td>-0.26</td>
<td>-0.07</td>
<td>-0.7</td>
<td>-0.06</td>
<td>-0.12</td>
<td>-0.75</td>
<td>-0.07</td>
</tr>
<tr>
<td>No Children Under 18</td>
<td>-0.01</td>
<td>0.14</td>
<td>-0.05</td>
<td>0.07</td>
<td>0.21</td>
<td>0.27</td>
<td>-0.01</td>
<td>-0.79</td>
<td>-0.02</td>
<td>0.18</td>
<td>-0.85</td>
<td>0.35</td>
<td>0.08</td>
<td>-1.01</td>
<td>0.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of People in House</td>
<td>0.39</td>
<td>0.58</td>
<td>0.64</td>
<td>0.35</td>
<td>0.44</td>
<td>0.39</td>
<td>0.75</td>
<td>-0.11</td>
<td>-0.01</td>
<td>-0.11</td>
<td>0.00</td>
<td>-0.35</td>
<td>0.06</td>
<td>-0.11</td>
<td>-0.25</td>
<td>-0.12</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>-0.1</td>
<td>-0.22</td>
<td>0.11</td>
<td>-0.22</td>
<td>0.18</td>
<td>-0.14</td>
<td>0.04</td>
<td>-0.08</td>
<td>0.16</td>
<td>-0.24</td>
<td>-0.2</td>
<td>-0.03</td>
<td>-0.23</td>
<td>0.01</td>
<td>0.2</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Income (25k increments)</td>
<td>0.09</td>
<td>0.02</td>
<td>0.16</td>
<td>0.22</td>
<td>0.21</td>
<td>-0.05</td>
<td>0.13</td>
<td>-0.06</td>
<td>-0.07</td>
<td>0.05</td>
<td>-0.1</td>
<td>0.04</td>
<td>0.3</td>
<td>0.05</td>
<td>-0.11</td>
<td>0.02</td>
<td></td>
</tr>
</tbody>
</table>

Coefficients in blue are statistically significant.

Figure 4.9: Florida Vacations Regression Coefficients (\( \Gamma \)): Demographic Variables

<table>
<thead>
<tr>
<th>Demographic</th>
<th>MK</th>
<th>Ep</th>
<th>AK</th>
<th>HS</th>
<th>US</th>
<th>IA</th>
<th>BG</th>
<th>Speed</th>
<th>Mobile</th>
<th>Early</th>
<th>Late</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>0.7</td>
<td>0.63</td>
<td>0.56</td>
<td>0.11</td>
<td>-0.25</td>
<td>1.03</td>
<td>-1.11</td>
<td>0.03</td>
<td>-0.15</td>
<td>0.09</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Age</td>
<td>-0.11</td>
<td>0.45</td>
<td>0.02</td>
<td>-0.16</td>
<td>-0.22</td>
<td>-0.59</td>
<td>0.03</td>
<td>-0.02</td>
<td>-0.08</td>
<td>-0.07</td>
<td>-0.1</td>
<td>0</td>
</tr>
<tr>
<td>No Children Under 18</td>
<td>-3.72</td>
<td>-2.09</td>
<td>-2.24</td>
<td>-2.23</td>
<td>-2.1</td>
<td>-3.8</td>
<td>-0.68</td>
<td>-0.06</td>
<td>-0.13</td>
<td>-0.16</td>
<td>-0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>Number of People in House</td>
<td>-0.28</td>
<td>-0.35</td>
<td>-0.01</td>
<td>-0.17</td>
<td>0.01</td>
<td>-0.39</td>
<td>0.32</td>
<td>0.05</td>
<td>-0.09</td>
<td>0.07</td>
<td>0.06</td>
<td>0</td>
</tr>
<tr>
<td>Education</td>
<td>0.46</td>
<td>0.49</td>
<td>0.33</td>
<td>0.21</td>
<td>0.3</td>
<td>0.55</td>
<td>0</td>
<td>0.08</td>
<td>0.03</td>
<td>-0.02</td>
<td>0.11</td>
<td>-0.01</td>
</tr>
<tr>
<td>Income (25k increments)</td>
<td>-0.27</td>
<td>-0.03</td>
<td>0.06</td>
<td>-0.27</td>
<td>-0.26</td>
<td>-0.21</td>
<td>-0.34</td>
<td>0.02</td>
<td>0.05</td>
<td>0.01</td>
<td>-0.08</td>
<td>0</td>
</tr>
</tbody>
</table>

Coefficients in blue are statistically significant.

The coefficients reported in both figures are similar in the sense that i) the significant coefficients are more prevalent on the left side of the table (brand intercepts) than on the right (product attributes); ii) there is a greater tendency for the demographic variables to affect all of the brand intercepts in the same way as compared to that which was found in our analysis of needs. For example, female respondents place a higher value on
4.4 Choice Simulator

Exploring the implications of conjoint models becomes tedious when heterogeneity is present. A separate coefficient vector $\beta_h$ is available for each respondent (h) in the survey that can be related to other sets of cross-sectional variables through the $\Gamma$ coefficient matrices. As the product attributes change, the coefficient vector changes and so does the resulting choice probabilities for the brands. Choice simulators are a useful tool for organizing and summarizing these calculations. A choice simulator is a spreadsheet application that uses the coefficient values to explore the implications of changes to attribute levels, needs and demographics on choices. The simulator for this book can be downloaded from:

http://themodellers.com/AnalyticsTool

Bob wanted to please his wife on Valentine’s Day and asked her where she would like to eat. “I love steak. Make it someplace with steak,” she said. He decided to choose a place on his own, and when they arrived there, he asked if she was happy with the restaurant. Her answer, “It’s okay, but I would have picked Chez Moo since it has a much nicer wine list. I like a good wine with my steak.” Poor Bob was frustrated and realized afterwards that he should have asked her about specific restaurant choices. He didn’t even think about how the wine choices would have influenced her preference. Bob would have figured this out if he had asked his wife to choose between different options and related her choices to the attributes of the restaurants. Bob should have read this book. (BB)

4.5 Summary

The success in linking needs to wants in the analysis in this chapter is due to two things:
1. The collection of preference data using a conjoint analysis that forces respondents to indicate their most preferred choice option. The data are analyzed using a multinomial logit model.

2. The use of a random-effects model with covariates describing needs state collected using a Pick any/J data format.

The first item above overcomes the tendency for respondents to not sufficiently discriminate their evaluations when data are collected on a rating scale (see Figure 3.1). Responses to questions of overall satisfaction tend to bunch up at the upper end of the scale and do not provide sufficient information for analyzing determinants of preferences (see Figures 3.3 and 3.4). Conjoint choice experiments require that respondents indicate their most preferred choice option, yielding greater information. In addition, the choice aspect of the conjoint design mimics what is actually done in the marketplace, making it easy and realistic for the respondent to answer.

The second item is integrated with the first through a hierarchical model specification, similar to the hierarchical model introduced in Chapter 1. Our model hierarchy can be expressed as:

1. Choose \( i \) if \( y_{i,h} = \max \{ y_{1,h}, y_{2,h}, \ldots, y_{N,h}, z \} \)
2. \( y_{i,h} = \beta_{0,i,h} + a_{1,i} \beta_{1,h} + a_{2,i} \beta_{2,h} + \cdots + a_{k,i} \beta_{k,h} + \varepsilon_{i,h} \)
3. \( \beta_h = \Gamma' n_h + \zeta_h ; \quad \zeta_h \sim \text{Normal}(0, \Sigma) \)

where the first equation, or level of the hierarchy, describes respondent choice from among a set of offerings with values \( y_{i,h} \). The second level of the hierarchy specifies how value is determined, and it can be combined with the first level to yield a logit model for the observed choices when we assume that the error term \( \varepsilon_{i,h} \) is distributed according to an extreme value distribution:

\[
\Pr (i)_h = \frac{\exp [\beta_{0,i,h} + a_{1,i} \beta_{1,h} + a_{2,i} \beta_{2,h} + \cdots + a_{k,i} \beta_{k,h}]}{1 + \sum_{j=1}^{N} \exp [\beta_{0,j,h} + a_{1,j} \beta_{1,h} + a_{2,j} \beta_{2,h} + \cdots + a_{k,j} \beta_{k,h}]} 
\]

The third level of the hierarchy describes the distribution of heterogeneity that relates needs \( (n_h) \) to wants \( (\beta_h) \). Measuring needs using a Pick any/J format reduces unwanted
scale effects. This system of equations is used to jointly estimate the model parameters using Bayesian methods described in Appendix C.

We find that our three-equation model leads to strong estimated relationships between needs and wants as measured through the regression coefficients \( \Gamma \). The \( \Gamma \) matrix associates respondent needs with model coefficients describing existing and potentially available offerings that are a combination of the measured product attributes. The coefficient matrix can be used to identify respondent segments for the brands and offerings. For Florida Vacations, Figure 4.7 indicates that Magic Kingdom is especially attractive to individuals expressing need states \{3,5,6,10\}, or any combination thereof, while Animal Kingdom is attractive to individuals with need states \{3,5,7,9\}. Strong positioning by a firm is indicated by large and significant coefficients, and unique positioning is indicated by a need being significantly associated with just one brand (e.g., Need 4 for Islands of Adventure). The strength and uniqueness of the significant \( \Gamma \) coefficients is therefore a diagnostic for brand positioning and competitive analysis.

It is important to remember that the \( \Gamma \) coefficients describe current associations made by consumers, but do not provide guidance on how these coefficients can be influenced. Changing \( \Gamma \) is the task of advertising, where need states are depicted and solutions are made available in branded offerings. We will discuss the issue of changing \( \Gamma \) later in Chapter 6. Moreover, each of the \( \Gamma \) coefficients represent an overall effect that connects the motivation of individuals to aspects of the brand, and is an aggregation of many effects such as the efficacy of a product attribute and beliefs about a brand possessing a sufficient quantity and quality of the attribute for it to be effective.

Brand positioning often involves work on beliefs and associations, and our analysis is not meant to replace these important variables. While conjoint analysis is found to work well for physical attributes that can be manipulated and that do not threaten the integrity of the product, it does not work well for psychological attributes that are essential ingredients to the identity of a brand. An example is the country of origin of an offering that considers itself an upscale foreign brand (e.g., Heineken Beer). It is difficult to imagine such offerings absent their country of origin, limiting the use of any analysis requiring its manipulation in an experimental setting. The material discussed in this chapter is therefore offered as a remedy for some, but not all problems of interest to marketing researchers. As with nearly all the material contained in this book, our exposition barely scratches the surface on the variation of models and variables that can be used to study product policy.
Discrete choice experiments allow us to study customer choice by observing how the customer behaves in a simulated choice situation, instead of by asking what they would do. The truth is, people aren’t always entirely honest on surveys. When asked, over 95% of people routinely report that they wash their hands after using the bathroom, but stealthy observers in public restrooms have consistently found that only 80-85% of people really do wash up. Why the disconnect? People don’t want to admit to behavior that might make them look bad, but sometimes inaccurate responses result from mood, honest forgetfulness, or even the desire to stick it to the person who interrupted their dinner with a survey. Observing behavior in a real (or realistically simulated) setting gives us a much better picture of what people will really choose to do in a given situation, and that’s the kind of information a company needs to make effective decisions. (CF-B)
4.6 Homework

1. Use the product optimizer in the IDT to optimize your brand offering relative to the base case. Assume that your optimization goal is to maximize preference share, and consider the following:
   a) Identify which customers want which flavors (ice cream) or attributes (Florida vacation) for your brand. Relate these preferences to the target segments identified in Chapter 2.
   b) Do preferences differ by segment?
   c) Recall the research you conducted on your brand earlier in Chapter 1. How does your optimized offering change if you only consider your key competitor(s)? Which set of competitive brands should you use in your analysis?
   d) Relate needs to wants within your target segments. How do the results differ from the aggregate associations displayed in Figures 4.6 and 4.7?

2. Form a 95% credible interval of respondent-level coefficients for your brand intercept using Figure 4.4 or 4.5.

3. Conduct a tradeoff analysis by converting each of the (average) attribute coefficients to its price equivalent using information from Figure 4.4 or 4.5. For ice cream, the flavor coefficients reflect the average increase in value relative to vanilla. For Florida vacations, the attribute coefficients are measured relative to their absence.

4. What do you learn from the demographic coefficients in Figures 4.8 and 4.9?

5. Continue your journal for the class by recording your thoughts about the following issues:
   a) What did you learn from this chapter?
   b) Where are you confused?
   c) What do you believe to be true from the class material?
   d) What do you question or not totally believe?
   e) What learning can you apply to your brand?
   f) What next piece of analysis would you conduct?

6. What vignette topic is needed to better explain, illustrate or apply the chapter material? Where would it be located? What would you say?
Chapter 5

Pricing Analysis

... and wants to demand.

Price analysis deals with understanding the role of prices, or more broadly, costs, on demand. Our analysis in this chapter extends the product analysis of Chapter 4 so that demand is more than just a discrete 0-1 outcome. Whereas product analysis focuses on how needs translate into wants, pricing examines how wants translate into sales. We introduce the economic concept of utility in this chapter, and employ a model structure that nests the multinomial logit model as a special case when demand is discrete. We begin by examining the limitations of standard regression analysis, where we find that pricing models can quickly require too many parameters for estimation. This motivates the use of economic theory to impose greater structure on our analysis, while ensuring that pricing effects are reasonable.

5.1 Introduction

Every transaction in the marketplace involves receiving something and giving something up. The product we receive gives us utility and improves our state of being. In return, we give up our money, time, attention and other resources so that we have the right to acquire and use the product. In a broad sense, price analysis involves all the things that we give up. In some cases, costs are easy to identify and are primarily related to money. In other cases, costs are difficult to quantify and play a nominal or screening role in a transaction. For example, people either do or do not feel assured that a transaction at an ATM is secure. People either do or do not incur a tolerable emotional cost of having
direct contact with a recipient of a charitable donation. In either event, though, there are limits to the resources we are willing to expend consistent with the idea of goal-directed behavior.

We begin our discussion of pricing by first considering a simple aggregate model of demand. We then turn our attention to economic models of choice that allow analysis to proceed at a more detailed, disaggregate level.

In many ways, a pricing analysis is similar to the analysis we all perform when we decide where to go to lunch. Of course, in order to go to lunch, we need to consider the price of the meal (monetary cost) and the benefit of eliminating hunger (utility). However, when deciding on lunch locations, we also consider many other costs and utilities, such as: how much we like the food; the number of times we've recently eating there; the time and fuel needed to get to the restaurant; the time spent waiting for the food to arrive, and; the health costs for eating a potentially unhealthy lunch. There's a lot more to pricing than just money. (KP)

5.2 Pricing With Regression Models

A simple model of demand is based on the regression model where demand for a product is related to the prices of all the brands in the product category. It is often convenient to convert the quantity demand and prices to natural logarithms:

\[ \ln y_i = \beta_{i,0} + \beta_{i,1} \ln p_1 + \beta_{i,2} \ln p_2 + \cdots + \beta_{i,i} \ln p_i + \beta_{i,j} \ln p_j + \cdots + \varepsilon_i \]

where \( p_j \) is the price of brand \( j \), and \( \beta_{i,j} \) measures the influence of the \( j^{th} \) price on demand for the \( i^{th} \) brand. A regression equation can be specified with the logarithm of demand for each brand as the dependent variable, resulting in a matrix of price coefficients with elements \( \beta_{i,j} \).

A property of regression models is that the regression coefficients measure the expected change in the dependent variable for a unit change in the independent variable, holding fixed the values of the other independent variables. Using partial derivatives to express the change in the variables, we have:
5.2. PRICING WITH REGRESSION MODELS

\[ \beta_{i,j} = \frac{\Delta \ln y_i}{\Delta \ln p_j} = \frac{\partial y_i / y_i}{\partial p_j / p_j} = \frac{\% \text{ change in } y_i}{\% \text{ change in } p_j} \]

Thus, a log-log regression yields regression coefficients that can be interpreted as elasticities. An elasticity is an economic measure of price sensitivity where values greater than 1.0 correspond to elastic demand, and values less than 1.0 are inelastic. As the price of an offering increases, its own demand is expected to decrease while the demand of competitive offerings should increase. The own elasticities, \( \beta_{i,i} \), are expected to be negative, and products that are substitutes, such as brands within a product category, the cross-price elasticities, \( \beta_{i,j} \), are expected to be positive.

---

Does an increase in the price of a brand have to result in a decrease in its demand? Not necessarily. Consider Frances, a young professional woman looking to purchase a quality perfume. She would like to find the best deal for her money, but she is not familiar with the category and lacks product information. Frances may use price as a cue for quality, reasoning that if a perfume can command a high price and still be profitable, it must be good. So she selects the most expensive brand out of her options. Most of the time, however, people judge the quality of a good based on its attributes, and would rather pay less than more for an offering. (EE)

Positive price elasticities can also arise with luxury good due to the effects of status-seeking. When I was in Korea, I was very surprised by Chanel’s marketing strategy. They almost doubled their product price within a three year period, and the demand for Chanel rapidly increased. People who did not know Chanel were willing to buy the product because they believe the high price represented high quality. Nowadays, you have to be waitlisted to buy a classic Chanel bag. (YJP)

The price elasticity can be used to set prices if we make some assumptions about costs and profits. If we assume that profits for a brand are related to demand as follows:
Profits = \pi_i = y_i (p_i - mc_i)

where \( y_i \) is demand for the \( i^{th} \) brand as in the regression equation above, \( p_i \) is the price of the brand and \( mc_i \) denotes marginal costs. Taking the derivative of profit and setting it equal to zero to find the optimal price yields:

\[
\frac{\partial \pi_i}{\partial p_i} = 0
\]

\[
= y_i + \left( \frac{\partial y_i}{\partial p_i} \right) (p_i - mc_i)
\]

\[
= y_i + (\partial y_i / \partial p_i) p_i - (\partial y_i / \partial p_i) mc_i
\]

\[
= y_i p_i / y_i + (\partial y_i / \partial p_i) (p_i / y_i) p_i - (\partial y_i / \partial p_i) (p_i / y_i) mc_i
\]

\[
= p_i + \beta_{i,i} p_i - \beta_{i,i} mc_i
\]

\[
p_i = mc_i \left( \frac{\beta_{i,i}}{1 + \beta_{i,i}} \right)
\]

and

\[
\frac{\partial^2 \pi_i}{\partial p_i^2} = 1 + \beta_{i,i}
\]

The optimal price is expressed as a percentage markup over marginal cost, \( \beta_{i,i} / (1 + \beta_{i,i}) \). This percentage is optimal as long as it reflects a point of maximum profits, or if \( (1 + \beta_{i,i}) < 0 \). Since a product’s own elasticity is negative, this indicates that an optimal price only occurs in the above formulation when demand is elastic, i.e., when \( \beta_{i,i} < -1 \). For example, if the product’s own elasticity is measured to be equal to -2, i.e., \( \beta_{i,i} = -2 \), then the optimal percentage markup is \(-2 / (1 - 2) = 2\), or a 100% markup over marginal costs. If the marginal cost of a product is $1.00, then its optimal price is $2.00.
Bob the baker is the only baker in town, and he’s having a hard time setting his prices. He has been experimenting with his price for a cookie and the change in demand has been all over the place. Bob would really like to know what he should charge to get the optimal value out of his luckily monopolistic situation. Bob recorded the respective volumes and prices for his cookies in the table below.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Volume</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Volume/ Low Price</td>
<td>5000 cookies</td>
<td>$1</td>
</tr>
<tr>
<td>Low Volume/ High Price</td>
<td>100 cookies</td>
<td>$10</td>
</tr>
</tbody>
</table>

People loved Bobs cookies, but they were certainly taken aback by the high price. Clearly there was some limit to how much they would pay. From these extremes, Bob will find the price elasticity \( \epsilon \) for his cookies.

\[
\begin{align*}
\text{Volume} & = (5000 + 100)/2 = 2550 \\
\text{Price} & = (1 + 10)/2 = 5.5 \\
\text{Change/Average} & = (4900/2550) = 1.92 \\
\text{Elasticity} & = \beta_i = (1.92)/(-1.64) = -1.17
\end{align*}
\]

With this measure of price elasticity, Bob will be able to determine how much to mark his cookies up over cost. Optimal percentage markup \( m_{ci} = (\beta_i/(1 + \epsilon)) = ((-1.17)/(1 - 1.17)) = 6.88 \) or 688% over his marginal costs. Knowing that it only costs him $0.47 to make a cookie, Bobs optimum price is easy to find: 6.88(.47) = $3.23 per Cookie. (PP)

An aggregate analysis of demand can provide some guidance to firms seeking to improve profits by combining two key features: a model of how demand changes with changes in prices and a profit function to be optimized. Consider, for example, a retailer who wants to maximize the vector of prices across an entire product line. A system of equations can be built up from a single regression equation by stacking the equations on top of each other:
\[
\begin{bmatrix}
\ln y_1 \\
\ln y_2 \\
\vdots \\
\ln y_n
\end{bmatrix} = \begin{bmatrix}
\beta_{0,1} \\
\beta_{0,2} \\
\vdots \\
\beta_{0,n}
\end{bmatrix} + \begin{bmatrix}
\beta_{1,1} & \beta_{1,2} & \cdots & \beta_{1,n} \\
\beta_{2,1} & \beta_{2,2} & \cdots & \beta_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
\beta_{n,1} & \beta_{n,2} & \cdots & \beta_{n,n}
\end{bmatrix} \begin{bmatrix}
\ln p_1 \\
\ln p_2 \\
\vdots \\
\ln p_n
\end{bmatrix} + \begin{bmatrix}
\varepsilon_1 \\
\varepsilon_2 \\
\vdots \\
\varepsilon_n
\end{bmatrix}
\]

or

\[
\ln y = \beta_0 + B \ln p + \varepsilon
\]

where \( B = \{\beta_{i,j}\} \) is the elasticity matrix. The profit function for the retailer is:

\[
\text{Profit} = \pi = \sum_{i=1}^{n} y_i (p_i - mc_i)
\]

and optimal prices are identified as those that simultaneously satisfy the set of first-order conditions:

\[
\frac{\partial \pi}{\partial p_j} = 0
\]

\[
= y_j + \sum_{i=1}^{n} \left( \frac{\partial y_i}{\partial p_j} \right) (p_i - mc_i)
\]

\[
= y_j + \sum_{i=1}^{n} \beta_{i,j} \left( \frac{p_j}{y_i} \right) (p_i - mc_i)
\]

This expression does not lend itself to a simple solution for an optimal price, or guarantee that there exists a unique set of optimal prices. However, these first-order conditions can be used to intelligently search for a set of prices that attempt to maximize profits. We will return to the issue of optimization later in Chapter 7.

A limitation of aggregate regression models is that there is no guarantee that the regression estimates conform to a coherent demand system for which optimal pricing implications exist. In many applied studies, the algebraic signs of price coefficients can be theoretically incorrect, particularly for cross-price effects where estimates are based on a limited amount of data. Without constraints, there is no guarantee that elasticity estimates are in the elastic range.

Figure 5.1 displays scatterplots for six popular brands of stick margarine sold in a Chicago-area supermarket chain. The data are weekly sales and prices spanning a two-year period. The data are displayed in logarithmic form, and the best fitting regression line is included in the plot, showing a downward sloping line for each of the brands. As price increases, expected demand decreases. We also see that the points in the plots
are stacked up along specific values of prices. This is because supermarket pricing often uses prices that end in the number nine (e.g., $0.79). The dependent variable is the logarithm of the total number of units (four 8-ounce sticks) sold in a specific pricing zone in Chicago, where consumers faced the same set of shelf prices.

Figure 5.1: Demand Curves for Stick Margarine

Table 5.1 reports the matrix of estimated elasticities $B$. The diagonal elements in the table are the brand’s own price elasticities $\beta_{i,i}$, which range from -4.812 for Parkay to -2.276 for the House Brand. Thus, Parkay margarine is the most price elastic while
the House Brand is the least price elastic. The optimal percentage markup for Blue Bonnet is \( \beta_{i,i}/(1 + \beta_{i,i}) = -4.812/(1 - 4.812) = 1.26 \) or 26%, assuming it is a monopoly. Similarly, the optimal percentage markup for Imperial margarine would be 78% under the monopoly assumption.

Table 5.1: Margarine Elasticity Estimates \((B)\)

<table>
<thead>
<tr>
<th>Brand</th>
<th>Imperial</th>
<th>Parkay</th>
<th>Blue Bonnet</th>
<th>Land O' Lakes</th>
<th>Chiffon</th>
<th>House Brand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imperial</td>
<td>-3.598</td>
<td>1.188</td>
<td>0.416</td>
<td>0.818</td>
<td>1.667</td>
<td>0.293</td>
</tr>
<tr>
<td>Parkay</td>
<td>0.491</td>
<td>-4.812</td>
<td>2.006</td>
<td>0.630</td>
<td>-0.879</td>
<td>-0.319</td>
</tr>
<tr>
<td>Blue Bonnet</td>
<td>0.529</td>
<td>0.641</td>
<td>-3.636</td>
<td>0.860</td>
<td>0.583</td>
<td>0.585</td>
</tr>
<tr>
<td>Land O' Lakes</td>
<td>1.097</td>
<td>0.945</td>
<td>1.753</td>
<td>-4.147</td>
<td>0.586</td>
<td>1.022</td>
</tr>
<tr>
<td>Chiffon</td>
<td>0.514</td>
<td>-0.071</td>
<td>-0.260</td>
<td>0.509</td>
<td>-3.828</td>
<td>0.291</td>
</tr>
<tr>
<td>Store Brand</td>
<td>0.372</td>
<td>0.778</td>
<td>1.239</td>
<td>0.734</td>
<td>1.307</td>
<td>-2.276</td>
</tr>
</tbody>
</table>

A problem with using regression analysis to estimate price elasticities is that cross-elastocities, \( \beta_{i,j} \) tend to be inaccurately estimated (see Allenby (1989)). While the diagonal entries in Table 5.1 are expected to be negative, reflecting an inverse relationship between a brand’s price and its own demand, the off-diagonal entries should be positive: An increase in the price of a competitor’s brand should lead to an increase in demand of my brand. Yet four of the thirty (i.e., 13%) off-diagonal elements in Table 5.1 are estimated to be negative, with some of these estimated to be statistically significant.

The entry \(-0.879\) in the second line of the table reads that a 1% increase in the price of Chiffon margarine leads to a 0.879% decrease in the sales of Parkay margarine. This result is not reasonable because a price increase in Chiffon should result in an increase in demand for Parkay, not a decrease in demand. The primary reason for the negative estimates is the large number of parameters that are estimated with a fixed amount of data. Negative cross-elasticity estimates are especially problematic when determining the source of volume due to price decreases – e.g., identifying from whom sales are being taken.

Consider, for example, a regression model applied to the Ice Cream survey data. There are seven brands of ice cream, ten flavors and three different package sizes (see Figure 4.1). This results in the need for \( 7 \times 10 \times 3 = 210 \) different “units” of analysis, assuming that all flavors and sizes are available for each brand. The resulting elasticity matrix is enormous \((210 \times 210 = 44,100 \text{ elements})\), and cannot be estimated without imposing restrictions on the parameters of the model.
In addition to problems encountered with too many parameters, often times the data used to obtain regression estimates is not very informative. The estimates of Table 5.1 are obtained from tens of thousands of customers who were all exposed to the same set of prices. This is not the case with survey data, where the design of the choice task often has different respondents reacting to different prices. In both the Ice Cream and Florida Vacations surveys, there were 16 different versions of the questionnaire that were administered to create sufficient variation in the data for model estimation. When this occurs, it is no longer possible to aggregate the data and use the aggregated choices as the dependent variable. Analysis must take place at a disaggregate level.

The Affordable Care Act was passed, in part, to lower healthcare expenditures. Healthcare costs are the product of both quantity of services demanded and price. If individuals have to pay a greater proportion of their healthcare costs, the idea is that less services would be demanded and thus costs will lower. The problem is that the demand for some healthcare services is much more inelastic than the demand for others.

Consider the effects on an expansive health insurance plan adding a $25 co-payment to visits to the emergency department, mental health therapists, and acupuncturists. If a patient needs emergent medical services, she will present to the emergency department regardless of the $25 additional fee, whereas demand for an acupuncturist - which is much more discretionary in nature - would decline much more.

In designing a health insurance plan, it is important to understand what types of services are more necessary and thus price inelastic, and which are more discretionary and thus price elastic. Marketing analysts, too, must assess which consumer behaviors are likely to adapt to changes in price. (JS)

Table 5.2 displays the demand data from the Florida Vacations survey for the first respondent in the survey. These data are the volumetric responses provided at the bottom of Figure 4.3 and extend the single most preferred data reported in Table 4.1. The data are very sparse, with the most frequently observed entry being the number zero. The data take on only four different values (0,1,2,3), and it is difficult to imagine this short record of preferences providing a reliable estimate of preferences and price sensitivities without imposing additional structure on the analysis. Such structure is the point of incorporating some underlying theory into the analysis. The purpose of any theory is
to parsimoniously explain a phenomena. Regression analysis is often not well suited for the development of theory because it does not impose sufficient restrictions when used to explain the data. For example, allowing for 44,100 coefficients to model pricing effects is too flexible. In the next section we rely on economic theory to give structure to our analysis.

Table 5.2: Florida Vacation Data - First Respondent’s Demands

<table>
<thead>
<tr>
<th>Choice</th>
<th>Magic Kingdom</th>
<th>Epcot Center</th>
<th>Animal Kingdom</th>
<th>Hollywood Studios</th>
<th>Universal Studios</th>
<th>Islands of Adventure</th>
<th>Busch Gardens</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

5.3 Pricing With Economic Models

Analysis of Goal-Directed Behavior

The basic model for goal-directed behavior is a model of constrained maximization. For individuals, what is being maximized is utility. For firms, the objective is to maximize output of some kind, such as profits, sales, impressions or the ability of consumers to recall their brand. We will develop our model for individual utility maximization, knowing that the same mathematical structure applies to firm-level decisions. When working with firm-level analysis, the utility function becomes what is known as a “production function.”

Utility is defined as a scalar measure of preference over a set of alternatives, with the alternatives serving as instruments to improve the state of being of an individual.
The utility function maps inputs, which are the decision options faced by a consumer, to a one-dimensional measure that provides an ordinal rank of the choice alternatives. Utility can be thought of as an attempt to quantify preferences for goods that increase the well-being of a consumer. Inputs can exist in a variety of forms, such as the quality of goods for purchase or decisions about where and how long to attend graduate school. The utility function can be of general form, subject to the constraint that its first derivative, or marginal utility, is positive and decreasing in the arguments of the function (e.g., diminishing marginal returns).

Mathematically, we seek to:

$$\max u(x) \text{ subject to } p'x \leq E$$

where $x$ denotes a vector of quantities, $p$ denotes a vector of prices and $E$ is the budgetary allotment, or expenditure. $E$ is sometimes called the “income constraint” and is often confused with the notion of a household’s annual income. For logical consistency and a well-defined optimization problem, $E$ can only refer to the budget allocated to the collection of alternatives under study (see Kim et al. (2002) and Chandukala et al. (2008)).

Even though it’s true that I love ice cream, I couldn’t possibly eat ice cream for every meal – or even for two meals in a row. This is the general notion behind marginal utility. While marginal utility is literally defined as the change in utility as quantity increase, it is better described as that third pint of ice cream.

The first pint is delicious and refreshing. The second pint is also delicious, but becomes more filling. That third pint, though, starts to become much less delicious and much more filling at the same time. So much so that I don’t want any more. So, while my desire for ice cream will never go away, it does decrease when I have too much. This is exactly how marginal utility works — it never becomes negative, but should always decline as quantity increases.

The next time someone asks you about marginal utility, ask them to join you for a pint of ice cream ... or three. (MM)

The conditioning introduced by studying choices in a particular category cannot be ignored in the formulation of the budget constraint. Thus, a better interpretation for $E$ is the maximum expenditure a consumer is willing to make in the product category, which is different from the annual household income. This implies the observed choices
of a consumer are affected by the budgetary allotment regardless of the magnitude of the prices, affecting both low-priced items such as toothbrushes and high-priced items like automobiles.

Utility functions have the following properties - they are positively valued, (weakly) increasing in $x$ and have non-increasing marginal returns (e.g., diminishing marginal returns). There are instances in the marketing literature where utility functions are said to have a “U” shape, where consumers prefer either small quantities or large quantities, and instances where utility functions are assumed to have an inverted U shape, where moderate amounts of quantity are preferred. These instances refer to what is known as the indirect utility function, not the utility function. The indirect utility function is obtained by solving for the maximum attainable utility given the budgetary allotment. Most choice models in marketing lack budgetary allotment implications, making it easy to mis-characterize utility functions as having either a U or an inverted U shape.

The Smiths are traveling to Florida with a budget of $5,000. This budget includes hotels, food, park admissions, and all other costs involved in the trip. However, their budget for park admissions is only $1,000. As a theme park operator, we want to know how the Smiths maximize their utility based on choices in the theme park category, and not make the mistake that they are willing to tradeoff theme park admissions and food. This is why, as a marketer, it is important to recognize that $E$ is a just portion of a total budget. (KT)

**Constrained Utility Maximization**

The general solution to the problem of associating observed responses to a process of constrained maximization involves a process of first forming a function that combines the utility function and budget constraint with a parameter known as a Lagrangian multiplier:

$$L = u(x) - \lambda(p'x - E)$$

The purpose of the Lagrangian multiplier ($\lambda$) is to ensure that the derivative of $u(x)$ is proportional to the derivative of the constraint ($p'x - E$) at the point of optimality. Differentiating with respect to $x$ we obtain what are known as the Kuhn-Tucker first-order conditions:
5.3. **PRICING WITH ECONOMIC MODELS**

\[
\begin{align*}
  u_j - \lambda p_j &= 0 \quad \text{if } x_j > 0 \\
  u_j - \lambda p_j &< 0 \quad \text{if } x_j = 0
\end{align*}
\]

where \( x \) is the vector of observed optimal demand, and \( u_j \) is the derivative of the utility function with respect to \( x_j \). Rearranging terms, we can express the above as:

\[
\begin{align*}
  \text{if } x_j > 0 \text{ and } x_k > 0 \text{ then } & \frac{u_j}{p_j} = \frac{u_k}{p_k} \text{ for all } j \text{ and } k \\
  \text{if } x_j > 0 \text{ and } x_k = 0 \text{ then } & \frac{u_j}{p_j} > \frac{u_k}{p_k} \text{ for all } j \text{ and } k
\end{align*}
\]

Think of utility as happiness, and goal-directed behavior as the pursuit of that happiness. For some the goal in the game of life is to consume as much happiness as possible, for others it is to produce as much happiness as possible. In either case there are rules that constrain us: we live short lives, with only 24 hours in each day and only a few decades to learn, work and earn money. We are faced with an astounding array of options to help us attain our goal, yet we have limited time, energy, money, and interest. Most of our life is spent NOT choosing and NOT buying as a result of these constraints. What we do choose reflects our best guess on what will help us reach our goal given our limited resources. (CF-B’12)

The Kuhn-Tucker conditions state a general principle of optimality associated with constrained maximization problems when the arguments are non-negative (i.e., either positive or zero). If we observe two choice options to be positively valued, then the bang-for-the-buck is equal when allocation is optimal. If a choice option is not chosen \( (x_k = 0) \), then the bang-for-the-buck for the “not” chosen good is less than that for the goods that are chosen. The bang-for-the-buck is the ratio of marginal utility to price, expressed as \( u_j/p_j \). We will employ the Kuhn-Tucker conditions to relate our observed marketing data to model parameters.
The Kuhn-Tucker conditions apply to any form of constrained allocation. I have a wife, two kids, and work full time so I only have about four hours each week of free time. I enjoy playing basketball the most, and allocate three hours of my time doing that, leaving another hour for other activities like football. I used to only play basketball, but got tired of the repetition and started playing some football for a change of pace. I found that the last hour of basketball was better spent playing another sport, and so I reallocated my time so that the marginal benefit per hour of both sports is equal. (CW)

A Utility Model for Volume Projections

The discrete choice conjoint model discussed in Chapter 4 was used to understand how product attributes affect choice. Choices changed as the levels of the product attributes changed, and this variation was used to estimate the utility part-worths of the products analyzed. The conjoint model allowed us to put the worth of changes in attribute levels on a common scale so that statements could be made about their relative importance to the consumer. Thus, the model was useful for understanding the determinants of utility.

The discrete choice model, however, is limited in its usefulness for understanding how price affects purchase quantities. In this section we introduce a model that makes use of volumetric data obtained by asking respondents to indicate their expected usage under various pricing scenarios. The model described below nests the logit model as a special case, therefore providing a generalization to standard discrete choice analysis. Consider the utility function:

\[ u(x, z) = \sum_{k=1}^{K} \frac{\psi_k}{\gamma} \ln (\gamma x_k + 1) + z \]

where \( x \) is a vector of choice alternatives and \( z \) is an outside good that represents other uses of one’s money. We assume that the price of the outside good is $1.00, so \( z \) represents both the consumption quantity and the expenditure associated with the outside good. The marginal utility associated with this model is of the form:
It might be helpful to step away from the math and view the Lagrange multiplier process more visually.

Farmer Ben would like to build a new pen for his chickens. This pen will be attached to one side of his barn so the chickens can get in and out of the weather. Farmer Ben has 500 feet of fence available. Given that he wants the pen to be as large as possible, how long should the sides of the pen be?

Instead of building or sketching many pens then calculating the area, Farmer Ben can find the solution to this problem using the Lagrange Multiplier technique. Using the diagram at the left as a guide, the area of the pen is \( A = xy \) and the constraint, the amount of fence available, is \( 2x + y \leq 500 \).

These two equations can be combined and graphed. The vertical axis is the pen area given a length of \( x \). You can see that the maximum pen area is achieved when \( x \) is equal to 125 feet. It was fairly easy to graph this problem since there were only two equations and two dimensions, but imagine how complex it would be to graph a problem with more dimensions, such as maximizing profit given quantity, flavor, ingredient cost, and price.

The same solution is found using the Lagrange multiplier technique, but it is found more quickly by using derivatives. The Lagrange multiplier method can be a real time saver. (ET)
CHAPTER 5. PRICING ANALYSIS

\[ u_k(x, z) = \frac{\partial u(x, z)}{\partial x_k} = \frac{\psi_k}{\gamma x_k + 1} \]

and

\[ u_z(x, z) = \frac{\partial u(x, z)}{\partial z} = 1 \]

Marginal utility is therefore diminishing in \( x \) and constant in \( z \). We assume that \( \psi_k \), the marginal utility associated with the first unit of consumption, is specified as a linear compensatory model, similar to that discussed in Chapter 4. Thus, \( \psi_k = \exp[a_k'\beta + \varepsilon] \) where \( a_k \) denote the attributes of the \( k^{th} \) alternative and \( \beta \) are coefficients for their part-worths. Below we assume that the error term, \( \varepsilon \), is distributed Extreme Value (0,\( \sigma \)):

\[ \varepsilon \sim EV(0, \sigma) \]

This assumption allows us to compare results from this new utility model to that of the logit model discussed in Chapter 4. The error term also allows the model to reflect data that is non-deterministic, possibly because of omitted variables.

The coefficient \( \gamma \) is a parameter that governs the rate of satiation in the model, which is estimated from the data. When \( \gamma \) is small, then marginal utility declines slowly, and when it is larger the rate of satiation increases. Figure 5.2 displays the relationship between marginal utility and the purchase quantity for different values of \( \gamma \).

200 million boxes of Girl Scout cookies are purchased each year, with Thin Mints accounting for more than 25% of all cookie box orders. One reason for their popularity is that people just don’t get tired of eating them. How many did you purchase last year before you have had your fill and wanted to add other flavors to your order? Chances are that your \( \gamma \) for Thin Mints is pretty low. (BB)
Mary and Bill are a couple who recently decided to go to a Columbus Clippers baseball game. Both enjoyed their first game—they liked seeing the game in person while enjoying hot dogs, peanuts and the beautiful Columbus weather. After their first game Bill suggested they go to another one the next night and Mary agreed. This time Bill enjoyed the game almost as much as the first night but Mary was less enthusiastic. She really wanted to try the new restaurant near their home. For some reason they decided to go back to the game a third straight night and Bill still loved it. This time Mary brought a book to read and was thinking she could be doing a million things other than attending a stupid baseball game. Mary satiated more quickly (larger $\gamma$), and her marginal utility is now as close to zero as the chances of the Browns winning the Superbowl next year. (CK)

Consumers are assumed to maximize utility subject to a budget constraint (recall that $p_z = $1.00):
$p'x + z \leq E$

where $p$ is the vector of prices, $x$ is the vector of demand quantities, $z$ is the amount of an outside good assumed to be priced as $1.00$ and $E$ is the budgetary expenditure. The utility maximizing solution can be found by using the Kuhn-Tucker conditions:

if $x_k > 0$ then \[ \frac{\psi_k}{\gamma x_k + 1} = \lambda p_k = p_k \]

if $x_k = 0$ then \[ \frac{\psi_k}{\gamma x_k + 1} < \lambda p_k = p_k \]

The identity $z = E - p'x$ is obtained from the budgetary allotment, and $\lambda = 1$ is obtained by applying the Kuhn-Tucker conditions to the quantity $z$ and noting that the price of the outside good is $1.00$.

Rearranging terms in the last set of equations results in an explicit expression for observed demand $x$:

\[
\begin{align*}
    x_k &= \frac{\psi_k - p_k}{\gamma p_k} & \text{for } \psi_k > p_k \\
    x_k &= 0 & \text{for } \psi_k \leq p_k
\end{align*}
\]

Figure 5.3 displays demand curves assuming that $\psi_k = 8.0$ for various rates of satiation. Regardless of the value of the satiation parameter, demand is zero when price exceeds $8.00$, or the bang is less than the buck.

The expression for demand ($x_k$) contains one unknown parameter, $\gamma$, not found in the traditional discrete choice model discussed in Chapter 4. Demand for the $k^{th}$ alternative is increasing in the conjoint estimates $\psi_k = \exp[\alpha_k' \beta]$ and decreasing in the satiation parameter $\gamma$ and price $p_k$. Demand is zero unless the marginal utility $\psi_k$ is greater than the price $p_k$. Thus, demand estimates from this specification are reasonable in that they are never negative nor violate other regularity conditions.
The next time a man asks me why I buy so many pairs of shoes, I’ll tell them that my satiation rate ($\gamma$) for shoes is low. For women, if a shoe doesn’t perfectly match the outfit then satiation doesn’t occur. What men see as brown shoes, women see as brown, leather shoes with two-inch heels with an open toe, and satiation only happens when I have more than one pair of brown, leather, open toed shoes with heels. And, just so you know, you can never have enough black shoes ($\gamma=0$), so men – don’t even try to understand how they’re different! (LK)

**Incorporating Budgetary Effects**

An advantage of the volumetric demand model described above is that it leads to an explicit expression for demand estimates that are well behaved. A disadvantage is that the demand quantities for a good depend only on its own price and not the prices of competing goods. This is due to utility being represented in an additive, separable form over the choice alternatives, coupled with the utility of the outside good not changing.
Additive separability is an assumption that implies that each good contributes to an individual’s well-being separately, with each acting as a substitute for another. That is, Magic Kingdom acts as a tool for addressing the needs of an individual independent from the presence of any of the other theme parks. While a person may prefer to go to Magic Kingdom than any of the other theme parks, the other parks are capable of addressing the same needs, although possibly at a much diminished level. Another way of stating this is that the theme parks do not act as complements to each other in the way that some goods do (e.g., a sprinkler and garden hose, or potato chips and dip) that are jointly consumed. Complementary goods are usually jointly consumed while substitute goods are not, although in either case consumers will be observed to purchase multiple offerings because they grow tired (i.e., satiated) of consuming just one good.

Competitive effects can be incorporated into our model of substitutes by allowing the outside good to satiate:

\[ u(x, z) = \sum_k \frac{\psi_k}{\gamma} \ln (\gamma x_k + 1) + \ln(z) \]

with marginal utility for the outside good equal to:

\[ u_z = \frac{\partial u(x, z)}{\partial z} = \frac{1}{z} \]

Since \( z = E - p'x \), the quantity consumed of the outside good is dependent on the prices of all goods, as is the marginal utility of the outside good. When prices \( (p) \) are lowered, demand \( (x) \) increases at an exceeding rate if demand is elastic, and the amount of the outside good \( (z) \) consumed declines because of the budgetary constraint. Lower prices lead to greater consumption of the inside goods at the expense of the outside good. As \( z \) declines, its marginal utility increases, and a new set of demand quantities is needed to satisfy the Kuhn-Tucker conditions of constrained utility maximization. Thus, competitive forces act through the budgetary allotment \( (E) \).

Mathematically, the effect of a non-linear outside good is to make the Lagrangian multiplier a function of prices:

\[ \lambda = \frac{u_z}{p_z} = \frac{1}{z} = \frac{1}{E - p'x} \]

where the second equality is true because the price of the outside good is normalized to one \( (p_z = 1) \). This leads to Kuhn-Tucker conditions that are a function of all prices:

\[ \text{if } x_k > 0 \text{ then } \frac{\psi_k}{\gamma x_k + 1} = \lambda p_k = \frac{p_k}{E - p'x} \]
if \( x_k = 0 \) then \( \frac{\psi_k}{\gamma x_k + 1} < \lambda p_k = \frac{p_k}{E - p'x} \)

These equations cannot be inverted to obtain an explicit expression for demand as was done earlier when the outside good was assumed to have linear utility. However, numerical (computational) methods can be used to solve for the demand vector for any given set of parameters and prices. These methods are employed in the choice simulator used to make demand predictions.

In life, we are governed by our constraints as much as our freedoms. This is particularly clear when comparing the life of my five-month old daughter to my own. She has every freedom available to her. She can sleep when she wants, eat when she wants, and even go to the bathroom literally wherever she wants. I, on the other hand, am limited in my daily routine by numerous constraints. I have very little freedom. I have a family that relies on my income, so I must spend many hours a week at my job when there are countless other things I would rather be doing. If I want to go play a sport, I have to spend time stretching before AND after to avoid injury as my body has begun its slow deterioration. Last but not least, I’m sure my freedom-loving daughter will one day decide the time is right to be married. At that point, two individuals will become one, while at the same time a little more joy will be taken from me as I then have to deal with an undeserving son-in-law.

As with life, marketing is also governed by a series of constraints. Today, it is impossible to cater to every individual perfectly. Strong marketers focus on segments rather than entire markets. There is no way to create a product that fulfills the need of every individual in the market. Costs would simply put a company out of business if they tried to customize the perfect product for every individual. Understanding budgetary effects and other constraints allows us realistically deal with the facts of life. (NS)
5.4 Analysis

Ice Cream

Figure 5.4 displays the mean ($\mu$) and covariance matrix ($\Sigma$) of the random-effects distribution for the demand model where the outside good is specified in logarithmic form. The logarithmic specification allows for the presence of competitive price effects. The first column of the figure displays estimates of the mean of the distribution, and the estimated covariance matrix takes up the rest of the figure. The estimates can be compared to that of the logit model in Figure 4.4, although some caution should be exercised because the logit model can only handle 0-1 data, whereas the demand model can handle greater quantities. This results in the brand coefficients shifting to be more negative relative to the outside good. In addition, we find that the diagonal elements of the covariance matrix, highlighted in red, are smaller than in the logit model, which indicates that less of the variation of the data is attributed to unobserved heterogeneity. That is, the demand model provides a better fit to the data.

Figure 5.4: Ice Cream Coefficients ($\mu$, $\Sigma$): Random-Effects Model with $E$

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Mean</th>
<th>BB</th>
<th>BBun</th>
<th>HD</th>
<th>Bry</th>
<th>B&amp;J</th>
<th>SB</th>
<th>VB</th>
<th>Choc</th>
<th>C&amp;C</th>
<th>Neo</th>
<th>VFR</th>
<th>Ore</th>
<th>RR</th>
<th>CC</th>
<th>CCCD</th>
<th>ln(g)</th>
<th>ln(s)</th>
<th>ln(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dreyer’s</td>
<td>-4.26</td>
<td>2.22</td>
<td>1.4</td>
<td>1.29</td>
<td>1.23</td>
<td>1.44</td>
<td>1.03</td>
<td>1.34</td>
<td>-0.01</td>
<td>-0.64</td>
<td>0.79</td>
<td>-0.35</td>
<td>-0.47</td>
<td>-0.74</td>
<td>-0.71</td>
<td>-0.86</td>
<td>-0.92</td>
<td>-0.21</td>
<td>0.18</td>
</tr>
<tr>
<td>Blue Bell</td>
<td>-4.48</td>
<td>0.57</td>
<td>2.71</td>
<td>1.66</td>
<td>1.04</td>
<td>1.27</td>
<td>0.92</td>
<td>1.67</td>
<td>-0.06</td>
<td>-0.63</td>
<td>0.76</td>
<td>-0.23</td>
<td>-0.5</td>
<td>-0.68</td>
<td>-1</td>
<td>-0.95</td>
<td>-0.97</td>
<td>-0.36</td>
<td>0.22</td>
</tr>
<tr>
<td>Blue Bunny</td>
<td>-4.31</td>
<td>0.56</td>
<td>0.66</td>
<td>2.36</td>
<td>1.09</td>
<td>1.33</td>
<td>1.09</td>
<td>1.61</td>
<td>-0.03</td>
<td>-0.42</td>
<td>0.93</td>
<td>-0.24</td>
<td>-0.4</td>
<td>-0.74</td>
<td>-0.79</td>
<td>-0.81</td>
<td>-0.99</td>
<td>-0.26</td>
<td>0.22</td>
</tr>
<tr>
<td>Haagen-Dasz</td>
<td>-3.94</td>
<td>0.56</td>
<td>0.43</td>
<td>0.48</td>
<td>2.67</td>
<td>1.48</td>
<td>1.71</td>
<td>1.08</td>
<td>0</td>
<td>-0.49</td>
<td>0.72</td>
<td>-0.2</td>
<td>-0.44</td>
<td>-0.53</td>
<td>-0.71</td>
<td>-0.59</td>
<td>-0.82</td>
<td>0.04</td>
<td>0.29</td>
</tr>
<tr>
<td>Breyers</td>
<td>-3.67</td>
<td>0.65</td>
<td>0.52</td>
<td>0.58</td>
<td>0.67</td>
<td>2.61</td>
<td>1.3</td>
<td>1.55</td>
<td>-0.03</td>
<td>-0.71</td>
<td>0.88</td>
<td>-0.38</td>
<td>-0.59</td>
<td>-0.67</td>
<td>-0.97</td>
<td>-0.8</td>
<td>-1.01</td>
<td>-0.14</td>
<td>0.28</td>
</tr>
<tr>
<td>Ben &amp; Jerry’s</td>
<td>-3.75</td>
<td>0.5</td>
<td>0.4</td>
<td>0.51</td>
<td>0.84</td>
<td>0.62</td>
<td>1.92</td>
<td>1.02</td>
<td>-0.04</td>
<td>-0.51</td>
<td>0.64</td>
<td>-0.17</td>
<td>-0.34</td>
<td>-0.45</td>
<td>-0.64</td>
<td>-0.57</td>
<td>-0.02</td>
<td>0.31</td>
<td>-0.45</td>
</tr>
<tr>
<td>Store Brand</td>
<td>-4.25</td>
<td>0.55</td>
<td>0.62</td>
<td>0.64</td>
<td>0.44</td>
<td>0.63</td>
<td>0.45</td>
<td>2.72</td>
<td>-0.04</td>
<td>-0.58</td>
<td>0.93</td>
<td>-0.27</td>
<td>-0.53</td>
<td>-0.86</td>
<td>-0.96</td>
<td>-0.87</td>
<td>-1.07</td>
<td>-0.25</td>
<td>0.28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>ln(g)</th>
<th>ln(sigma)</th>
<th>ln(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(gamma)</td>
<td>2.16</td>
<td>-0.13</td>
<td>-0.21</td>
</tr>
<tr>
<td>ln(sigma)</td>
<td>-0.42</td>
<td>0.19</td>
<td>0.21</td>
</tr>
<tr>
<td>ln(E)</td>
<td>2.49</td>
<td>-0.34</td>
<td>-0.36</td>
</tr>
</tbody>
</table>

Covariances ($\sigma_{i,j}$) are reported above the diagonal, Correlations ($\rho_{i,j}$) below the diagonal.

The bottom portion of Figure 5.4 displays estimates of the new coefficients: $\ln(\gamma)$,
ln(\(\sigma\)) and ln(\(E\)). The new coefficients are estimated in logarithmic form to ensure that their anti-logs are positive, i.e., \(\lambda = \exp(\lambda^*)\) is always positive for any value of \(\lambda^*\), where \(\lambda^* = \ln(\lambda)\) is the estimated coefficient. On average, the satiation parameter for Ice Cream is \(\exp(-2.16) = 0.115\), indicating a relative slow rate of satiation (see Figure 5.2) for this product category. The scale parameter for the error term, \(\sigma\), is estimated to be \(\exp(-0.42) = 0.65\), which is in the same range as the price coefficient reported for the logit model. Thus, on average, the logit model and continuous demand model indicate roughly the same amount of price sensitivity in consumer choices, although the continuous demand model allows for more than 0-1 demand. The estimate of the average budgetary allotment is \(\exp(2.49) = 12.06\), indicating that respondents have an average upper limit of $12.06 for the ice cream product category. This information is useful for understanding the proportion of wallet a brand currently has, and the upper limit of additional revenue a brand can possibly generate from product introductions and brand reformulation efforts.

Figure 5.5 displays the regression coefficients relating needs (Q9) measured with a Pick any/J format to the parameters of the continuous demand model. The regression coefficients are for the random-effects model that relate model parameters to needs:

\[
\theta_h = \Gamma' n_h + \zeta_h ; \quad \zeta_h \sim Normal (0, \Sigma)
\]

where \(\theta_h = (\beta_h, \ln(\gamma_h), \ln(\sigma_h), \ln(E_h))'\). Coefficients are highlighted in blue if they are “significant” in the sense that the estimate is more than two standard deviations from zero. There is general agreement between the significant coefficients in Table 5.5 and those reported in Table 4.6 for the logit model. For example, respondents reporting agreement with the statement “Ice cream helps make an event special” have more positive brand coefficients than respondents who do not agree with this statement, with the increase in the brand coefficients leading to a greater likelihood of purchasing more ice cream. Interestingly, respondents in agreement with the statement “Everyone loves ice cream” have smaller brand coefficients and larger flavor coefficients, indicating respondents agreeing with this statement are not brand sensitive, but are flavor sensitive.

The last three columns in Table 5.5 are especially interesting. These coefficients associate needs with satiation, the scale of the error term (which is similar to a price coefficient) and the level of expenditure. Respondents indicating agreement with the statement “Ice cream reminds me of my childhood” and “I am amazed at how tastes are created” have significantly greater coefficients for satiation. Recall that the average value of ln(\(\gamma\)) is -2.16, which translates to a satiation value of \(\gamma = 0.115\). The difference between respondents who agree with the first statement and those that don’t is 0.27, and this difference is equal to 0.55 for the second statement. These differences raise and lower
the value of satiation parameter $\gamma$ by 0.04 and 0.09 (see Figure 5.2). Similarly, average expenditure differs by about $2.00 and $3.50 for respondents agreeing and not agreeing with the first and last statements in Q9.

Figure 5.5: Ice Cream Regression Coefficients (T): Needs (Q9) with $E$

<table>
<thead>
<tr>
<th>Q9</th>
<th>Dry</th>
<th>BB</th>
<th>BBun</th>
<th>HD</th>
<th>Bry</th>
<th>B&amp;J</th>
<th>SB</th>
<th>VB</th>
<th>Choc</th>
<th>C&amp;C</th>
<th>Neo</th>
<th>Oreo</th>
<th>RR</th>
<th>CC</th>
<th>CCCD</th>
<th>ln(gam)</th>
<th>ln(sig)</th>
<th>ln(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ice cream helps me relax and enjoy life.</td>
<td>0.21</td>
<td>-0.21</td>
<td>-0.12</td>
<td>0.04</td>
<td>-0.02</td>
<td>0.05</td>
<td>-0.16</td>
<td>-0.11</td>
<td>-0.17</td>
<td>-0.07</td>
<td>-0.49</td>
<td>0.1</td>
<td>0.19</td>
<td>0.04</td>
<td>-0.04</td>
<td>-0.06</td>
<td>0.07</td>
<td>-0.07</td>
</tr>
<tr>
<td>2. Ice cream is a wholesome treat.</td>
<td>0.26</td>
<td>0.06</td>
<td>0.65</td>
<td>0.05</td>
<td>0.43</td>
<td>0.12</td>
<td>0.45</td>
<td>-0.15</td>
<td>-0.06</td>
<td>-0.25</td>
<td>0.03</td>
<td>-0.29</td>
<td>-0.27</td>
<td>-0.22</td>
<td>-0.05</td>
<td>-0.38</td>
<td>0</td>
<td>0.09</td>
</tr>
<tr>
<td>3. Ice cream gives me something fun to do.</td>
<td>-0.04</td>
<td>-0.23</td>
<td>-0.15</td>
<td>-0.17</td>
<td>-0.31</td>
<td>-0.3</td>
<td>-0.29</td>
<td>0.3</td>
<td>0.34</td>
<td>0.57</td>
<td>0.58</td>
<td>0.28</td>
<td>0.64</td>
<td>0.22</td>
<td>0.29</td>
<td>0.69</td>
<td>-0.04</td>
<td>-0.11</td>
</tr>
<tr>
<td>4. Ice cream provides relief from regular life.</td>
<td>-0.29</td>
<td>-0.03</td>
<td>-0.23</td>
<td>-0.23</td>
<td>-0.22</td>
<td>-0.11</td>
<td>-0.39</td>
<td>-0.01</td>
<td>0.25</td>
<td>0.28</td>
<td>0.38</td>
<td>0.09</td>
<td>0.17</td>
<td>0.08</td>
<td>0.33</td>
<td>0.48</td>
<td>-0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>5. Ice cream helps make an event special.</td>
<td>0.48</td>
<td>0.51</td>
<td>0.52</td>
<td>0.47</td>
<td>0.62</td>
<td>0.44</td>
<td>0.41</td>
<td>-0.04</td>
<td>-0.56</td>
<td>-0.17</td>
<td>-0.22</td>
<td>-0.04</td>
<td>-0.3</td>
<td>-0.25</td>
<td>-0.29</td>
<td>-0.55</td>
<td>0.07</td>
<td>0.02</td>
</tr>
<tr>
<td>6. Ice cream reminds me of my childhood.</td>
<td>-0.25</td>
<td>-0.03</td>
<td>-0.23</td>
<td>-0.14</td>
<td>0.01</td>
<td>0.09</td>
<td>-0.08</td>
<td>0.05</td>
<td>0.18</td>
<td>-0.06</td>
<td>-0.33</td>
<td>-0.18</td>
<td>0.26</td>
<td>-0.02</td>
<td>-0.08</td>
<td>-0.15</td>
<td>0.37</td>
<td>0.01</td>
</tr>
<tr>
<td>7. Ice cream is ice cream.</td>
<td>0.52</td>
<td>0.37</td>
<td>0.91</td>
<td>0.21</td>
<td>0.06</td>
<td>0.1</td>
<td>1.09</td>
<td>-0.33</td>
<td>-0.33</td>
<td>-0.27</td>
<td>-0.26</td>
<td>-0.38</td>
<td>-0.02</td>
<td>-0.23</td>
<td>-0.34</td>
<td>-0.09</td>
<td>0.21</td>
<td>-0.07</td>
</tr>
<tr>
<td>8. I am amazed at how tastes are created.</td>
<td>0.17</td>
<td>0.32</td>
<td>-0.09</td>
<td>0.27</td>
<td>0.37</td>
<td>0.27</td>
<td>0.24</td>
<td>-0.13</td>
<td>0.06</td>
<td>-0.31</td>
<td>0.13</td>
<td>0.09</td>
<td>-0.12</td>
<td>-0.27</td>
<td>-0.28</td>
<td>-0.2</td>
<td>0.55</td>
<td>0.16</td>
</tr>
<tr>
<td>9. I enjoy trying new flavors.</td>
<td>-0.35</td>
<td>-0.11</td>
<td>-0.27</td>
<td>-0.16</td>
<td>-0.19</td>
<td>0.06</td>
<td>-0.49</td>
<td>-0.17</td>
<td>-0.2</td>
<td>0.45</td>
<td>0.19</td>
<td>0.48</td>
<td>0.15</td>
<td>0.27</td>
<td>0.1</td>
<td>0.28</td>
<td>-0.02</td>
<td>0.05</td>
</tr>
<tr>
<td>10. I love having lots of flavors to choose.</td>
<td>-0.36</td>
<td>-0.38</td>
<td>-0.38</td>
<td>-0.43</td>
<td>-0.58</td>
<td>-0.4</td>
<td>-0.45</td>
<td>0.32</td>
<td>0.35</td>
<td>0.41</td>
<td>0.28</td>
<td>0.05</td>
<td>0.45</td>
<td>0.64</td>
<td>0.57</td>
<td>0.51</td>
<td>0.15</td>
<td>-0.02</td>
</tr>
<tr>
<td>11. Ice cream is good for entertaining.</td>
<td>0.09</td>
<td>-0.29</td>
<td>-0.18</td>
<td>-0.13</td>
<td>-0.22</td>
<td>-0.21</td>
<td>0.13</td>
<td>-0.07</td>
<td>-0.01</td>
<td>-0.17</td>
<td>-0.06</td>
<td>-0.21</td>
<td>-0.15</td>
<td>-0.2</td>
<td>-0.24</td>
<td>0.08</td>
<td>0.01</td>
<td>0.06</td>
</tr>
<tr>
<td>12. Everyone loves ice cream.</td>
<td>-0.36</td>
<td>-0.36</td>
<td>-0.36</td>
<td>-0.22</td>
<td>-0.42</td>
<td>-0.37</td>
<td>-0.49</td>
<td>0.09</td>
<td>0.44</td>
<td>0.47</td>
<td>0.1</td>
<td>0.13</td>
<td>0.53</td>
<td>0.32</td>
<td>0.25</td>
<td>0.52</td>
<td>0.17</td>
<td>-0.09</td>
</tr>
</tbody>
</table>

Coefficients in blue are statistically significant.

The relationship of the brand, flavor and demand model parameters to demographic variables is reported in Table 5.6. As reported earlier, the demographic variables tend to affect all the brands similarly – e.g., the presence of many people in the house leads to a devaluing of all brands, which is similar to the effect displayed for higher levels of age and education. Flavor is relatively unaffected except for its relationship to age, with newer flavors tending not to be preferred by older respondents. Age and education are also negatively associated with satiation (i.e., older, wealthier respondents do not grow tired of flavors as quickly) and have higher levels of expenditure.

Figure 5.6: Ice Cream Regression Coefficients (T): Demographics with $E$

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Dry</th>
<th>BB</th>
<th>BBun</th>
<th>HD</th>
<th>Bry</th>
<th>B&amp;J</th>
<th>SB</th>
<th>VB</th>
<th>Choc</th>
<th>C&amp;C</th>
<th>Neo</th>
<th>Oreo</th>
<th>RR</th>
<th>CC</th>
<th>CCCD</th>
<th>ln(gam)</th>
<th>ln(sig)</th>
<th>ln(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>0.22</td>
<td>0.18</td>
<td>0.19</td>
<td>0.07</td>
<td>0.09</td>
<td>0.11</td>
<td>0.2</td>
<td>0.15</td>
<td>0.01</td>
<td>0.09</td>
<td>0.04</td>
<td>0.17</td>
<td>-0.27</td>
<td>0.02</td>
<td>-0.05</td>
<td>-0.15</td>
<td>0.12</td>
<td>-0.03</td>
</tr>
<tr>
<td>Age</td>
<td>-0.16</td>
<td>-0.09</td>
<td>-0.1</td>
<td>-0.18</td>
<td>-0.11</td>
<td>-0.23</td>
<td>-0.03</td>
<td>-0.05</td>
<td>0.02</td>
<td>-0.24</td>
<td>-0.07</td>
<td>0.01</td>
<td>-0.24</td>
<td>-0.05</td>
<td>-0.07</td>
<td>-0.23</td>
<td>-0.15</td>
<td>-0.03</td>
</tr>
<tr>
<td>No Children Under 18</td>
<td>0.07</td>
<td>-0.15</td>
<td>-0.23</td>
<td>-0.05</td>
<td>-0.26</td>
<td>0.05</td>
<td>0.07</td>
<td>0.37</td>
<td>-0.16</td>
<td>-0.15</td>
<td>0.02</td>
<td>-0.04</td>
<td>-0.39</td>
<td>0.27</td>
<td>0.1</td>
<td>-0.16</td>
<td>0.28</td>
<td>0.05</td>
</tr>
<tr>
<td>Number of People in House</td>
<td>-0.25</td>
<td>-0.27</td>
<td>-0.29</td>
<td>-0.33</td>
<td>-0.29</td>
<td>-0.29</td>
<td>-0.21</td>
<td>0.11</td>
<td>0.01</td>
<td>0.15</td>
<td>0.1</td>
<td>-0.04</td>
<td>0.1</td>
<td>0.03</td>
<td>0.05</td>
<td>0.14</td>
<td>-0.02</td>
<td>-0.04</td>
</tr>
<tr>
<td>Education</td>
<td>-0.24</td>
<td>-0.32</td>
<td>-0.36</td>
<td>-0.16</td>
<td>-0.27</td>
<td>-0.12</td>
<td>-0.34</td>
<td>0.04</td>
<td>-0.04</td>
<td>0.16</td>
<td>-0.08</td>
<td>-0.03</td>
<td>0.09</td>
<td>-0.08</td>
<td>0.02</td>
<td>0.17</td>
<td>-0.06</td>
<td>-0.02</td>
</tr>
<tr>
<td>Income (25k increments)</td>
<td>0.04</td>
<td>0.04</td>
<td>0.08</td>
<td>0.1</td>
<td>0.13</td>
<td>-0.02</td>
<td>0.03</td>
<td>-0.02</td>
<td>-0.05</td>
<td>-0.03</td>
<td>-0.08</td>
<td>0.02</td>
<td>0.09</td>
<td>0.01</td>
<td>-0.05</td>
<td>0.03</td>
<td>0.02</td>
<td>0</td>
</tr>
</tbody>
</table>

Coefficients in blue are statistically significant.

It is important to remember that the regression coefficients in Figures 5.5 and 5.6 report the expected change in model coefficients for changes in the need state or in the demographic variable. We have previously seen examples indicating that these coefficients can be misleading because they do not reflect the expected changes for a specific segment.
5.4. ANALYSIS

of respondents. The final assessment of the usefulness of the needs and demographics for understanding variation in demand should be made with the use of the IDT that allows investigation at a more disaggregate level.

Florida Vacations

Figure 5.7 displays the mean ($\mu$) and covariance matrix ($\Sigma$) of the random-effects distribution for the demand model calibrated on the Florida Vacation data. The outside good in this model was specified in logarithmic form so that the cross-price effects are non-negligible. The first column in the figure reports the mean of the random effects distribution and the remaining columns are the covariance matrix. The diagonal elements of the covariance matrix, corresponding to the variance of random effects, are shaded in red. Off-diagonal elements are the covariances in parameter estimates across respondents.

Similar to that found with the Ice Cream data, the brand coefficients are smaller than in the logit model (Figure 4.5) due to modeling quantity versus discrete (0-1) choices. The relative rank of the brand coefficients, however, is nearly identical to that found with the logit model, confirming that the difference in the brand coefficients is due to the outside good. The most striking difference, however, is the mean estimates for the product attributes of the speed ticket, mobile app and early/late entry to the park. Whereas these attributes were found to be significant in the logit model, their value in the volumetric demand model are nearly zero. Moreover, the diagonal elements of the covariance matrix associated with these attributes is also zero, indicating that few respondents are interested in these features.

The difference in the estimated importance of the attributes in the continuous demand model (Figure 5.7) and the logit model (Figure 4.5) highlights a distinction between discrete and continuous choice models. The logit coefficients are useful for understanding a discrete 0-1 choice where the respondent indicates their preference for a single occasion of use. The continuous demand estimate reflects intended respondent behavior over the course of, say, an entire week. For respondents planning a family vacation, intending to make visits to multiple theme parks, the benefit of a few extra hours afforded by early/late entry is close to zero, as is the speed ticket and the mobile app. An explanation for this is that the respondent finds theme parks fatiguing, and these attributes make no differences in the parks they plan to attend when considering a week-long vacation. In contrast, a person going to the park for just one day (e.g., a person already living in Florida who will drive in for a day trip) does find the extra features useful and beneficial. We find that the attributes are unimportant within the context of a long-term family vacation, but much more important for respondents deciding to visit the park for a stand-alone vacation day.
CHAPTER 5. PRICING ANALYSIS

Figure 5.7: Florida Vacations Coefficients ($\mu, \Sigma$): Random-Effects Model with $E$

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Mean</th>
<th>MK</th>
<th>Ep</th>
<th>AK</th>
<th>HS</th>
<th>US</th>
<th>IA</th>
<th>BG</th>
<th>Speed</th>
<th>Mobile</th>
<th>Early</th>
<th>Late</th>
<th>ln(g)</th>
<th>ln(s)</th>
<th>ln(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magic Kingdom</td>
<td>-1.37</td>
<td>1.27</td>
<td>0.89</td>
<td>0.91</td>
<td>0.87</td>
<td>0.76</td>
<td>0.82</td>
<td>0.56</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.67</td>
<td>-0.01</td>
<td>-0.12</td>
</tr>
<tr>
<td>Epcot</td>
<td>-1.37</td>
<td>0.71</td>
<td>1.23</td>
<td>0.92</td>
<td>0.88</td>
<td>0.85</td>
<td>0.72</td>
<td>0.78</td>
<td>-0.05</td>
<td>-0.03</td>
<td>-0.01</td>
<td>-0.02</td>
<td>0.63</td>
<td>0</td>
<td>-0.19</td>
</tr>
<tr>
<td>Animal Kingdom</td>
<td>-1.71</td>
<td>0.68</td>
<td>0.7</td>
<td>1.42</td>
<td>0.95</td>
<td>0.8</td>
<td>0.85</td>
<td>0.85</td>
<td>-0.06</td>
<td>-0.04</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.77</td>
<td>-0.05</td>
<td>-0.09</td>
</tr>
<tr>
<td>Hollywood Studios</td>
<td>-1.96</td>
<td>0.64</td>
<td>0.66</td>
<td>0.66</td>
<td>1.55</td>
<td>1.04</td>
<td>0.93</td>
<td>0.83</td>
<td>-0.05</td>
<td>-0.03</td>
<td>-0.02</td>
<td>-0.01</td>
<td>0.68</td>
<td>-0.03</td>
<td>-0.1</td>
</tr>
<tr>
<td>Universal Studios</td>
<td>-1.66</td>
<td>0.57</td>
<td>0.64</td>
<td>0.56</td>
<td>0.72</td>
<td>1.42</td>
<td>1.1</td>
<td>1.05</td>
<td>-0.04</td>
<td>-0.02</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.51</td>
<td>0.02</td>
<td>-0.24</td>
</tr>
<tr>
<td>Islands of Adventure</td>
<td>-1.93</td>
<td>0.56</td>
<td>0.5</td>
<td>0.51</td>
<td>0.59</td>
<td>0.71</td>
<td>1.71</td>
<td>0.89</td>
<td>-0.03</td>
<td>-0.02</td>
<td>0</td>
<td>0.01</td>
<td>0.56</td>
<td>0</td>
<td>-0.19</td>
</tr>
<tr>
<td>Busch Gardens</td>
<td>-2.07</td>
<td>0.35</td>
<td>0.5</td>
<td>0.49</td>
<td>0.62</td>
<td>0.48</td>
<td>2</td>
<td>-0.04</td>
<td>-0.02</td>
<td>0.01</td>
<td>-0.01</td>
<td>0.32</td>
<td>0.05</td>
<td>-0.34</td>
<td></td>
</tr>
</tbody>
</table>

Covariances ($\sigma_{i,j}$) are reported above the diagonal, Correlations ($\rho_{i,j}$) below the diagonal.

I recently took a four-day vacation to Disney World with my three nephews. The first day of the trip, they were very excited and we hurried to wait in line for the rides and tried to see all the shows. By the fourth day, however, we were tired! We got to the park late and only rode three roller coasters before calling it a day. For our vacation, extended hours and fast passes would not have added much value to our tickets because we had so much time to explore and enjoy the park. In contrast, my girlfriend went to Disney World with her husband for just one day and placed a high value on extended hours and fast passes. Because her time was much more limited, she was willing to invest the extra money to ensure she would enjoy more of the parks attractions.

You can see these situations reflected in Figure 5.7. All the coefficients for speed, mobile, early and late extended hours are close to zero when attributes are measured using the volumetric demand model (four days in Disney World) instead of the preference criteria (one day in Disney World). This makes sense because if you have multiple days to enjoy the park you do not need the extended hours and fast pass attributes that allow guests to be at the park all day every day. But if you only have a single day, these attributes make a difference in your ability to enjoy all the parks attractions and are more valued. (KB)

The average estimated satiation parameter ln($\gamma$) is -0.3, translating into an average
level of satiation of $\gamma = 0.74$, which is much larger than the average value of satiation for ice cream. Similarly, the average level of maximum expenditure $E$ is estimated to be $\exp(6.08) = $437 per person. Variation exists around this average level of expenditure, with the variance of random effects of the $\ln(E)$ parameter equal to 0.56. The 95% interval for respondent expenditures ranges from 

$$(\mu - 2\sigma, \mu + 2\sigma) = (6.08 - 2\sqrt{0.56}, 6.08 + 2\sqrt{0.56})$$

or

$$(98, 1952)$$

which is quite large, indicating that expenditures in the product category are very heterogeneous. This range is the expected amount of money spent for theme park tickets, excluding money spent on other vacation items such as transportation, hotels and meals.

Figure 5.8: Florida Vacation Regression Coefficients ($\Gamma$): Needs (Q8) with $E$

<table>
<thead>
<tr>
<th>Q8</th>
<th>MK</th>
<th>Ep</th>
<th>AK</th>
<th>HS</th>
<th>US</th>
<th>IA</th>
<th>BG</th>
<th>Speed</th>
<th>Mobile</th>
<th>Early</th>
<th>Late</th>
<th>In(gam)</th>
<th>In(sig)</th>
<th>In(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I go on vacation to relax and enjoy life.</td>
<td>-0.06</td>
<td>0.14</td>
<td>0.11</td>
<td>0.11</td>
<td>-0.05</td>
<td>-0.16</td>
<td>0.06</td>
<td>0.03</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0.02</td>
<td>-0.04</td>
</tr>
<tr>
<td>2. I want a wholesome vacation experience.</td>
<td>0.23</td>
<td>0.13</td>
<td>0.23</td>
<td>0.2</td>
<td>0.06</td>
<td>-0.05</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>-0.01</td>
<td>-0.03</td>
<td>0.03</td>
<td>0.02</td>
<td>-0.03</td>
</tr>
<tr>
<td>3. I go on vacation to have fun with family/friends.</td>
<td>0.32</td>
<td>0.17</td>
<td>0.35</td>
<td>-0.04</td>
<td>0.1</td>
<td>0.24</td>
<td>0.14</td>
<td>-0.02</td>
<td>0</td>
<td>0.03</td>
<td>0.03</td>
<td>0.19</td>
<td>-0.03</td>
<td>0</td>
</tr>
<tr>
<td>4. Vacations provide relief from regular life.</td>
<td>0.03</td>
<td>0.04</td>
<td>0.16</td>
<td>0.1</td>
<td>0.06</td>
<td>0.32</td>
<td>0.05</td>
<td>0</td>
<td>0.02</td>
<td>-0.03</td>
<td>-0.03</td>
<td>0.08</td>
<td>-0.04</td>
<td>0.02</td>
</tr>
<tr>
<td>5. Vacations are where lasting memories are made.</td>
<td>0.19</td>
<td>0.1</td>
<td>0.16</td>
<td>0.07</td>
<td>-0.07</td>
<td>-0.11</td>
<td>-0.31</td>
<td>-0.03</td>
<td>-0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.08</td>
<td>-0.14</td>
<td>0.25</td>
</tr>
<tr>
<td>6. Vacations remind me of my childhood.</td>
<td>0.27</td>
<td>0.15</td>
<td>0.08</td>
<td>0.2</td>
<td>0.04</td>
<td>0.24</td>
<td>-0.25</td>
<td>0.02</td>
<td>0.03</td>
<td>-0.01</td>
<td>-0.04</td>
<td>0.06</td>
<td>0.07</td>
<td>-0.03</td>
</tr>
<tr>
<td>7. I don’t have strong feelings about where to go.</td>
<td>0.08</td>
<td>0.14</td>
<td>0.11</td>
<td>-0.1</td>
<td>0.15</td>
<td>-0.08</td>
<td>0.2</td>
<td>-0.01</td>
<td>0</td>
<td>0</td>
<td>-0.01</td>
<td>0.02</td>
<td>-0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>8. I am amazed at how vacation attractions work.</td>
<td>-0.01</td>
<td>-0.04</td>
<td>-0.02</td>
<td>0.05</td>
<td>-0.06</td>
<td>0.06</td>
<td>0.13</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.02</td>
<td>0.03</td>
<td>0.08</td>
<td>-0.05</td>
<td>0.13</td>
</tr>
<tr>
<td>9. I enjoy trying new attractions.</td>
<td>-0.12</td>
<td>0.04</td>
<td>0.14</td>
<td>0</td>
<td>0.02</td>
<td>0.25</td>
<td>0.33</td>
<td>0</td>
<td>-0.02</td>
<td>0.03</td>
<td>0.03</td>
<td>0.01</td>
<td>-0.05</td>
<td>-0.02</td>
</tr>
<tr>
<td>10. I love having a variety of attractions.</td>
<td>0.19</td>
<td>0.1</td>
<td>0.05</td>
<td>0.3</td>
<td>0.36</td>
<td>0.16</td>
<td>0.11</td>
<td>0</td>
<td>0.03</td>
<td>0</td>
<td>-0.01</td>
<td>0.32</td>
<td>-0.06</td>
<td>0.16</td>
</tr>
<tr>
<td>11. Vacations help me enjoy my family/friends.</td>
<td>-0.04</td>
<td>0.09</td>
<td>-0.03</td>
<td>-0.02</td>
<td>-0.02</td>
<td>0</td>
<td>-0.09</td>
<td>0.01</td>
<td>-0.04</td>
<td>0</td>
<td>0</td>
<td>0.12</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>12. Everyone loves going on a Florida vacation.</td>
<td>0.06</td>
<td>0.02</td>
<td>-0.05</td>
<td>-0.05</td>
<td>0.08</td>
<td>-0.02</td>
<td>0.06</td>
<td>-0.01</td>
<td>0.02</td>
<td>0</td>
<td>0</td>
<td>0.03</td>
<td>0.02</td>
<td>-0.04</td>
</tr>
</tbody>
</table>

Coefficients in blue are statistically significant.

Figure 5.8 displays the regression coefficients between needs and model parameters. In contrast to the results found with the Ice Cream data, these coefficients are very different from those reported for the logit model in Figure 4.5. The brand coefficients displayed on the left side of the figure are much smaller than those associated with the logit model, implying that the needs are not strongly related to brand preference. The relationship to product attributes is non-existent because the attributes are almost uniformly evaluated to be zero, implying that respondents do not want them. Probably the most interesting results are associated with the demand model estimates: Respondents are found to grow
tired of the theme parks (i.e., satiate more quickly), and are willing to spend more money in the category when “lasting memories” and “having a variety of attractions” are indicated as important motivations.

Figure 5.9: Florida Vacation Regression Coefficients (Γ): Demographics with $E$

<table>
<thead>
<tr>
<th>Demographic</th>
<th>MK</th>
<th>Ep</th>
<th>AK</th>
<th>HS</th>
<th>US</th>
<th>IA</th>
<th>BG</th>
<th>Speed</th>
<th>Mobile</th>
<th>Early</th>
<th>Late</th>
<th>ln(gam)</th>
<th>ln(sig)</th>
<th>ln(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>0.36</td>
<td>0.33</td>
<td>0.41</td>
<td>0.21</td>
<td>0.15</td>
<td>0.31</td>
<td>-0.23</td>
<td>-0.05</td>
<td>-0.02</td>
<td>0.05</td>
<td>0.07</td>
<td>0.43</td>
<td>-0.09</td>
<td>-0.07</td>
</tr>
<tr>
<td>Age</td>
<td>-0.11</td>
<td>-0.07</td>
<td>-0.13</td>
<td>-0.15</td>
<td>-0.15</td>
<td>-0.18</td>
<td>-0.16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-0.03</td>
<td>-0.03</td>
<td>0.32</td>
</tr>
<tr>
<td>No Children Under 18</td>
<td>0.22</td>
<td>0.3</td>
<td>0.21</td>
<td>0.36</td>
<td>0.22</td>
<td>0.02</td>
<td>0.56</td>
<td>-0.02</td>
<td>-0.01</td>
<td>-0.02</td>
<td>0.01</td>
<td>0.05</td>
<td>0.14</td>
<td>-0.31</td>
</tr>
<tr>
<td>Number of People in House</td>
<td>0.02</td>
<td>-0.03</td>
<td>-0.03</td>
<td>0.04</td>
<td>0.03</td>
<td>-0.02</td>
<td>0.06</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.19</td>
</tr>
<tr>
<td>Education</td>
<td>-0.04</td>
<td>-0.05</td>
<td>-0.03</td>
<td>-0.06</td>
<td>-0.05</td>
<td>0.01</td>
<td>-0.08</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>-0.08</td>
<td>0.32</td>
</tr>
<tr>
<td>Income (25k increments)</td>
<td>-0.03</td>
<td>-0.02</td>
<td>0</td>
<td>-0.05</td>
<td>-0.06</td>
<td>-0.07</td>
<td>-0.09</td>
<td>0</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0</td>
<td>0.04</td>
<td>-0.01</td>
<td>0</td>
</tr>
</tbody>
</table>

Coefficients in blue are statistically significant.

Finally, Figure 5.9 reports regression coefficients for the demographic variables. The coefficients are similar for all the theme park coefficients, with female respondents perceiving greater value for most of the brand names, and older respondents perceiving less value. Female respondents are also found to satiate quicker, while we find that expenditure increases with age, education and the number of people living in the house, while decreases with the presence of children in the home.

### 5.5 Summary

The advantage of using an economic model for pricing analysis is that demand and volume projections are guaranteed to be reasonable in the sense of being downward sloping, having higher demand when expenditures are larger, and having lower demand when satiation is higher. These advantages are not present in regression models. However regression models are easier to use and do not require sophisticated methods of estimation. From a strictly pragmatic perspective, a reasonable approach to pricing analysis would be to first try analysis with a regression model, and then move to a more formal analysis based on economic theory if the results are counter-intuitive.

Economic theory, as with any good theory, leads to a parsimonious description of demand. Just three additional parameters are needed (i.e, satiation ($\gamma$), error scale ($\sigma$) and expenditure ($E$)) to extend an analysis of discrete choice to an analysis of volume. In contrast, the regression model requires $K^2$ price-effect parameters to characterize the effects of price on demand where $K$ is the number of distinct offerings. In the Ice Cream study, there are seven brands, ten flavors and three package sizes requiring potentially 44,100 cross-effect parameters to fully characterize price effects for all potential offerings.
5.5. SUMMARY

The presence of goal-directed behavior on the part of individuals and firms requires the use of models that reflect a process where some form of resource conservation, or strategic behavior, is present. The resources that individuals and firms expend in an effort to better themselves can take many forms (Satomura et al. (2011)), including:

\[-2cm\]

\[\begin{align*}
  p'x & \leq E & \text{Budget} \\
  q'x & \leq Q & \text{Quantity} \\
  t'x & \leq T & \text{Time} \\
  th'x & \leq Th & \text{Thoughts} \\
  s'x & \leq S & \text{Sensitivities} \\
  c'x & \leq C & \text{Calories}
\end{align*}\]

Multiple constraints are likely binding all the decisions we make, and the result is that only a few of the many choice options available to an individual are actually considered for purchase, and only a few of the variables that could possibly explain the goals of an individual are relevant at a specific point in time. It is too costly, in a broad sense, to search out, learn, consider, weigh and experience all the products available. People ignore most brands, ignore most messages and ignore most of the efforts made by firms to attract and acquire their business.

Advances in computing power and algorithms, especially Bayesian Markov chain Monte Carlo (MCMC) methods, have enabled the development of models that reflect the presence of goal-directed behavior (see Appendix C). These models treat the “not,” or zero responses differently than the positive responses, offering a framework for relating model parameters to the observed data sensitive to most variables not being immediately relevant to an individual. This difference between “not” and “positive” responses is reflected in the Kuhn-Tucker conditions in an economic model.
Mrs. Gunther has dreamed of taking an African safari ever since she was a little girl. Unfortunately, she never went because her parents were poor and her family decided to go camping in Pennsylvania instead.

As an adult Mrs. Gunther owned a very successful bakery franchise. Now, she had more than enough money to vacation wherever she liked. Unfortunately, time constrained her choices as she only allowed herself to take 4 days of vacation a year over a 30 year period. She seriously considered a trip to Nairobi, but with such limited time she knew that her vacations had to be closer to home and she ended up going to Pennsylvania a lot.

Last year, Mrs. Gunther passed on her baking empire to her son because stress-related problems forced her to retire. Now Mrs. Gunther has all the time and money in the world. She would like to go on an African safari, except that she needs to be close to her doctors who are in, you guessed it, Pennsylvania. Mrs. Gunter, like many of us, is governed more by her contraints than by her desires. (DG)
5.6 Homework

1. What are the consumer costs, in a broad sense, associated with acquiring and using your brand? How are these costs different for a new user versus an experienced user. How might your brand overcome some of these costs?

2. Estimate the own price elasticity for your brand using the demand curve for volume. The estimate can be obtained by calculating the slope of the curve and converting the changes in $y$ (demand) and $x$ (price) to percentage changes. What is the optimal percentage markup over marginal costs assuming your brand is a monopolist? What are the pros and cons of setting prices using the % markup rule.

3. Determine the optimal price for your brand using the IDT. Does your answer change if you focus analysis on your segment or different demographic groups? If so, which price would you recommend?

4. Continue your journal for the class by recording your thoughts about the following issues:
   a) What did you learn from this chapter?
   b) Where are you confused?
   c) What do you believe to be true from the class material?
   d) What do you question or not totally believe?
   e) What learning can you apply to your brand?
   f) What next piece of analysis would you conduct?

5. What vignette topic is needed to better explain, illustrate or apply the chapter material? Where would it be located? What would you say?
Chapter 6

Advertising Analysis

Incorporating brand belief and consideration effects.

The purpose of advertising is to tell people something they didn’t know or to remind them of something they had forgotten. Its goal is to influence the manner in which consumers think about the choice options they have in front of them. In this chapter we consider the effect of advertising – its influence on brand consideration and its influence on the determinants of brand value – in the context of the Ice Cream and Florida Vacations surveys. We find the use of cross-sectional survey data to be insufficient for fully studying the effect of advertising, and discuss additional analysis that is needed for advertising analytics.

6.1 Introduction

Decisions do not just happen. They are an outcome of a process that is usually pre-mediated and sometimes spontaneous. A good way of understanding how advertising influences consumer decision-making is to first understand how consumers make decisions. In our earlier discussion of market segmentation, the distinction between understanding where a person is “coming from” and where they are “going to” was discussed. To understand the role of advertising, it is important to formalize and expand upon this distinction.

Figure 6.1 provides an overview of a behavioral episode that is useful for marketing. It starts, on the left, with personal and environmental systems coming together to produce one or more motivating conditions within an individual. These motivating conditions are
what prompt a person to expend their resources in the form of time, money, attention, effort and work to make an adjustment between themselves and their environment. The desire to make such an adjustment occurs when we bring our car into the repair shop, when we exercise in the gym, when we plan a family vacation and when we attend our children’s soccer games on the weekend. In each case, we are prompted to act for one or more of the motivating reasons discussed in Chapter 2. These motivations are what are generally referred to as “needs” in the marketing literature. We prefer to refer to them in more specific terms that reflect the context of the situation that is prompting action.

This summer my car died without warning, leaving me to find alternate transportation without any time to search for new information. Personal and environmental systems led me to be motivated to move away from an uncomfortable state of being (i.e., sharing a car with my spouse) and to prevent a potential problem (i.e. the mess we would be in if said spouse’s vehicle also died).

Suddenly, I was hyper-aware of the plethora of car ads on television, on the radio, on my computer, on billboards. Car ads were everywhere! These ads helped trigger information that was stored deep in the recesses of my brain about which features (attributes) I like to have in a car: manual transmission, AWD for Ohio winters, highly-rated safety features, heated seats, roomy enough for four, gently used, preferably blue. As my list of features came into focus, I began honing in on specific ads and came to the conclusion that only two makes would fulfill my requirements: BMW and Subaru.

With BMW and Subaru in my consideration set, I scoured the internet for used cars that matched my list of desired features, were within a reasonable distance from Columbus and fit within my budget. Oddly enough, I located one BMW and one Subaru, both blue, at the same lot in Cincinnati. Since both cars fit my list perfectly, my choice (action) came down to which car I preferred driving. In this situation, the winner was Subaru.

Now that I have been driving the ‘Ru’ for several months, I have become a bit of a brand loyalist. I am quite certain that my next car will also be a Subaru! (MP)

The presence of a motivating condition leads to a desire to rectify the imbalance a person experiences with their current and/or anticipated environment. A person may see that it might rain in the afternoon and be prompted to look for a way to stay dry. They
Figure 6.1: A Model of Action and Brand Use
also might want to have company on Saturday night and not be lonely, or might want to get into shape for an upcoming school reunion. They might want relief from a more immediate problem that might be physical in nature (e.g., something to eat or drink), or perhaps they might be seeking relief from boredom. For any of these conditions that describe where a person is coming from, they might look for help in the marketplace in the form of branded instruments that are potentially effective at helping them adjust to their perceived situation. What the person really wants are the attributes and resulting benefits from their use to affect change, and these are offered for sale in the form of branded offerings.

Advertising affects a person’s cognitive associations between motivating conditions and desired attributes, and between attributes and brands. The first association is used to convince an individual that a brand should be considered for potential purchase, and the second affects the perceived value of the brand. Our analysis of the effects of advertising is therefore organized in two parts:

- The influence of advertising on brand awareness and consideration.
- The influence of advertising on brand value.

There is a growing body of literature in marketing documenting the manner in which media advertising affects choices. We know the following about advertising’s effects:

1. The response function to advertising is concave. That is, there are diminishing returns to advertising.

2. Long-term effects are conditional on the existence of short-term effects.

3. People learn quickly and forget slowly.

4. Effect sizes are larger for new products than for established products.

5. A hierarchy of effects is not universally accepted, i.e., thinking, feeling and then doing is one of many possible routes that advertising has on sales.
Imagine one day you are driving to work and you see an advertisement that promotes a toothpaste by emphasizing its attribute of building strong teeth:

1. There are diminishing returns to advertising – it gets your attention immediately, but in a week or so you won’t even realize it is there. The effect of your first initial impressions is significantly higher than later ones.

2. Long-term effects are conditional on the existence of short term effects – unless the effect of the advertisement is strong enough for you to remember today, the advertisement won’t impact your purchases tomorrow.

3. People learn quickly and forget slowly – things that grab people’s attention are either relevant or unexpected. People pick up on the idea of this image fairly quickly.

4. Effect sizes are larger for new products than for established products – its effect on sales depends on the novelty of building strong teeth.

5. A hierarchy of effects isn’t universally accepted – you might forget about this brand but purchase it later because of a sales promotion, and then remember the ad when you look at the packaging. Thinking and feeling might come after doing.
Have you ever noticed how well-known brands want to be “New and Improved”? This is because effect sizes are larger for new products than for established products. When they advertise, brands need to keep telling people what’s new so that new connections and associations are made in consumer’s minds. There aren’t many reasons to pay attention to an ad unless it tells us something new. The amusing part is that, although the product is new and improved, it often has the same great taste, smell and effect as it did before. You don’t want to mess with a winning formula! (LK)

Correlation versus Causation

Our analysis of cross-sectional data obtained from the Ice Cream and Florida Vacations surveys does not permit the study of temporal aspects of advertising. Our survey focuses on established brands for which brand beliefs are likely well established for all respondents, and it is unlikely that a purely survey-based measurement, even if collected over multiple points in time, would provide enough information to fully study the causal effects of advertising. Cross-sectional data is capable of revealing associations (e.g., correlations) among variables, but is not capable of determining which variable causes another to take on large or small values. Determining causal effects requires data collected across time, where it is possible to determine a temporal sequence of events. Without a temporal sequence, it is not possible to know which variable comes first and hence which variable might cause another to occur.

While it is tempting to interpret the relationships reported below as causal, it is important to remember that they are in fact correlational. A positive relationship between
advertising recall and brand usage can be due to advertising reminding respondents to consider and purchase a brand, or it might equivalently be due to purchased brands being remembered more often because they are used. Similarly, a large regression coefficient ($\Gamma$) relating advertising to a brand intercept may be due to advertising influencing brand value, or by brand value enhancing advertising recall. If advertising recall is higher for consumers who are looking for discounts and are willing to be flexible in what they purchase then the advertising coefficient may be estimated to be negative instead of positive. Additionally, time-dependent information is needed to determine which of many possible explanations is correct.

Consider, for example, the analysis of a set of respondents that are heterogeneous in their sensitivity to price. Suppose that price sensitive respondents exhibit low brand loyalty and are very willing to switch brands to get a lower price. Suppose that these respondents are measured to have small brand intercepts. Suppose also that price insensitive respondents exhibit greater brand loyalty in the face of price promotions, and have larger brand intercepts. Finally, let us assume that greater price sensitivity leads to greater attention to advertisements as respondents search for the best deals. This relationship is shown graphically in Figure 6.2.

Figure 6.2: A Hypothetical Price-Advertising Relationship

An analysis of the relationship between advertising recall and brand value, as measured in terms of the size of the brand intercepts, would reveal a negative association for these respondents – i.e., respondents who report seeing many advertisements for a brand are likely price sensitive and have small estimated brand intercepts. This negative association is a true association, and the level of brand recall can be used to predict the size of the brand intercept (e.g., low recall is associated with large brand intercepts and high brand value) for a respondent. However, we know that the association is driven by a respondent’s level of price sensitivity. Increasing the number of advertisements broadcast to a respondent will not, in this example, increase their price sensitivity because the causal relationship works in the opposite direction.
Causation is mistaken for correlation in survey research whenever an across-respondent association is interpreted as a within-respondent association. Causation is an explanation of within-respondent effects and changes. It is never appropriate to measure an across-respondent association in the data and assume that this same association is present at a finer level of analysis, unless it can be convincingly reasoned that alternative explanations are not present.

A golf company wanted to survey its customers to verify their claims that their new golf clubs would make golfers better. But when they analyzed their data, they found that the golfers using the newest clubs actually played worse than the golfers using older clubs. Were the new clubs causing people to be worse golfers? Not if the worst golfers were the ones buying the new clubs to improve their game, and the good golfers were already happy with the clubs they were using. (CK)

We encountered this issue earlier in Chapter 3 when measuring drivers of brand satisfaction, where the presence of scale-usage effects were seen to inflate cross-sectional correlations. Cross-sectional associations do not imply that changing the “drivers” will necessarily lead to an improvement in overall satisfaction for any specific person because other variables (e.g., scale-usage patterns) may be affecting both the driver scores and scores for overall satisfaction. This problem is reduced in the conjoint analysis in Chapter 4 because brand values ($\beta_{0,i}$) were estimated from an experiment in which the product descriptions were manipulated, and driver information was obtained from associations between the estimated conjoint part-worths and respondent answers to need-state questions that are driven by factors outside the experiment. Thus, scaling effects were eliminated, as were many other confounding explanations.

I recently heard this tidbit on the news: marriages in which household chores are shared equally between husband and wife are more likely to end in divorce then those in which the woman does more. The newscasters were discussing the study, saying that perhaps men would hear the results and stop doing housework under the name of “saving the marriage.” However amusing it might be, this reaction is not an appropriate one based on the results of this study. There may be other reasons why couples who share household chores equally have higher divorce rates, such as having joint careers that are stressful and time consuming. So, men – keep vacuuming and remember to support your spouse! (SC)
The detection of within-respondent changes requires data collected at different points in time. While methods exist to accelerate the measured learning resulting from advertising, the requirement of multiple observations over time is something that will not change. In a conjoint experiment, prices change across choice tasks, and one can predict the effect of price changes when data are collected across different sets of prices. Advertising is harder to manipulate in conjoint studies because it affects a respondent’s cognitive associations and is not easily forgotten.

One way of measuring the causal effect of advertising is with a vector autoregressive VAR(X) model. This model is used to measure the relationship among multiple time series, such as advertising expenditures and sales. A simple version of the VAR(X) model for two time series, sales \( (y_{1,t}) \) and advertising \( (y_{2,t}) \) is:

\[
\begin{bmatrix}
  y_{1,t} \\
  y_{2,t}
\end{bmatrix} =
\begin{bmatrix}
  \alpha_{1,1} & \alpha_{1,2} \\
  \alpha_{2,1} & \alpha_{2,2}
\end{bmatrix}
\begin{bmatrix}
  y_{1,t-1} \\
  y_{2,t-1}
\end{bmatrix} +
\begin{bmatrix}
  \beta_{1,1} & \beta_{1,2} \\
  \beta_{2,1} & \beta_{2,2}
\end{bmatrix}
\begin{bmatrix}
  x_{1,t} \\
  x_{2,t}
\end{bmatrix} +
\begin{bmatrix}
  \varepsilon_{1,t} \\
  \varepsilon_{2,t}
\end{bmatrix}
\]

where \( x_{1,t} \) and \( x_{2,t} \) are other variables (e.g., price, seasonality) that might also affect sales and advertising. In long-hand, the model for sales is:

\[
y_{1,t} = \alpha_{1,1}y_{1,t-1} + \alpha_{1,2}y_{2,t-1} + \beta_{1,1}x_{1,t} + \beta_{1,2}x_{2,t} + \varepsilon_{1,t}
\]

and the model for advertising is:

\[
y_{2,t} = \alpha_{2,1}y_{1,t-1} + \alpha_{2,2}y_{2,t-1} + \beta_{2,1}x_{1,t} + \beta_{2,2}x_{2,t} + \varepsilon_{2,t}
\]

A test for the causal effect of advertising on sales is \( H_0 : \alpha_{1,2} \geq 0 \) and \( \alpha_{2,1} = 0 \). That is, last year’s advertising \( (y_{2,t-1}) \) affects current sales \( (y_{1,t}) \), while last year’s sales \( (y_{1,t-1}) \) does not affect current advertising \( (y_{2,t}) \).

If advertising budgets are set as a percentage of last year’s sales (i.e., \( \alpha_{2,1} > 0 \)), then this test will likely be rejected by the data.

Recent research on the short-term effect of advertising on sales indicates that the primary effect is through the construction of a respondent’s consideration set (Gilbride and Allenby (2004), Terui et al. (2011)). Effects of advertising on brand beliefs are slower to occur, particularly for mature brands such as those studied in our analysis. There are many forms of advertising, with many possible mediums for transmission. The analysis presented in the next section is a very short introduction to the measurement of some of
advertising’s effects.

My wife and I are expecting a baby. We find ourselves paying more attention to baby products than we used to, and we have higher recall for brands and products that are related to babies. An analyst measuring this data can erroneously announce that exposing people to these ads is making them have babies, i.e., there is a causal effect. In fact the opposite is true. We first became pregnant, then we started to self-select ourselves into the marketing messages for baby needs, and then began considering which of those satisfy our pre-existing need conditions. Correlation is not causation! (JC)

6.2 Brand Consideration

A simple model of brand consideration relies on a binary logit model:

\[
\Pr(\text{Consideration}) = \Pr(\delta_1 \ln(Ad + 1) + \varepsilon > \text{threshold})
\]

\[
= \frac{\exp[\delta_0 + \delta_1 \ln(Ad + 1)]}{1 + \exp[\delta_0 + \delta_1 \ln(Ad + 1)]}
\]

where “Ad” is the number of advertising exposures seen or heard by an individual within a specified period of time (e.g., the past month). The second equality is true when the error term \(\varepsilon\) is distributed Extreme Value. When the number of advertisements seen is equal to zero, the probability of consideration is set to a baseline level determined by the coefficient \(\delta_0\). As the number of advertising exposures increase, the logarithm in the equation gives rise to diminishing returns.

Diminishing returns occur all the time and are usually measured with a logarithmic function. Loudness is measured in units of decibels (dB), which has a logarithmic relationship to power. The Richter scale for earthquake measurement, the measurement of entropy in thermodynamics and the measurement of pH for acidity are all measured logarithmically, where just-noticeable differences between stimuli is proportional to the magnitude of the stimuli. In marketing, we also find that effects are dependent on baseline magnitudes, and use logarithmic functions to describe their relationships. (MY)
Three things in life are certain: death, taxes, and a touching Budweiser Clydesdale commercial during the Super Bowl. Many spectators view the Super Bowl solely for the commercials, which receive more Monday office talk among larger, non-football fans than the game itself.

Whether or not an individual viewer watching a Super Bowl Clydesdale commercial suddenly purchases more Budweiser beers after the game, as a whole, these horses have become an international symbol of the brand. The initial impact of the commercial debuting during the game increases the brand recognition and consideration among targeted audiences. Super Bowl commercials, requiring millions of dollars in production and displaying, are then played several times over during regular television broadcasts. While fans will continue to react to the commercial, additional exposures to the advertising do not produce that initial impact. Returns on the commercial diminish in accordance with its widespread familiarity and loss of novelty. (JS)

The consideration probability acts as a multiplier to the demand equations, so that the expected demand at a price is the consideration probability multiplied by the value of forecasted demand:

\[ E[\text{Demand}] = \Pr(\text{Consideration}) \times \text{Forecasted Demand} \]

where \( E \) stands for “expectation.” The model coefficients \( (\delta) \) characterize the association between the number of advertisements recalled by a respondent and the likelihood of purchasing the brand in the recent past. The direction of causality is not implied by the model – the model simply associates the consideration probability, as defined, with the number of advertisements recalled. This relationship can be used to predict brand consideration as a function of brand recall, assuming that the underlying relationship between advertising recall and consideration remains the same.

Table 6.1: Ice Cream Consideration Coefficients (S8)

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Dreyer’s</th>
<th>Blue Bunny</th>
<th>Blue Bell</th>
<th>Breyers</th>
<th>Ben and Jerry’s</th>
<th>Haagen-Dazs</th>
<th>Store Brand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ( (\delta_0) )</td>
<td>-1.280</td>
<td>-0.755</td>
<td>-1.554</td>
<td>-0.217</td>
<td>-0.608</td>
<td>-1.533</td>
<td>0.680</td>
</tr>
<tr>
<td>Slope ( (\delta_1) )</td>
<td>0.892</td>
<td>0.514</td>
<td>1.040</td>
<td>0.584</td>
<td>0.519</td>
<td>0.617</td>
<td>0.171</td>
</tr>
</tbody>
</table>

Note: All of the slope coefficients are significant at \( p < 0.01 \) except for the Store Brand.
Table 6.1 reports coefficient estimates for the Ice Cream survey. Brand consideration is determined using question S8 of the survey, which asks respondents to indicate the brands purchased in the past six months. We reason that brands in a person’s consideration set should be purchased eventually because prices in the category change frequently, and if a brand has not been purchased in the recent past then it likely is not seriously considered. The number of advertising exposures is obtained by summing the entries provided in question Q13 of the survey. Thus, we do not distinguish between a TV advertisement and advertisements using other mediums such as radio or print newspaper, although it is possible to do so.

Figure 6.3: Ice Cream Brand Consideration (S8)

All of the estimated relationships between advertising exposure and consideration set construction are significant \((p < 0.01)\) except for the Store Brand. A plot of the relationship between advertising exposure and the probability of brand consideration is provided in Figure 6.3. All of the estimated relationships are concave to the origin, indicating diminishing marginal value for additional advertising. The Store Brand, for example, has an estimated probability of consideration equal to 0.7 for nearly all advertising exposures. The other brands in the analysis show steeper and more significant relationships.
6.2. BRAND CONSIDERATION

Consideration set analysis for the Florida Vacations survey is summarized in Table 6.2 and plotted in Figure 6.4. The slope coefficients are generally smaller than in the Ice Cream category, but continue to indicate a positive association between the number of exposures and the likelihood that a theme park would be visited.

Table 6.2: Florida Vacations Consideration Coefficients (S10)

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Magic Kingdom</th>
<th>Epcot</th>
<th>Animal Kingdom</th>
<th>Hollywood Studios</th>
<th>Universal Studios</th>
<th>Islands of Adventure</th>
<th>Busch Gardens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($\delta_0$)</td>
<td>-0.800</td>
<td>-0.596</td>
<td>-0.807</td>
<td>-0.880</td>
<td>-0.867</td>
<td>-0.626</td>
<td>-1.708</td>
</tr>
<tr>
<td>Slope ($\delta_1$)</td>
<td>0.381</td>
<td>0.483</td>
<td>0.235</td>
<td>0.491</td>
<td>0.404</td>
<td>0.326</td>
<td>0.641</td>
</tr>
</tbody>
</table>

Note: All of the slope coefficients are statistically significant at $p < 0.01$ except Animal Kingdom.

Figure 6.4: Florida Vacation Brand Consideration (S10)

For both the Ice Cream and Florida Vacation product categories, we find evidence of a positive relationship between brand consideration and advertising exposure serving to modify forecasted demand estimates to obtain expected demand, $E[\text{Demand}]$. If we
further assume that the direction of causality moves from advertising recall to brand
consideration, and not from consideration to recall, then the estimated relationship can
be used to help predict sales by manipulating the level of advertising. As discussed earlier,
this is an assumption made outside of, or apart from, the evidence available in the data.
Data from cross-sectional surveys are not informative about causal mechanisms.

Knowledge of a causal mechanism is desirable, but not always available for predicting
the influence of one variable on another. A less restrictive requirement is the assumption
that the explanatory variable is independently, or exogenously, determined. This means
that the value of the independent variable is not determined by the other variables
in the relationship, such as consideration affecting recall, or by any of the coefficients
themselves. An example of an independent variable in marketing is a person’s age, which
is determined apart from the product category under study. Another example is the price
variable used in the conjoint experiment studied in Chapters 4 and 5 where price was
manipulated as a variable in the conjoint study.

When a variable is independently determined, conditional analysis can explore the
effects of alternative levels of a variable on the dependent variable and select the level
that maximizes the level of the outcome variable. Understanding the true causal mecha-
nism is not needed for making predictions so long as the relationship among the variables
(e.g., correlational structure) remains the same. When a variable is not independently
determined, its values are also an outcome of the system of study and is called an “endoge-
nous” variable. In the absence of knowing the true causal mechanism between advertising
recall and brand consideration, an assumption that recall is determined independent, or
“exogenous,” of consideration is sufficient for counter-factual predictions to be valid.

There is an old maxim – you can’t control another person, but you can control
yourself. This concept sheds some light on the difference between endogenous and
exogenous variables. Endogenous variables are determined from within a system
and can not be completely controlled, such as the reactions of individuals and
firms to the things that we do. Exogenous variables are determined apart from the
system of study, acting as inputs that lead us in our behavior and can sometimes
be controlled. Examples of exogenous variables include our age, income, gender
and even our needs when we view them as describing our “upstream” behavior
as in Chapter 2. When we take the initiative to do the right thing, we act for
exogeneous reasons. (LG)
6.3 Brand Value

The effect of advertising on brand value, and its resulting influence on preference, is another way that advertising can affect choice. Brands have value because of beliefs that people hold about them, which are related to the attributes and benefits they are assumed to possess. Brand beliefs can refer to physical aspects of a product that can be manipulated in a conjoint experiment, and can also refer to psychological attributes that are so imbedded in the character of a brand that they change too slowly to be studied in a cross-sectional survey. Automobiles have horsepower and acceleration characteristics that can change across models and years, and psychological characteristics such as “rugged” and “elegant” that typically take years to establish in a person’s mind. Examples of brand beliefs are provided in Table 6.3 for various products.

Table 6.3: Examples of Brand Beliefs

<table>
<thead>
<tr>
<th>Brand Beliefs</th>
<th>Digital Cameras</th>
<th>Blue Jeans</th>
<th>Cell Phones</th>
<th>Soft Drinks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Beliefs</td>
<td>Reliable</td>
<td>Rugged</td>
<td>Easy To Use</td>
<td>Quenches Thirst</td>
</tr>
<tr>
<td>Psychological Beliefs</td>
<td>Techy</td>
<td>Cool</td>
<td>Sophisticated</td>
<td>Fun</td>
</tr>
</tbody>
</table>

In general, there are three ways that advertising can affect brand value (see also Chandukala et al. (2011a)):

1. A direct effect of advertising on brand intercept coefficients and attribute coefficients ($\Gamma$).

2. An effect that is mediated by brand beliefs.

3. An effect that is moderated by brand beliefs.

1. Direct Effect

The simplest model for studying the effect of advertising on choice is to directly relate exposure information to choice model parameters through the distribution of heterogeneity:
The logarithmic specification for advertising is used to reflect diminishing marginal returns. That is, the effect of an additional exposure diminishes as the number of exposures increases. Mathematically, we can see this reflected in the expression for the first derivative of brand belief with respect to advertising:

$$\beta_h = \gamma_0 + \gamma_1 \ln(Ad + 1) + \zeta_h \quad ; \quad \zeta_h \sim \text{Normal}(0, \Sigma)$$

A graph of this relationship is provided in Figure 6.6 for $\gamma_0 = -1$ and $\gamma_1 = 1$. Each additional unit of advertising exposure has diminishing but positive marginal association to the model parameters. It is expected that advertising has a positive effect on brand preference, and therefore the estimated coefficients are positive.

Measuring the effect of advertising is complicated by the fact that people forget slowly, making it difficult to determine which advertisements have convinced us that a brand is desirable. It would be easier if respondents were like Dory, the character in the movie “Finding Nemo,” who suffers from memory loss. Her brand preferences and advertising recall would not be affected by earlier advertising campaigns and past usage experiences. For people with better memories, changes in brand beliefs and preferences are slower to occur and the effects of advertising must be measured over time.
2. Mediation through Brand Beliefs

An alternative model for advertising effects is to assume that advertising affects brand value by changing brand beliefs. Brand beliefs act as a mediating variable for the effects of advertising, and analysis is developed in two stages: i) the effect of advertising on brand beliefs and ii) the effect of brand beliefs on brand value.

From Advertising Exposure to Brand Beliefs

The first portion of a mediation model would use a regression model where brand beliefs are regressed on advertising exposure:

\[ bb = \kappa_0 + \kappa_1 \ln(Ad + 1) + \varepsilon \]

with \( bb \) denoting a specific brand belief, such as those contained in Q11 in the Ice Cream survey or Q10 in the Florida Vacations survey. The variable \( Ad \) is the total number of advertisement exposures as reported in questions Q13 (Ice Cream) or Q14 (Florida
Vacations) of the surveys. The logarithmic specification for advertising is used to reflect diminishing marginal returns.

This analysis explicitly assumes away much information about the content of the advertisement seen by respondents, including types of persuasion used and the specific needs and wants addressed. These aspects are important to consider when developing advertisements, and are not part of the model above. Instead, the model implicitly assumes that these aspects do not change over the brand advertisements recalled by respondents. Thus, the model expressed above has a limited goal of understanding the relationship between execution quantity and brand purchase. Changing brand beliefs is a first step toward this goal.

From Brand Beliefs to Brand Value

The second portion of the model relates brand beliefs to brand value. We can measure brand value as the brand intercepts in a choice model, as it pertains to psychological attributes of a brand that are difficult to change in the short run.

\[
\beta_0 = \delta_0 + \delta_1 bb + \varepsilon
\]

where “bb” are the brand beliefs collected on the 7-point scale in questions Q11 in the Ice Cream survey and Q10 in the Florida Vacations survey. The value \( \beta_0 \) was previously measured through either the choice model of Chapter 4 or the continuous demand model in Chapter 5.

Recall from Chapter 4 that brand value is related to choice through two constructs: i) a model of preference formation and ii) a decision rule, whose combination leads to a model of choice. The model of preference formation was:

\[
y_i = \beta_{0,i} + \beta_1 a_{1,i} + \beta_2 a_{2,i} + \cdots + \beta_k a_{k,i} + \varepsilon_i
\]

where \( a \) denotes the brand attributes and \( \beta_{0,i} \) is the intercept for brand \( i \). The decision rule was:

Choose \( i \) if \( y_i = \max \{y_1, y_2, \cdots, y_n, z\} \)

where \( z \) denotes the outside good, or the decision to not purchase at this time. That is, purchases take place when some \( y_i > z \) and do not occur otherwise.
6.3. BRAND VALUE

In 1984, Apple and IBM fiercely competed for their share of the personal computer market. However, they each took different advertising strategies. IBM used the direct method to increase brand value, but Apple used the mediation approach.

Using the direct approach, IBM’s 1984 TV commercial used four senior citizens to inform the viewer of IBM’s low cost and ease of use.

Apple used the mediation approach during their famous “1984” commercial that aired during the Super Bowl on January 22, 1984. In one interpretation of the commercial, 1984 used the unnamed heroine to represent the coming of the Macintosh as a means of saving humanity from conformity. Another interpretation is that the bluish gray hues and televised Big Brother represented IBM, and Apple attempts to taking down IBM in the personal computer market. Either interpretation specifically influences the viewers brand belief of Apple, which in turns influences brand value.

It’s interesting to note that IBM stopped making personal computers in 2002, and as of April 2013, Apple is third in terms of PC market share and the only PC manufacture of the top three that is growing market share. (KP)
The Ace Tool Company started selling a new line of “Handy Lady” tools – screw drivers, hammers, pliers, etc., – all in fresh feminine colors with soft-grip handles that conformed to a woman’s hands and came in a fashionable tool box. They targeted their advertising to self-help HGTV shows and special displays within hardware stores and hardware departments. Unfortunately, their sales were extremely low despite analysis showing growing unmet demand in the market. A survey showed that those not buying the product felt more comfortable purchasing a brand they had used before, and viewed the Handy Lady as being too feminine. They also did not trust that a pink hammer would work as well as a blue one. For those who did buy the product, they loved the feminine features but seldom ventured into a hardware store where the product was sold.

Ace’s advertising was successful at establishing positive beliefs about Handy Lady tools, but it was not reaching women who occasionally needed tools and valued fashion and being feminine. Knowing about these mediating variables led Ace to change their advertising focus to fashion magazines, women centered television shows and product placement in home decor and craft material sections rather than solely in the hardware department. The results were fabulous – The Ace Tool Company hit a bulls-eye. (BB)

Assuming that the error term distributed Extreme Value resulted in the multinomial choice model:

\[ \Pr (i) = \frac{\exp [\beta_{0,i} + a_{1,i} + a_{2,i} + \cdots + a_{k,i}]}{1 + \sum_{j=1}^{N} \exp [\beta_{0,j} + a_{1,j} + a_{2,j} + \cdots + a_{k,j}]} \]

where we assume the utility of the outside good is zero, i.e., \( z = 0 \), to allow estimation of the remaining coefficients. Combining the elements of our model we can obtain an estimate of the influence of a brand’s advertising on demand using the following relationship:
6.3. BRAND VALUE

\[
\frac{\partial \Pr (i)}{\partial \text{Ad}_i} = \frac{\partial \Pr (i)}{\partial \beta_{0,i}} \times \frac{\partial \beta_{0,i}}{\partial \text{bb}_i} \times \frac{\partial \text{bb}_i}{\partial \text{Ad}_i} \\
= \frac{\partial \Pr (i)}{\partial \beta_{0,i}} \times \delta_{1,i} \times \frac{\kappa_{1,i}}{\text{Ad}_i + 1}
\]

Measuring the mediating effect of advertising and brand preferences helps ensure that advertising effects are what they seem to be. The measurement of a direct effect can yield insignificant associations when the advertisement may in fact be working to change some of the beliefs about the brand.

Suppose that as a child you ate a bowl of Kraft Macaroni and Cheese and got sick, and to this day you avoid it like the plague because it reminds you of an aching stomach. Kraft’s advertising may still be effective at changing your beliefs about the brand’s other attributes, such as its nutritional value, even if you don’t buy the brand and its advertising appears to fail.

Mediation analysis can reveal if advertising is partially effective – effective at changing some brand beliefs but not necessarily the ones with greatest influence on purchases. A challenge in conducting a mediation analysis is including all of the belief variables that lead to preference to identify the ones that stand in the way of purchases. (CW)

3. Moderation by Brand Beliefs

A third way of modeling the effect of advertising is to assume that the strength of the direct relationship between brand value and advertising exposure is enhanced by the strength of brand beliefs. We begin with the standard model relating brand intercepts to advertising:

\[ \beta_h = \gamma_0 + \gamma_1 \ln(Ad + 1) + \zeta_h \quad ; \quad \zeta_h \sim \text{Normal}(0, \Sigma) \]

and assume that the parameter \( \gamma_1 \) is related to brand beliefs:

\[ \gamma_1 = \eta_0 + \eta_1 \text{bb} \]
and, substituting, we arrive at a model with an interaction effect:

\[ \beta_h = \gamma_0 + \eta_0 \ln(Ad + 1) + \eta_1 (\ln(Ad + 1) \times bb) \]

It is helpful to think about these relationships visually. (MB)

---

**Analysis**

Our analysis of the direct effects of advertising on brand value reveals many coefficient estimates indicating that an increase in the reported recall of advertising is often negative (see Figures 6.7 - 6.8). This is especially problematic in the Florida Vacations data. Negative estimates across all coefficients indicates that respondents who report exposure to many advertisements have smaller brand intercepts and, typically, place smaller value on the brand attributes. Similar effects are found using the volumetric demand model of Chapter 5.
We conclude from this analysis that it is especially problematic to assume that analysis conducted across respondents is valid for inferring within-respondent effects. It is highly doubtful that an increase in advertising will lead to a respondent wanting a product less instead of more. Our analysis illustrates the difficulty, if not the impossibility, of understanding temporal effects without having access to temporal data.

Coefficients in blue are statistically significant.
Figure 6.8: Florida Vacations Logit Regression Coefficients (\( \Gamma \)): ln(Ad + 1)

<table>
<thead>
<tr>
<th>Advertising In(Ad+1)</th>
<th>MK</th>
<th>Ep</th>
<th>AK</th>
<th>HS</th>
<th>US</th>
<th>IA</th>
<th>BG</th>
<th>Speed</th>
<th>Mobile</th>
<th>Early</th>
<th>Late</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magic Kingdom</td>
<td>-0.76</td>
<td>-0.6</td>
<td>-0.58</td>
<td>-0.72</td>
<td>-0.84</td>
<td>-0.64</td>
<td>-0.75</td>
<td>0.03</td>
<td>0.08</td>
<td>-0.12</td>
<td>-0.12</td>
<td>0.01</td>
</tr>
<tr>
<td>Epcot Center</td>
<td>-1.07</td>
<td>-0.42</td>
<td>-1</td>
<td>-1.32</td>
<td>-1.32</td>
<td>-1.59</td>
<td>-1.11</td>
<td>0.16</td>
<td>0.07</td>
<td>-0.21</td>
<td>-0.19</td>
<td>0.01</td>
</tr>
<tr>
<td>Animal Kingdom</td>
<td>0.3</td>
<td>0.14</td>
<td>0.35</td>
<td>0.12</td>
<td>-0.07</td>
<td>0.06</td>
<td>-0.06</td>
<td>0.1</td>
<td>-0.12</td>
<td>0.12</td>
<td>0.11</td>
<td>0.01</td>
</tr>
<tr>
<td>Hollywood Studios</td>
<td>0</td>
<td>0.07</td>
<td>-0.12</td>
<td>0.03</td>
<td>-0.54</td>
<td>-0.39</td>
<td>-0.5</td>
<td>-0.01</td>
<td>0.3</td>
<td>0.13</td>
<td>0.11</td>
<td>0</td>
</tr>
<tr>
<td>Universal Studios</td>
<td>0.12</td>
<td>-0.21</td>
<td>0.46</td>
<td>0.7</td>
<td>0.65</td>
<td>0.47</td>
<td>0.67</td>
<td>0.13</td>
<td>0.01</td>
<td>-0.03</td>
<td>-0.22</td>
<td>0</td>
</tr>
<tr>
<td>Islands of Adventure</td>
<td>-0.4</td>
<td>-0.39</td>
<td>-0.59</td>
<td>-0.63</td>
<td>-0.14</td>
<td>0.72</td>
<td>-0.52</td>
<td>0.03</td>
<td>-0.07</td>
<td>0.02</td>
<td>0.08</td>
<td>0</td>
</tr>
<tr>
<td>Busch Gardens</td>
<td>-1.2</td>
<td>-0.82</td>
<td>-0.71</td>
<td>-0.58</td>
<td>-0.09</td>
<td>-1.22</td>
<td>0.9</td>
<td>0.16</td>
<td>0.2</td>
<td>0.01</td>
<td>0.18</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

Coefficients in blue are statistically significant.

Analysis involving models based on mediating and moderating effects proved to have similar results, and is not reported here because the within-respondent interpretation of the coefficients is not plausible. As mentioned earlier, the investigation of advertising effects requires longitudinal data analysis where levels of advertising and demand change through time.

Prior to studying marketing and reading this text, one may find it logical to define a segment of customers in terms of demographic variables (e.g., gender, age, education, and/or income). However, we know now that prospects are best defined in terms of need states. Nonetheless, demographic variables play an important role in the development of an integrated marketing program. While need states are the basis of market segments and serve to guide the content of advertisements, the demographic make-up of the market segments influences which media channels and vehicles are used. In the end, an advertisement must be placed within a channel, and demographics come back to play an important role in advertising.

(JM)

### 6.4 Summary

Advertising analysis requires longitudinal data collected over multiple points in time so that a cause-effect relationship can be measured. The analysis presented in this chapter discusses and illustrates problems in using cross-sectional data alone, where correlations are erroneously assumed to indicate causal relationships.
6.4. SUMMARY

Advertising is known to affect brand consideration (Terui et al. (2011)), and is most effective when informing consumers of something they already don’t know that is also important to them. Respondents are known to more readily engage their attention when advertisements “speak” to their needs. Finally, the fact the consumers forget slowly makes changes in brand associations slow to occur once consumers learn what to expect from a brand.

The effect of our results on the interactive decision tool (IDT) is two-fold. First, we assume that brand consideration characterized by Figures 6.3 and 6.4 are valid and modify the demand estimates multiplicatively to obtain $E[\text{Demand}]$.

Second, since advertising is known to be effective when it speaks to people about their needs, we assume that advertising exposures increase the brand intercept in the demand models of Chapters 4 and 5 in proportion to the size of the $\Gamma$ coefficients in 4.6 and 4.7 for preferences, or the $\Gamma$ coefficients in 5.5 and 5.8 for demand estimates. Thus, in the short run, it is advantageous to advertise to a brand’s strength.

Finally, while customer needs play an important role in the development of advertisements, media purchases are based on demographic variables describing viewership. Advertising therefore cannot be implemented without using demographic variables. Our analysis of market segmentation in Chapter 2 lead us to the creation of segments based on needs, and the interactive decision tool (IDT), provides us with the demographic make-up of the resulting segments. Of course, demographics do not map cleanly onto needs, so there always exists some overlap or spillage in attempting to reach the target audience. For this reason the selection of market segments will always result in spillover effects to non-market targets.

There are three ways that to analyze the relationship between advertising exposure and brand preference. The mere exposure to advertising for Huggies brand diapers may lead a mother to want the product. Alternatively, the ads may convince her that Huggies diapers have certain qualities not present in other diapers by showing the on-screen “leak” test, leading her to change her beliefs and preference for the brand. Or, she may have pre-conceived beliefs about Huggies that are strengthened by the ads. And all three of these approaches require within-person data over time in order to provide valid conclusions. (TT)
6.5 Homework

1. Using an actual advertisement for your brand, provide a causal story of its effect using at least one mediating variable. Diagram your causal story similar to Figure 6.2. How can across-person correlations be confused with within-person causation if you only have access to cross-sectional data?

2. Assume that you have been given a $20 Million advertising budget. Use the interactive decision tool (IDT) to answer the following:
   a) From Figure 6.3 or 6.4, do you expect advertising to have a large or small effect on your brand’s consideration? Is this confirmed in the IDT?
   b) From Tab 1 of the IDT, what media channels do you think would be best to use? Is this confirmed in Tab 6 for Awareness and Consideration?
   c) Compare and comment on differences in advertising effectiveness using preferences (Chapter 4) shares versus volume (Chapter 5). Are advertising expenditures more effective or less effective when they “speak” to the needs of a segment?
   d) Does advertising result in a positive return on investment (ROI)?

3. Advertising is effective when it speaks to people about their needs, and works best for brands that are responsive to those needs.
   a) For which needs is your brand responsive?
   b) What segments are known to have these needs? Consider alternative segmentation solutions to find one or more segments that are of interest to you.
   c) How should you allocate your advertising to maximize profits?
   d) How does your optimal allocation change based on demand estimates using preferences (Chapter 4) versus volume (Chapter 5)?

4. Continue your journal for the class by recording your thoughts about the following issues:
   a) What did you learn from this chapter?
   b) Where are you confused?
   c) What do you believe to be true from the class material?
   d) What do you question or not totally believe?
e) What learning can you apply to your brand?

f) What next piece of analysis would you conduct?

5. What vignette topic is needed to better explain, illustrate or apply the chapter material? Where would it be located? What would you say?
Chapter 7

Optimization

Coordinating product, price and promotion.

Optimizing the marketing mix requires the ability to balance the effect of one decision variable, such as a particular advertising expenditure, against other uses of a firm’s resources. This requires an integrated model of consumer choice in which all decision variables work to produce demand and, ultimately, profits. The advent of Excel-based decision tools make it practical to predict the effects of changing product design, prices and advertising on sales. In this chapter we learn how to balance these effects to achieve optimal profits.

7.1 Introduction

Determining optimal levels of expenditure across marketing’s decision variables implicitly assumes a firm has access to a limited amount of resources that it desires to spend. This leads to trade-offs of expenditures across decision variables. Sometimes the optimal expenditure for a variable is zero, indicating that spending the first budget dollar on an option leads to less profits than spending the last dollar on another. While this can happen when the decision options are expressed at a disaggregate level, it rarely happens when considering budgets for broad classes of decisions. Internet advertising, for example, may best be zero for a certain venture, but it is very doubtful that a firm would find it optimal to not advertise at all across all mediums.

Limited resources expresses itself in the form of constraints on the decision, much like the presence of a budget constraint in Chapter 5. But, instead of a consumer maximizing utility subject to a budget constraint, the problem here is of a firm maximizing profits
subject to resource constraints. It is important to consider whether the effects of product attributes, price and advertising on demand have increasing or decreasing marginal effects on profits. If a decision variable has increasing marginal effects on profits, then the profit maximizing strategy would be to allocate all resources to the variable producing greatest return and zero resources to the remainder. However, this is rarely observed in practice, giving testimony to the fact that effects have diminishing, not increasing marginal return on profits.

Consider, for example, the effect of price on profits for a monopolist. We learned in our discussion at the beginning of Chapter 5 that in a simple log-log regression model, the estimated regression coefficients are elasticities, and that an optimal price for a brand did not exist unless the elasticity was less than negative one. Elastic demand is an example of a relationship with increasing returns – a one percent decrease in price leads to a change in demand that is greater than one percent. But while the effect of price on demand is elastic, this does not imply that price has increasing marginal effects on profits. Figure 7.1 provides an illustration for the regression model:

$$\ln(y_i) = \beta_0 + \beta_1 \ln(price_i) + \varepsilon_i$$

and

$$\text{Profit} = \pi_i = y_i (price_i - mc_i)$$

If we assume that $\beta_0 = 10$ and marginal cost is $2.00, then it can be shown that profit is maximized at a price of $4.00 when $\beta_1 = -2$. As shown in Figure 7.1, the optimal price actually declines as demand becomes more elastic. It is therefore important to remember that the objective when allocating funds to product, pricing and promotion decisions is to maximize profits, not demand.

We begin our discussion of optimization by considering a simple example that is easy to visualize, and then more formally develop concepts around the types of problems found in marketing that can be addressed with Excel-based decision tools. We then go on to develop the profit functions to be maximized for the Ice Cream and Florida Vacations markets.

### 7.2 A Simple Example

Let us assume that a manufacturer must decide on two aspects of their product: i) a level of quality and ii) a level of advertising. Quality can be something like gas mileage for a car, or the number of years a wine might age. Advertising is thought of as broadcasting a message to a population. Sales ($y_i$) is a function of both, people want better gas mileage
(x_1) and will buy the automobile only if they know about it through advertising (x_2). To simplify our analysis, we assume that price is not a variable (i.e., it is fixed), and that firms may pick any positive level of quality and advertising. Let sales be equal to the number of units of quality and the number of units of advertising allocated:

\[ y = x_1 \times x_2 \]

and suppose that a unit of quality costs twice as much as a unit of advertising, with a total budget of 100 units:

\[ 2x_1 + x_2 \leq 100 \]

The optimal allocation of quality and advertising units to maximize sales (y) subject to the constraint can be found by forming an expression that is a combination of the original sales function and the constraint:

\[ L = x_1 \times x_2 - \lambda(2x_1 + x_2 - 100) \]

where the parameter \( \lambda \) is known as the “Lagrangian multiplier.” The purpose of \( \lambda \) in the new function is to ensure that the new maximizing solution conforms to a solution in
the original sales function, having a first derivative that is proportional to the constraint. Figure 7.2 shows the line $2x_1 + x_2 = 100$. The set of feasible solutions to the maximization problem must be below the line. This set of feasible solutions is known as the “budget set.”

Figure 7.2: Budget Set

Gas prices in 2012 are predicted to reach $5/gallon by summer, a 57\% increase in the last six months. Some consumers are tightening their belts, looking to spend less in many product categories. As gas prices go up, budget sets shrink and consumers are priced out of many markets. (MN)

Figure 7.3 adds dashed contour lines to the budget set. The contour lines are the set of $\{x_1, x_2\}$ that yield a specific sales level. The bottom contour traces the values of $x_1$ and $x_2$ that yield a sales level of 500, such as $\{50, 10\}$, $\{49, 10.2\}$, $\{48, 10.4\}$, etc. The input values $\{x_1, x_2\}$ for quality and advertising that maximize sales within the budget.
set are the values \{25, 50\}. The sales contour is tangent to the budget line at this point, and the maximum value of attainable sales is 2500.

Figure 7.3: Budget Set and Sales Contours

We can solve for the optimal values of \(x_1\) and \(x_2\) by differentiating the function \(L\) above in the usual way, and setting the derivatives to zero:

\[
\frac{\partial L}{\partial x_1} = 0 = x_2 + 2\lambda \\
\frac{\partial L}{\partial x_2} = 0 = x_1 + \lambda \\
\frac{\partial L}{\partial \lambda} = 0 = 2x_1 + x_2 - 100
\]

where we find from the first equation that \(x_2 = -2\lambda\), from the second equation that \(x_1 = -\lambda\) and from the third that \(\lambda = -25\). Substituting the solution of the third equation into the other two gives us \(\{x_1, x_2\} = \{25, 50\}\) as the optimal value. This
example illustrates how a constrained maximization problem can be made to resemble a standard maximization problem by using Lagrangian multipliers.

Now that we have covered all the sections in the book, it should be apparent that all of the factors need to come together and work in conjunction with one another. Changing one thing affects another and what seems like a logical change may have the opposite results than what was expected. Think of a balloon and when you squeeze it. It will distort and transform into various shapes depending on where you put the pressure and it may not distribute the air within to the location that you might think. But as you plan and think about where to apply the pressure, you will get the desired results, which in the balloon example is the shape of a particular animal, and in marketing, optimization! (BK)

7.3 Optimization for Marketing

Determining the optimal value of input variables requires an objective function that translates the inputs, or decision variables, into an output measure. We will make use of the multinomial logit model of Chapter 4, and the volumetric demand model of Chapter 5, where attributes and prices serve as inputs to product demand estimates. We will modify the demand estimates from these models in two ways according to the discussion of Chapter 6:

1. An advertising effect that creates a consideration set, where the probability of consideration is used to modify the demand estimate such that expected demand is the product of consideration probability and the original demand estimate.

2. An advertising effect that modifies the regression coefficients in $\Gamma$ that relates needs to wants. Details of this modification are discussed below.

A generic sales function can be written as:

$$y = f(\text{attributes, prices, advertising})$$

where $y$ is the vector of demand estimates of all brands in the market. There are costs associated with offering specific attributes and advertising, and the profit for a brand is measured in terms of its own sales multiplied by its own contribution margin:

$$\text{Profit} = \pi_i = y_i(\text{price}_i - \text{mc}_i)$$
We assume that brand profit is maximized subject to expenditures on attributes and advertising. Generically, the profit maximizing solution is found by forming the function combining profits and costs:

\[ L = \pi_i - \lambda (cost_{attributes} + cost_{advertising} - E) \]

Taking derivatives leads to the expressions:

\[ \frac{\partial L}{\partial \text{price}_i} = 0 = \frac{\partial \pi_i}{\partial \text{price}_i} \]

\[ \frac{\partial L}{\partial \text{cost}_{attribute}} = 0 = \frac{\partial \pi_i}{\partial \text{cost}_{attribute}} - \lambda \]

\[ \frac{\partial L}{\partial \text{cost}_{advertising}} = 0 = \frac{\partial \pi_i}{\partial \text{cost}_{advertising}} - \lambda \]

The second two expressions point to a general feature of optimization: The “bang-for-the-buck” across decision options are equal when allocations are optimal:

\[ \frac{\partial \pi_i}{\partial \text{cost}_{attribute}} = \frac{\partial \pi_i}{\partial \text{cost}_{advertising}} = \lambda \]

where “bang” is the marginal profit and “buck” is the marginal cost. This expression holds exactly when costs can be allocated continuously, and hold approximately when they are discretely allocated as in the Ice Cream and Florida Vacations markets.

This suggests the following procedure for finding the optimal price and optimal expenditures on product attributes and advertising:

1. The interactive decision tool (IDT) will optimize a product or product line in terms of the product attributes and price. This optimization is done given a set of competitive offerings. An important part of product positioning involves anticipating competitor plans and actions.

   a) Determine a relevant set of competitors for your brand using Tab 1 of the IDT.

   b) Select a segmentation strategy (2-Stage or K-means) and market segments using Tab 2 of the IDT. Optimize your product offerings for each selected segment using Tab 4, holding price fixed at a moderate level and including only relevant competition in your calculations. Reconcile differences across the segment-specific solutions, and determine a product offering that will be attractive to all selected market segments.
c) Optimize profits for your brand by changing its price with Tab 5, and holding fixed the product offering.

2. For an overall advertising budget of $20 Million, develop an advertising campaign with Tab 6 of the IDT that considers:

   a) Different media.
   b) Different market segments.
   c) Different spends per medium and segment.

Select your advertising message based on the needs of each segment. Greater gains are realized when an advertising message is congruent with known concerns and interests of the individuals you wish to reach.

3. Explore alternative advertising allocations. The optimal allocation for your brand is the one that maximizes profits, i.e., no other allocation leads to higher profits. The optimal level of advertising can be identified using the “bang-for-the-buck” principle. It is profitable to continue to reallocate advertising dollars so long as profits increase.

Optimization is all about finding balance in the decision variables so that our objective is maximized. Without balance we find that we can reallocate our expenditures to improve our state of being, while with balance we find that there is no way to improve upon our solution. The problem is that our balance is upset by changes in our costs, consumers and the behavior of our competitors. Keeping balance is a continuous process. (QL)
Constrained optimization can also be done using Solver in Excel. The constrained problem discussed earlier can be set up in a worksheet as follows:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>X1</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>X2</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Constraint</td>
<td></td>
<td>2(X1)+X2</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>Sales</td>
<td>X1*X2</td>
<td>1250</td>
<td></td>
</tr>
</tbody>
</table>

Click on Solver in the data panel and set the constraint in the Solver Parameters as below. Make sure to select Max as required in this case.
Click the OK button and the Solver will show the answer in the Answer Report as below:

<table>
<thead>
<tr>
<th>Objective Cell (Max)</th>
<th>Final Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C5$ X1*X2</td>
<td>1250</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable Cells</th>
<th>Final Value</th>
<th>Integer</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B2$ X1</td>
<td>25</td>
<td>Contin</td>
</tr>
<tr>
<td>$B3$ X2</td>
<td>50</td>
<td>Contin</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Constraints</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$C4$ 2(X1)+X2 = $C4=$B2*2+$B3</td>
<td>Binding</td>
<td>0</td>
</tr>
</tbody>
</table>

From the above result we can see, the Solver has calculated the optimized data in the Final Value. (BL)
7.4 Concluding Thoughts

This book proposes the use of decision-based analytics to address marketing’s toughest problems. Decision-based analysis encompasses the data used to understand consumer behavior and the models used to inform decisions. It acknowledges and reflects behavior that is both goal-directed and discrete. Goal-directed means that data will invariably take on the value of zero more than other values because most things are of no interest to most people. We do not participate in most of the product categories with brands for sale, expose ourselves to most of the messages broadcast to inform and persuade us or shop at most of the retail outlets where products are made available. Analysis in marketing must therefore deal with the discreteness of human behavior that is driven by people seeking to improve their life with access to limited resources.

We began our discussion of decision-based analytics in Chapters 1 and 2 by introducing a simple model for data collected using a Pick any/J format. When we coupled this data format with a latent multivariate normal distribution, we found great flexibility in being able to summarize high-dimensional data for the purpose of market definition and market segmentation. The Pick any/J format is easy for respondents to use, but differs from the traditional format of using fixed-point scales (e.g., 7-point scales) to reflect finer levels of response. An examination of data collected using fixed-point scales in Chapter 3 showed us that this method of data collection is plagued by problems associated with respondents not using the scales in a uniform manner, and allowing responses to be uninformative for distinguishing among items. Responses have a tendency to bunch up near the top of the scale for most respondents, and correlations among items are uniformly moderate and significant, reflecting collinear responses.

A more formal model of decision making is introduced in Chapters 4 and 5 that is shown to overcome limitations of fixed-point scales. Data for these models reflect decisions that respondents would make if prices and product attributes were available, and provide an extension to the Pick any/J data format used in Chapters 1-3. These models provide a rational basis for making predictions of choices and demand for changes in a product configuration and competitive environment, while allowing for the presence of heterogeneous respondents. The models also integrate needs (i.e., motivations) and wants (i.e., product offerings) within the same model structure, providing for the quantification of behavior of different market segments in response to different product and pricing scenarios.

Limitations of the proposed model are discussed in Chapter 6 as advertising effects are introduced in the formation of a respondent’s consideration set and its effect on the importance of product attributes. We discuss the difference between correlation and causation, and the challenge of using cross-sectional data to make inferences about the
effects of advertising that are not easily manipulated in an experiment.

Chapter 7 shows how to tie the decision elements — the Seven Summits of Marketing Research — together using the concept of optimization. The spreadsheet decision tool automates the task of predicting the effects of changes to product, price and advertising level within the context of a competitive set of brands seeking to address the needs of potential customers. The spreadsheet produces demand and profit estimates, allowing the exploration of alternative marketing strategies that can be compared and ranked in terms of their effectiveness.

Throughout this book, we have attempted to provide a perspective on marketing and the goal-directed nature of people. We advocate collecting goal-directed data, using goal-directed models and utilizing spreadsheets for making goal-directed decisions. Doing so requires models and analysis that blends together the twin goals of inference and prediction. While we readily acknowledge that we have barely scratched the surface on the topics present in each of the chapters of this book, we hope our discussion serves as an example of how these twin goals can be achieved using recent developments in hierarchical (Bayesian) modeling and spreadsheet analysis.
7.5 Homework – Final Course Project

Each team is to propose a repositioning strategy for their brand to maximize profit. The final reports should touch on the following issues:

1. Discuss the current positioning of your brand and what you plan to change in response to a recent market development (e.g., some competitive news).

2. Whom do you select as your primary competition?

3. How will you communicate and exchange with consumers?

4. What is the context of consumption being addressed?

5. What needs are being targeted?

6. What market segments do you select? What are their names?

7. What attributes will you offer?

8. What price will you charge?

9. What is the level, content and targeting of your advertising?

The final report will be a maximum of three pages plus appendices. Your grade will be based on the completeness of your answers to the above questions, and your ability to relate your recommendation to both the IDT output and your conceptual understanding of the class material.
Appendix A: Questionnaires
Ice Cream

Version: FINAL: 8/2/2011

Study Team:
Jeff Brazell, Team Lead – jeff.brazell@themodellers.com, 801-290-3810
Matt Madden, Lead Methodologist – matt.madden@themodellers.com, 801-290-3815
Kristy Olaveson Allen, Lead Project Manager – Kristy.OlavesonAllen@themodellers.com, 801-290-3835
Melissa Flinn, Project Manager – Melissa.Flinn@themodellers.com, 801-290-3824
Matt Poulton, Lead Analyst – matt.poulton@themodellers.com, 801-290-3806

Study Objectives:
The purpose of this research is to identify consumer awareness, usage, attitudes, and behaviors related to purchasing ice cream.

Survey Quotas:

<table>
<thead>
<tr>
<th>Group</th>
<th>Definition</th>
<th>Quota</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purchase at least one container of ice cream in the past 30 days and eat ice cream at least once/month</td>
<td>600</td>
<td></td>
</tr>
</tbody>
</table>

Minimum of 50 respondents/occasion | S6 |
1=Routine dessert/snack
2=Special occasion (e.g., birthday)
3=Entertaining guests
4=Take to social gathering
5=Special treat

Minimum of 50 respondents/brand | S7 |
1=Dreyer’s
2=Blue Bunny
3=Blue Bell
4=Breyers
5=Ben & Jerry’s
6=Häagen-Dazs
7=Private Label/Store Brand

Total Sample | 600 |

Length: 20 minutes

Incidence: 80+% (98% of all households purchase ice cream, according to MakeIceCream.com)
Programmer Notes:

- Radio buttons indicate single-select questions.
- Check boxes indicate multi-select questions.
- Include "(Please select one answer.)" and "(Please select all that apply.)" for each single- and multi-select question, respectively.
- Allow only whole numbers in open-ended quantitative questions, except where noted.
- For all questions with randomized lists, anchor "Other" and "None of the above" in the bottom two positions.
- All programmer notes are in [BRACKETS, ALL CAPS, BOLD FONT].
- Besides programmer notes, respondents will not see anything in purple, bold text.
- If not otherwise specified, use the following default codes:

<table>
<thead>
<tr>
<th>Response</th>
<th>Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Don’t know/Not sure</td>
<td>97</td>
</tr>
<tr>
<td>Refused</td>
<td>98</td>
</tr>
<tr>
<td>Not applicable/None</td>
<td>99</td>
</tr>
</tbody>
</table>
**Introduction and Welcome Screen**

We appreciate your willingness to participate in our survey about ice cream. This survey is completely confidential; we will only use the information for research purposes. You should be able to complete the survey in about 15-20 minutes.

Simply answer each question by clicking inside the appropriate box. You will use the “Continue” button on your screen to move to the next question. Do not use the “Back” or “Forward” buttons on your web browser, as this may close down the survey. When you are finished, please click the “Finish” button. If you have any problems or questions, please click the “Survey Support” link at the top of the page. We will respond within 24 hours.

**Screener**

**[BASE: ALL RESPONDENTS]**

S1. How much of your household’s grocery shopping do you do?

- O1. All or most of it
- O2. About half of it
- O3. Less than half **[TERMINATE]**

**[BASE: ALL RESPONDENTS]**

S2. Which of the following best describes how often you, or someone in your household, typically purchase a container of ice cream?

- O1. Once a day
- O2. 3-5 times per week
- O3. Once a week
- O4. 2-3 times per month
- O5. Once a month
- O6. Less than once a month **[TERMINATE]**

**[BASE: ALL RESPONDENTS]**

S3. How frequently do you, yourself, eat ice cream?

- O1. Once a day or more often
- O2. 2-3 times a week
- O3. Once a week
- O4. 2-3 times a month
- O5. Once a month
- O6. Every 2-3 months **[TERMINATE]**
- O7. Less often than every 2-3 months **[TERMINATE]**
- O8. Never **[TERMINATE]**

**[MUST BUY AND EAT AT LEAST ONE A MONTH OR MORE (S2=1-5 and S3=1-5), OTHERWISE TERMINATE]**

© The Modellers.

---

**Ice Cream**

3 – Ice Cream
S4. Please indicate how many containers of ice cream (excluding individual treats or bars) you have purchased in the past month at a store. If you haven’t purchased any ice cream in the past month, please enter a zero.

[NUMERIC TEXT BOX] [TERMINATE IF S4=0]

S5. What is the first brand that comes to mind when you think about ice cream that you would buy at a grocery or specialty store?

[TEXT BOX] [BASE: ALL RESPONDENTS]

S6. Please select the occasions below that you purchase ice cream for? Select all that apply.

[RANDOMIZE]

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Minimum n= 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>For a routine dessert or snack for yourself or family</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>For a special occasion such as a birthday</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>For entertaining guests at your home</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>To take to a social gathering/party</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>As a special treat or reward for yourself</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Other (Specify)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>None of the above</td>
<td></td>
</tr>
</tbody>
</table>

[BASE: ALL RESPONDENTS]

S7. Which of the following brands of ice cream had you heard of before participating in this survey? Please check all that apply.

[RANDOMIZE LIST]

<table>
<thead>
<tr>
<th></th>
<th>Brand</th>
<th>Minimum n= 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dreyer’s</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Blue Bunny</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Blue Bell</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Breyers</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Ben &amp; Jerry’s</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Haagen-Daz</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Private Label/Store Brand (e.g., Kroger, Private Selection, Market Pantry, etc.)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>None of the above</td>
<td></td>
</tr>
</tbody>
</table>

[EXCLUSIVE] [TERMINATE] [DO NOT RANDOMIZE]
[BASE: ALL RESPONDENTS]

S8. Which of the following **brands** of ice cream have you **purchased** in the past six months? Please check all brands that apply.

[SHOW ONLY THOSE SELECTED IN S6]

<table>
<thead>
<tr>
<th></th>
<th>Brand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dreyer’s</td>
</tr>
<tr>
<td>2</td>
<td>Blue Bunny</td>
</tr>
<tr>
<td>3</td>
<td>Blue Bell</td>
</tr>
<tr>
<td>4</td>
<td>Breyers</td>
</tr>
<tr>
<td>5</td>
<td>Ben &amp; Jerry’s</td>
</tr>
<tr>
<td>6</td>
<td>Häagen-Dazs</td>
</tr>
<tr>
<td>7</td>
<td>Private Label/Store Brand(e.g., Kroger, Private Selection, Market Pantry, etc.)</td>
</tr>
<tr>
<td>99</td>
<td>None of the above</td>
</tr>
</tbody>
</table>

[EXCLUSIVE] [TERMINATE] [DO NOT RANDOMIZE]
**Main Questionnaire**

[BASE: ALL RESPONDENTS]

Q1. Of the brands that you have purchased in the past six months, how **satisfied** are you with that ice cream brand? Please use a scale from 1 to 7, where 1 means “extremely dissatisfied” and 7 means “extremely satisfied.” You may use any number between 1 and 7.

[SHOW ONLY THOSE SELECTED IN S8]

<table>
<thead>
<tr>
<th></th>
<th>Extremely dissatisfied</th>
<th>Extremely satisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[BASE: ALL RESPONDENTS]

Q2. Of the ice cream brands you have purchased in the past six months, how likely are you to purchase each brand again? Please use a scale from 1 to 7, where 1 means “not at all likely to purchase again” and 7 means “extremely likely to purchase again.” You may use any number between 1 and 7.

[SHOW ONLY THOSE SELECTED IN S8]

<table>
<thead>
<tr>
<th></th>
<th>Not at all likely to purchase again</th>
<th>Extremely likely to purchase again</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX A: QUESTIONNAIRES

[BASE: ALL RESPONDENTS]

Q3. Of the brands that you have purchased in the past six months, how likely are you to recommend each ice cream brand? Please use a scale from 1 to 7, where 1 means “not at all likely to recommend” and 7 means “extremely likely to recommend.” You may use any number between 1 and 7.

[SHOW ONLY THOSE SELECTED IN S8]

<table>
<thead>
<tr>
<th>Brand</th>
<th>Not at all likely to recommend</th>
<th>Extremely likely to recommend</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dreyer’s</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Blue Bunny</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Blue Bell</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Breyers</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Ben &amp; Jerry’s</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Häagen-Dazs</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Private Label/Store Brand</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>

[BASE: ALL RESPONDENTS]

Q4. Which of the following brands of ice cream have you considered but have not purchased in the past six months? Please check all brands that apply.

[SHOW ONLY THOSE SELECTED IN S7 AND NOT SELECTED IN S8]

1. Dreyer’s
2. Blue Bunny
3. Blue Bell
4. Breyers
5. Ben & Jerry’s
6. Häagen-Dazs
7. Private Label/Store Brand
8. None of the above [EXCLUSIVE] [DO NOT RANDOMIZE]
[BASE: THOSE WHO CONSIDERED BUT DID NOT PURCHASE ICE CREAM BRANDS IN Q4. REPEAT FOR EACH BRAND SELECTED IN Q4.]

Q5. What were the reasons you did not purchase [restore brand from Q4] after considering it?

[RANDOMIZE LIST]

- [ ] Requires too much effort
- [ ] Too expensive
- [ ] Contains artificial flavors and ingredients
- [ ] Too fattening
- [ ] Guests wouldn’t like it
- [ ] Didn’t see the flavor I wanted
- [ ] Not the quality I wanted
- [ ] Just chose a different brand
- [ ] None of the above

[EXCLUSIVE] [DO NOT RANDOMIZE]

[BASE: ALL RESPONDENTS]

Q6. Where do you typically purchase ice cream? Please select all the places you typically purchase it.

- [ ] Traditional supermarket (such as Albertsons, Safeway, Vons, Stop and Shop, etc.)
- [ ] Natural Organic store (such as Whole Foods, Wild Oats, etc.)
- [ ] Specialty supermarket (such as Trader Joe’s)
- [ ] Discount Mass Merchandise Store/Supermarket (such as Target, Walmart, etc.)
- [ ] Wholesale/Club store (such as Sam’s Club, Costco, BJ’s, etc.)
- [ ] Convenience store (such as 7-Eleven, ampm, etc.)
- [ ] Ice cream specialty store – pre-packaged containers only (such as Marble Slab, Cold Stone Creamery, Baskin Robbins)
- [ ] Other (Specify) ____________________________
**APPENDIX A: QUESTIONNAIRES**

[BASE: ALL RESPONDENTS]

Q8. Of the brands you have purchased, please select all of the occasions for which you purchased that brand.

[ONLY SHOW BRANDS SELECTED IN S8 AND OCCASIONS SELECTED IN S6]

<table>
<thead>
<tr>
<th>Brand</th>
<th>For a routine dessert or snack</th>
<th>For a special occasion such as a birthday</th>
<th>For entertaining guests at home</th>
<th>To take to a social gathering or party</th>
<th>As a special treat or reward for yourself</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Dreyer’s</td>
<td>☐ 1</td>
<td>☐ 2</td>
<td>☐ 3</td>
<td>☐ 4</td>
<td>☐ 5</td>
</tr>
<tr>
<td>2 Blue Bunny</td>
<td>☐ 1</td>
<td>☐ 2</td>
<td>☐ 3</td>
<td>☐ 4</td>
<td>☐ 5</td>
</tr>
<tr>
<td>3 Blue Bell</td>
<td>☐ 1</td>
<td>☐ 2</td>
<td>☐ 3</td>
<td>☐ 4</td>
<td>☐ 5</td>
</tr>
<tr>
<td>4 Breyers</td>
<td>☐ 1</td>
<td>☐ 2</td>
<td>☐ 3</td>
<td>☐ 4</td>
<td>☐ 5</td>
</tr>
<tr>
<td>5 Ben &amp; Jerry’s</td>
<td>☐ 1</td>
<td>☐ 2</td>
<td>☐ 3</td>
<td>☐ 4</td>
<td>☐ 5</td>
</tr>
<tr>
<td>6 Häagen-Dazs</td>
<td>☐ 1</td>
<td>☐ 2</td>
<td>☐ 3</td>
<td>☐ 4</td>
<td>☐ 5</td>
</tr>
<tr>
<td>7 Private Label/Store Brand</td>
<td>☐ 1</td>
<td>☐ 2</td>
<td>☐ 3</td>
<td>☐ 4</td>
<td>☐ 5</td>
</tr>
</tbody>
</table>
Discrete Choice Task

Intro: The next part of the survey will be similar to when you are on a typical shopping trip and looking for ice cream. In the store, you find several different types of ice cream that come in different flavors, sizes, brands, and are different prices. After considering the options, please tell us how many of each item you would purchase, then select the ONE that is your favorite.

Between first and second choice question: On the next few screens we’re going to show you a new set of ice cream with different features, which you can think of as ice cream available on a different day or at a different store. All you have to do is tell us how many of each you would buy and which one is your favorite. We will do this several times so we can learn about what types of ice cream you like.

Heading for second through last choice questions: Please tell us how many of each ice cream product you would purchase on a typical shopping trip and which is your favorite.
**Example Shelf Task**

Please treat this screen as if you were shopping at your local grocery store for ice cream. Please indicate how many of each product you would purchase and which of these products is your *favourite* product.

**Scenario 1 of 12**

<table>
<thead>
<tr>
<th>Option 1</th>
<th>Option 2</th>
<th>Option 3</th>
<th>Option 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dreyer's Vanilla</td>
<td>Dreyer's Chocolate</td>
<td>Blue Bell Bourbon</td>
<td>Blue Bell Fudge Pecan</td>
</tr>
<tr>
<td>(4 servings)</td>
<td>(8 servings)</td>
<td>(8 servings)</td>
<td>(4 servings)</td>
</tr>
<tr>
<td>$1.99</td>
<td>$2.49</td>
<td>$3.33</td>
<td>$2.99</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Option 5</th>
<th>Option 6</th>
<th>Option 7</th>
<th>Option 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Häagen-Dazs Vanilla Block</td>
<td>Häagen-Dazs Mulligan</td>
<td>Häagen-Dazs Salted Caramel</td>
<td>Häagen-Dazs Fudge Ripple</td>
</tr>
<tr>
<td>(16 servings)</td>
<td>(4 servings)</td>
<td>(8 servings)</td>
<td>(1 servings)</td>
</tr>
<tr>
<td>$2.89</td>
<td>$2.49</td>
<td>$3.49</td>
<td>$2.99</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Option 9</th>
<th>Option 10</th>
<th>Option 11</th>
<th>Option 12</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Storebrand</em> Vanilla</td>
<td><em>Storebrand</em> Mulligan</td>
<td><em>Storebrand</em> Cookies and Cream</td>
<td><em>Storebrand</em> Mint Chip</td>
</tr>
<tr>
<td>(16 servings)</td>
<td>(4 servings)</td>
<td>(8 servings)</td>
<td>(1 servings)</td>
</tr>
<tr>
<td>$2.49</td>
<td>$4.59</td>
<td>$3.53</td>
<td>$4.99</td>
</tr>
</tbody>
</table>

**Of the options above, please specify your *favourite* product.**

- [ ] Option 1
- [ ] Option 2
- [ ] Option 3
- [ ] Option 4
- [ ] Option 5
- [ ] Option 6
- [ ] Option 7
- [ ] Option 8
- [ ] Option 9
- [ ] Option 10
- [ ] Option 11
- [ ] Option 12

**How many of each option above would you purchase?**

<table>
<thead>
<tr>
<th>Option 1</th>
<th>Option 2</th>
<th>Option 3</th>
<th>Option 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**(CALCULATE) Total Price**

**(Count(s) Summary)**

Total cost of ice cream purchase $_____

Respondent must click this if not purchasing ice cream.
### Main Questionnaire (continued)

**[BASE: ALL RESPONDENTS]**

**Q9.** The list below includes reasons some people choose to buy ice cream. Please select the items that apply to you or others in your household.

**[RANDOMIZE]**

<table>
<thead>
<tr>
<th></th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ice cream helps me relax and enjoy life.</td>
</tr>
<tr>
<td>2</td>
<td>Ice cream is a wholesome treat.</td>
</tr>
<tr>
<td>3</td>
<td>Ice cream gives me something fun to do with family and/or friends.</td>
</tr>
<tr>
<td>4</td>
<td>Ice cream provides relief from regular life.</td>
</tr>
<tr>
<td>5</td>
<td>Ice cream helps make an event special.</td>
</tr>
<tr>
<td>6</td>
<td>Ice cream reminds me of my childhood.</td>
</tr>
<tr>
<td>7</td>
<td>I’m not picky – ice cream is ice cream.</td>
</tr>
<tr>
<td>8</td>
<td>I am amazed at how ice cream tastes are created.</td>
</tr>
<tr>
<td>9</td>
<td>I enjoy trying new flavors of ice cream.</td>
</tr>
<tr>
<td>10</td>
<td>I love having lots of flavors to choose from.</td>
</tr>
<tr>
<td>11</td>
<td>Ice cream is good for entertaining.</td>
</tr>
<tr>
<td>12</td>
<td>Everyone loves ice cream.</td>
</tr>
</tbody>
</table>
| 99| None of the above                                           | **[EXCLUSIVE; ALWAYS SHOW LAST]**
**APPENDIX A: QUESTIONNAIRES**

[BASE: ALL RESPONDENTS]

**Q10.** Now we would like to explore your attitudes and opinions about ice cream generally. Please select all of the statements below that you feel describe you.

[RANDOMIZE]

| 1 | I only eat ice cream on special occasions. |
| 2 | I eat ice cream all the time. |
| 3 | I only buy ice cream when it is on sale. |
| 4 | I buy the ice cream I want, regardless of price. |
| 5 | I buy the same flavor of ice cream every time. |
| 6 | I buy the same brand of ice cream every time. |
| 7 | I buy ice cream in bulk. |
| 8 | I prefer smaller containers of ice cream that may be more expensive. |
| 9 | I look for ice cream every time I go to the store. |
| 10 | I always buy low fat ice cream. |
| 11 | I look for the healthiest ice cream on the shelf. |
| 12 | I avoid artificial colors and flavors. |
| 13 | I prefer to eat all-natural/organic ice cream. |

[EXCLUSIVE; ALWAYS SHOW LAST]

| 99 | None of the above |
[BASE: ALL RESPONDENTS]

**Q11.** From what you have heard or experienced, please indicate how well each statement describes [INSERT BRAND FROM S7].

[SHOW ONE BRAND AT A TIME FROM THOSE SELECTED IN S7– RANDOMIZE STATEMENTS FOR EACH NEW SURVEY BUT KEEP IN SAME ORDER FOR EACH RESPONDENT.]

<table>
<thead>
<tr>
<th>Does not Describe at all</th>
<th>Describes extremely well</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Relaxed</td>
<td>Relaxing</td>
</tr>
<tr>
<td>Wholesome</td>
<td>Funk</td>
</tr>
<tr>
<td>Exciting</td>
<td>Premium — uses better quality ingredients</td>
</tr>
<tr>
<td>Memorable</td>
<td>Buy as a special treat but not regularly</td>
</tr>
<tr>
<td>Good enough to keep on hand for regular consumption</td>
<td>Interesting</td>
</tr>
<tr>
<td>Tastes better than most brands</td>
<td>Offers a wide variety of flavors</td>
</tr>
<tr>
<td>Enjoyable</td>
<td>Best value for the price</td>
</tr>
<tr>
<td>Natural/organic</td>
<td>Low calorie</td>
</tr>
<tr>
<td>Great for the whole family</td>
<td>Great for guests</td>
</tr>
</tbody>
</table>

[BASE: ALL RESPONDENTS]

**Q12.** Where do you typically get information about food and nutrition? Please select all that apply.

[RANDOMIZE]

- [ ] TV
- [ ] Radio
- [ ] Print newspaper
- [ ] Print magazine
- [ ] Online – website
- [ ] Online – social media
- [ ] Other (Specify) __________________________________________________________________________
- [ ] None of the above [EXCLUSIVE]
[BASE: ALL RESPONDENTS]

Q13. Now we’d like you to think about advertising you may have seen or heard for various brands of ice cream. For each brand, please indicate about how many advertisements you have seen or heard in the past month in the following types of media.

[SHOW THIS QUESTION ONCE FOR EACH OF THE ICE CREAM BRANDS AWARE OF – SELECTED IN S7]

<table>
<thead>
<tr>
<th>MEDIA</th>
<th>INSERT BRAND FROM S7</th>
</tr>
</thead>
<tbody>
<tr>
<td>TV</td>
<td>__________1</td>
</tr>
<tr>
<td>Radio</td>
<td>__________2</td>
</tr>
<tr>
<td>Print newspaper</td>
<td>__________3</td>
</tr>
<tr>
<td>Print magazine</td>
<td>__________4</td>
</tr>
<tr>
<td>Online – website</td>
<td>__________5</td>
</tr>
<tr>
<td>Online – social media</td>
<td>__________6</td>
</tr>
<tr>
<td>Store flier</td>
<td>__________7</td>
</tr>
<tr>
<td>In store/On shelf</td>
<td>__________8</td>
</tr>
</tbody>
</table>

[BASE: ALL RESPONDENTS]

Q14. Please record the approximate number of hours you spend each day in the following activity. If you want to record a fraction of an hour, please show this with a decimal (e.g., .5 hrs).

<table>
<thead>
<tr>
<th>MEDIA</th>
<th><em><strong>.</strong></em></th>
</tr>
</thead>
<tbody>
<tr>
<td>TV</td>
<td>___.___1</td>
</tr>
<tr>
<td>Radio</td>
<td>___.___2</td>
</tr>
<tr>
<td>Print newspaper</td>
<td>___.___3</td>
</tr>
<tr>
<td>Print magazine</td>
<td>___.___4</td>
</tr>
<tr>
<td>Online – website</td>
<td>___.___5</td>
</tr>
<tr>
<td>Online – social media</td>
<td>___.___6</td>
</tr>
<tr>
<td>Other (Specify)</td>
<td>___.___7</td>
</tr>
<tr>
<td>None of the above</td>
<td>[EXCLUSIVE]</td>
</tr>
</tbody>
</table>
Demographics

[BASE: ALL RESPONDENTS]

D1. Please select your gender.

- Male
- Female

[BASE: ALL RESPONDENTS]

D2. Into which group does your age fall as of your last birthday?

- Under 18
- 18-24
- 25-34
- 35-44
- 45-54
- 55-64
- 65 or older

[BASE: ALL RESPONDENTS]

D3. Which of the following best describes your marital status? (Please select one response.)

- Single, not living with a domestic partner
- Single, living with a domestic partner
- Married
- Divorced
- Separated
- Widowed
- Prefer not to say

[BASE: ALL RESPONDENTS]

D4. Are there any children under the age of 18 currently living with you, either on a part-time or full-time basis?

- Yes
- No

[BASE: ALL RESPONDENTS]

D5. How many people, including yourself, live in your household?

Enter number of people: _______ [WHOLE NUMBER ONLY. RANGE: 1-20]
APPENDIX A: QUESTIONNAIRES

[BASE: ALL RESPONDENTS]

D6. Which of the following best describes your highest level of education? (Please select one response.)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High school or less [NOTES]</td>
</tr>
<tr>
<td>2</td>
<td>Trade/technical school</td>
</tr>
<tr>
<td>3</td>
<td>Some college or Associate’s degree</td>
</tr>
<tr>
<td>4</td>
<td>Graduated college/Bachelor’s degree</td>
</tr>
<tr>
<td>5</td>
<td>Attended graduate school</td>
</tr>
<tr>
<td>6</td>
<td>Advanced degree (Master’s, Ph.D.)</td>
</tr>
</tbody>
</table>

[BASE: ALL RESPONDENTS]

D7. Which of the following best describes your current employment status? (Please select one response.)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Employed full-time (30+ hours) [NOTES]</td>
</tr>
<tr>
<td>2</td>
<td>Employed part-time</td>
</tr>
<tr>
<td>3</td>
<td>Not currently employed</td>
</tr>
<tr>
<td>4</td>
<td>Student</td>
</tr>
<tr>
<td>5</td>
<td>Retired</td>
</tr>
<tr>
<td>6</td>
<td>Homemaker</td>
</tr>
</tbody>
</table>

[BASE: ALL RESPONDENTS]

D8. Which of the following best describes your racial or ethnic background? (Please select one response.)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Caucasian [NOTES]</td>
</tr>
<tr>
<td>2</td>
<td>African-American</td>
</tr>
<tr>
<td>3</td>
<td>Hispanic or Latino</td>
</tr>
<tr>
<td>4</td>
<td>Asian/Pacific Islander</td>
</tr>
<tr>
<td>5</td>
<td>Other</td>
</tr>
<tr>
<td>6</td>
<td>Prefer not to say</td>
</tr>
</tbody>
</table>

[BASE: ALL RESPONDENTS]

D9. Which of the following categories best describes your annual household income before taxes? (Please select one response.)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Under $50,000 [NOTES]</td>
</tr>
<tr>
<td>2</td>
<td>$50,000 to just under $75,000</td>
</tr>
<tr>
<td>3</td>
<td>$75,000 to just under $100,000</td>
</tr>
<tr>
<td>4</td>
<td>$100,000 to just under $125,000</td>
</tr>
<tr>
<td>5</td>
<td>$125,000 to just under $150,000</td>
</tr>
<tr>
<td>6</td>
<td>$150,000 to just under $175,000</td>
</tr>
<tr>
<td>7</td>
<td>$175,000 to just under $200,000</td>
</tr>
<tr>
<td>8</td>
<td>$200,000 or more</td>
</tr>
<tr>
<td>9</td>
<td>Prefer not to say</td>
</tr>
<tr>
<td>10</td>
<td>Don’t know</td>
</tr>
</tbody>
</table>

© The Modellers.
[BASE: ALL RESPONDENTS]

D10. In what state do you reside?

Enter state: _______ [DROPDOWN BOX]

Closing

Thank you for participating in our survey!
**Vacation**


**Study Team:**
Jeff Brazell, Team Lead – jeff.brazell@themodellers.com, 801-290-3810  
Matt Madden, Lead Methodologist – matt.madden@themodellers.com, 801-290-3815  
Kristy Olaveson Allen, Lead Project Manager – Kristy.OlavesonAllen@themodellers.com, 801-290-3835  
Melissa Flinn, Project Manager – Melissa.Flinn@themodellers.com, 801-290-3824  
Matt Poulton, Lead Analyst – matt.poulton@themodellers.com, 801-290-3806

**Study Objectives:**
The purpose of this research is to identify consumer awareness, usage, attitudes, and behaviors related to purchasing ice cream.

**Survey Quotas:**

<table>
<thead>
<tr>
<th>Group</th>
<th>Definition</th>
<th>Quota</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have taken a vacation in the past year</td>
<td></td>
<td>600</td>
</tr>
<tr>
<td>Minimum of 50 respondents / travel purpose</td>
<td>S7</td>
<td></td>
</tr>
<tr>
<td>1=Annual family vacation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2=Family reunion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3=Special occasion (e.g., birthday)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4=School breaks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5=Holidays</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6=Just to get away</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Minimum of 50 respondents / location | S8 | |
| 1=Disney World – Magic Kingdom |  |
| 2=Disney World – Epcot |  |
| 3=Disney’s Animal Kingdom |  |
| 4=Disney’s Hollywood Studios |  |
| 5=Universal Studios Florida |  |
| 6=Universal’s Islands of Adventure |  |
| 7=Busch Gardens in Florida |  |

| Total Sample | 600 |

**Length:** 20 minutes

**Incidence:** ???
**Programmer Notes:**

- Radio buttons indicate single-select questions.
- Check boxes indicate multi-select questions.
- Include "(Please select one answer.)" and "(Please select all that apply.)" for each single- and multi-select question, respectively.
- Allow only whole numbers in open-ended quantitative questions, except where noted.
- For all questions with randomized lists, anchor “Other” and “None of the above” in the bottom two positions.
- All programmer notes are in **[BRACKETS, ALL CAPS, BOLD FONT].**
- Besides programmer notes, respondents will not see anything in purple, bold text.
- If not otherwise specified, use the following default codes:

<table>
<thead>
<tr>
<th>Response</th>
<th>Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Don’t know/Not sure</td>
<td>97</td>
</tr>
<tr>
<td>Refused</td>
<td>98</td>
</tr>
<tr>
<td>Not applicable/None</td>
<td>99</td>
</tr>
</tbody>
</table>
Introduction and Welcome Screen

We appreciate your willingness to participate in our survey about travel. This survey is completely confidential; we will only use the information for research purposes. You should be able to complete the survey in about 15-20 minutes.

Simply answer each question by clicking inside the appropriate box. You will use the “Next” button on your screen to move to the next question. Do not use the “Back” or “Forward” buttons on your web browser, as this may close down the survey. When you are finished, please click the “Finish” button. If you have any problems or questions, please click the “Survey Support” link at the top of the page. We will respond within 24 hours.

Screener

[BASE: ALL RESPONDENTS]

S1. Which best describes your decision making role as it relates to travel within your household?

1. I am the primary decision maker.
2. I share decision making responsibilities with someone else in my household.
3. I do not make or influence decisions about travel. [TERMINATE]

[BASE: ALL RESPONDENTS]

S2. Which of the following best describes how many times you typically travel each year for leisure purposes with family and/or friends?

1. Several times a year
2. At least once per year
3. At least once every two years
4. At least once every three to five years
5. Less than once every five years [TERMINATE]
6. [TERMINATE]

[BASE: ALL RESPONDENTS]

S3. When was the last time you personally planned and took a vacation with family and/or friends?

1. Within the past 6 months
2. At least 6 months but less than a year ago
3. At least one year but less than two years ago
4. At least two years but less than three years ago
5. At least three years but less than five years ago
6. More than five years ago [TERMINATE]
[BASE: ALL RESPONDENTS]

S4. Please indicate how many trips (excluding all business travel) you have taken in the past five years. If you haven’t been on vacation in the past five years, please enter a zero.

[NUMERIC TEXT BOX]
[TERMINATE IF S4=0]

[BASE: ALL RESPONDENTS]

S5. On average, how many days do you spend on a typical vacation?

[NUMERIC TEXT BOX]

[BASE: ALL RESPONDENTS]

S6. What is the first destination that comes to mind when you think about the ideal vacation with family or friends?

[TEXT BOX]

[BASE: ALL RESPONDENTS]

S7. Please select the occasions from the list below that you travel for? Select all that apply.

[RANDOMIZE]

☐ 1. An annual family vacation Minimum n= 50
☐ 2. A family reunion Minimum n= 50
☐ 3. A special occasion such as a birthday or anniversary Minimum n= 50
☐ 4. School breaks Minimum n= 50
☐ 5. Holidays Minimum n= 50
☐ 6. Just to get away Minimum n= 50
☐ 7. Other (Specify) ____________________________
☐ 99. None of the above [EXCLUSIVE] [TERMINATE]
### [BASE: ALL RESPONDENTS]

**S8.** Which of the following theme parks had you heard of before participating in this survey? Please check all that apply.

[**RANDOMIZE LIST**]

<table>
<thead>
<tr>
<th></th>
<th>Theme Park</th>
<th>Minimum n= 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Disney World – Magic Kingdom</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Disney World – Epcot</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Disney’s Animal Kingdom</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Disney’s Hollywood Studios</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Universal Studios Florida</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Universal’s Islands of Adventure (Includes Wizarding World of Harry Potter)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Busch Gardens in Florida</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>None of the above</td>
<td>[EXCLUSIVE] [TERMINATE] [DO NOT RANDOMIZE]</td>
</tr>
</tbody>
</table>

### [BASE: ALL RESPONDENTS]

**S9.** Which of the following theme parks have you **visited in the past five years**? Please check all brands that apply.

[**SHOW ONLY THOSE SELECTED IN S8**]

<table>
<thead>
<tr>
<th></th>
<th>Theme Park</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Disney World – Magic Kingdom</td>
</tr>
<tr>
<td>2</td>
<td>Disney World – Epcot</td>
</tr>
<tr>
<td>3</td>
<td>Disney’s Animal Kingdom</td>
</tr>
<tr>
<td>4</td>
<td>Disney’s Hollywood Studios</td>
</tr>
<tr>
<td>5</td>
<td>Universal Studios Florida</td>
</tr>
<tr>
<td>6</td>
<td>Universal’s Islands of Adventure (Includes Wizarding World of Harry Potter)</td>
</tr>
<tr>
<td>7</td>
<td>Busch Gardens in Florida</td>
</tr>
<tr>
<td>9</td>
<td>None of the above</td>
</tr>
</tbody>
</table>
[BASE: ALL RESPONDENTS]

**S10.** Which of the following theme parks do you **plan to visit in the next year or two**? Please check all brands that apply.

**[SHOW ONLY THOSE NOT SELECTED IN S8]**

<table>
<thead>
<tr>
<th></th>
<th>Which of the following theme parks do you plan to visit in the next year or two? Please check all brands that apply.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Disney World – Magic Kingdom</td>
</tr>
<tr>
<td>2</td>
<td>Disney World – Epcot</td>
</tr>
<tr>
<td>3</td>
<td>Disney’s Animal Kingdom</td>
</tr>
<tr>
<td>4</td>
<td>Disney’s Hollywood Studios</td>
</tr>
<tr>
<td>5</td>
<td>Universal Studios Florida</td>
</tr>
<tr>
<td>6</td>
<td>Universal’s Islands of Adventure (Includes Wizarding World of Harry Potter)</td>
</tr>
<tr>
<td>7</td>
<td>Busch Gardens in Florida</td>
</tr>
<tr>
<td>99</td>
<td>None of the above</td>
</tr>
</tbody>
</table>

[EXCLUSIVE] [TERMINATE IF S9=99] [DO NOT RANDOMIZE]

[IF BOTH S9 AND S10=99, THEN TERMINATE. IF ONLY 1 OF THE TWO=99, CONTINUE]
Main Questionnaire

[BASE: S9≠99]

Q1. Of the theme parks that you have visited in the past five years, how satisfied are you with that theme park brand? Please use a scale from 1 to 7, where 1 means "extremely dissatisfied" and 7 means "extremely satisfied." You may use any number between 1 and 7.

[SHOW ONLY THOSE SELECTED IN S9]

<table>
<thead>
<tr>
<th></th>
<th>Extremely dissatisfied</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disney World – Magic Kingdom</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>Disney World – Epcot</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>Disney’s Animal Kingdom</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>Disney’s Hollywood Studios</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>Universal Studios Florida</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>Universal’s Islands of Adventure (Includes Wizarding World of Harry Potter)</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>Busch Gardens in Florida</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
</tr>
</tbody>
</table>

[BASE: S9≠99]

Q2. Of the theme parks you have visited in the past five years, how likely are you to visit each brand again? Please use a scale from 1 to 7, where 1 means "not at all likely to visit again" and 7 means "extremely likely to visit again." You may use any number between 1 and 7.

[SHOW ONLY THOSE SELECTED IN S9]

<table>
<thead>
<tr>
<th></th>
<th>Not at all likely to visit again</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disney World – Magic Kingdom</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>Disney World – Epcot</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>Disney’s Animal Kingdom</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>Disney’s Hollywood Studios</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>Universal Studios Florida</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>Universal’s Islands of Adventure (Includes Wizarding World of Harry Potter)</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>Busch Gardens in Florida</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
</tr>
</tbody>
</table>
[BASE: S9≠99]

Q3. Of the theme parks that you have visited in the past five years, how likely are you to recommend each theme park? Please use a scale from 1 to 7, where 1 means “not at all likely to recommend” and 7 means “extremely likely to recommend.” You may use any number between 1 and 7.

[SHOW ONLY THOSE SELECTED IN S9]

<table>
<thead>
<tr>
<th>Theme Park</th>
<th>Not at all likely to recommend</th>
<th>Extremely likely to recommend</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Disney World – Magic Kingdom</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>2. Disney World – Epcot</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>3. Disney’s Animal Kingdom</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>4. Disney’s Hollywood Studios</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>5. Universal Studios Florida</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>6. Universal’s Islands of Adventure (Includes Wizarding World of Harry Potter)</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>7. Busch Gardens in Florida</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

[BASE: ALL RESPONDENTS]

Q4. Which of the following theme park brands have you considered but have not visited in the past five years? Please check all brands that apply.

[SHOW ONLY THOSE SELECTED IN S8 AND NOT SELECTED IN S9]

☑ 1. Disney World – Magic Kingdom
☑ 2. Disney World – Epcot
☑ 3. Disney’s Animal Kingdom
☑ 4. Disney’s Hollywood Studios
☑ 5. Universal Studios Florida
☑ 6. Universal’s Islands of Adventure (Includes Wizarding World of Harry Potter)
☑ 7. Busch Gardens in Florida
☐ 99 None of the above [EXCLUSIVE] [TERMINATE] [DO NOT RANDOMIZE]
APPENDIX A: QUESTIONNAIRES

[BASE: THOSE WHO CONSIDERED BUT DID NOT VISIT IN Q4. REPEAT FOR EACH BRAND SELECTED IN Q4.]

Q5. What were the reasons you did not visit [RESTORE BRAND FROM Q4] after considering it?

[RANDOMIZE LIST – KEEP IN THE SAME ORDER FOR EACH BRAND SHOWN]

- Did not have the time to visit
- Too expensive
- Too planned out and artificial
- Worried about family safety
- Poor location
- Did not offer the entertainment I was looking for
- Not the quality I wanted
- Just chose a different location
- None of the above

[EXCLUSIVE] [DO NOT RANDOMIZE]

Q6. When planning a vacation for your family and/or friends, what resources do you typically use to make reservations?

[MULTI-SELECT - RANDOMIZE]

- I go directly to the hotel/airline website to book a reservation.
- I use a travel search engine sites (Travelocity, Hotels.com, Orbitz, etc.).
- I use discount travel sites (Priceline, Travel Zoo, etc.).
- I call a travel agent/agency.
- Other (Specify) __________________________

Q7. Of the brands you have visited, please select all of the occasions for which you visited that theme park.

[ONLY SHOW OCCASIONS SELECTED IN S7 AND BRANDS SELECTED IN S9]

<table>
<thead>
<tr>
<th>Brand</th>
<th>Annual family vacation</th>
<th>Family Reunion</th>
<th>Special occasion</th>
<th>School break</th>
<th>Holiday</th>
<th>Just to get away</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disney World – Magic Kingdom</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Disney World – Epcot</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Disney’s Animal Kingdom</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Disney’s Hollywood Studios</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Universal Studios Florida</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Universal’s Islands of Adventure (Includes Wizarding World of Harry Potter)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Busch Gardens in Florida</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
Discrete Choice Task

In the next part of the survey, we would like you to consider taking your family on a trip to a Florida to visit theme parks such as Disney, Universal Studios, and Busch Gardens. Could you first please tell us how many people you would likely be traveling with and the number of days you would likely spend on a Florida theme park vacation.

[BASE: ALL RESPONDENTS]
T1. Number of people you would travel with: [NUMERIC TEXT BOX]

[BASE: ALL RESPONDENTS]
T2. Number of days you would spend on a Florida theme park vacation: [NUMERIC TEXT BOX]

Next we are going to show you multiple screens on which we will be showing you various vacation packages. Please look at each of the six options and consider the theme park and the available offerings. Then, select the option you most prefer and indicate how many days you would buy tickets for each theme park. Please take a moment to review the descriptions of the theme parks and offerings before continuing to the next activity.

INSERT PARK AND ATTRIBUTE DESCRIPTIONS — GLOSSARY

- **Disney World – Magic Kingdom** - Magic Kingdom theme park, one of 4 Theme Parks in Walt Disney World Resort, captures the enchantment of fairy tales with exciting entertainment, classic attractions, backstage tours and beloved Disney Characters. Designed like a wheel with the hub in front of Cinderella Castle, pathways spoke out across the 107 acres of Magic Kingdom theme park and lead to 6 whimsical lands.

- **Disney World – Epcot** - Epcot theme park, one of 4 Theme Parks at Walt Disney World Resort, sprawls across 300 acres—twice the size of Magic Kingdom theme park—and is divided into Future World and World Showcase. Future World is full of seasonal attractions—including one of the fastest attractions in all Disney Parks—as well as inspiring entertainment and shows, all of which focus on technological advancements, innovation, and wonder. World Showcase is a collective of pavilions that wrap around the World Showcase Lagoon. Inside the pavilions, find shops, attractions, and restaurants that represent the culture and cuisine of eleven countries.

- **Disney World – Animal Kingdom** - Disney's Animal Kingdom theme park, one of 4 Theme Parks at Walt Disney World Resort, is full of attractions, adventures and entertainment that reflect Walt Disney's dedication to nature and conservation, and in doing so, the Park leads the way in animal care, education and research. Home to more than 1,700 animals from 250 species and sprawling across 500 acres of lush landscape, Disney's Animal Kingdom theme park is the largest animal-themed park in the world!

- **Disney World – Hollywood Studios** - Disney's Hollywood Studios theme park, one of 4 Theme Parks in Walt Disney World Resort, offers behind-the-scenes glimpses of Hollywood-style action with live shows, thrilling attractions, backstage tours and special events that only happen in this Disney Park dedicated to entertainment. The glitz and glamour of the Hollywood Heydays from the 1930s and 1940s are captured by the neon, chrome, art deco and modern architecture throughout Disney's Hollywood Studios theme park.

- **Universal Studios Florida** - Go behind the scenes, beyond the screen, and jump right into the action of your favorite movies at Universal Studios®, the world’s premier movie and TV based theme park. At this real, working film and TV production facility you’ll find an amazing array of rides, shows, movie sets and attractions that put you right in the picture. You’ll enjoy themed dining and shopping, a variety of exciting special events throughout the year, and you might even catch a real film crew at work on the backlot. From entertaining shows that take you behind the scenes of the movie-making process to state-of-the-art rides that make you part of the action, nobody brings the magic of the movies . . . television . . . and music to life like Universal Studios.
• **Universal’s Islands of Adventure** - Around every bend is another epic adventure. Around every corner awaits another once-in-a-lifetime thrill. Take an unforgettable journey through the uniquely themed islands of Universal’s Islands of Adventure®, where the world’s most cutting edge rides, shows and interactive attractions bring your favorite stories, myths, cartoons, comic book heroes and children’s tales to life. The Wizarding World of Harry Potter™ is at Islands of Adventure.

• **Busch Gardens Florida** - Near Florida’s beautiful Gulf Coast, you’ll find yourself on the edge of Africa at Busch Gardens in Tampa Bay. Experience up-close encounters with amazing animals, take an unforgettable safari across the Serengeti Plains, test your courage on pulse-pounding roller coasters, take in a stage show, and colorful live entertainment all year long.

• **Premium speed ticket at theme park**: Allows you to save a place in line while you enjoy the rest of the theme park and return during a window of time to enjoy minimal wait times.

• **Mobile app for wait times, character meetings, etc**: Allows you to see what times and activities from wherever you may be in the park.

• **Special extended hours**: Allows you to enjoy the park before or after regular closing hours.

[Pipe in costs based on the number of people they would travel with in T1 and have them allocate the number of days based on the number of days they would spend on a Florida vacation in T2]
Main Questionnaire (continued)

[BASE: ALL RESPONDENTS]

Q8. Now we would like to explore your attitudes and opinions about travel generally. Please select all of the statements below that you feel describe you.

[RANDOMIZE]

<table>
<thead>
<tr>
<th></th>
<th>Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I go on vacation to relax and enjoy life.</td>
</tr>
<tr>
<td>2</td>
<td>I want a wholesome vacation experience.</td>
</tr>
<tr>
<td>3</td>
<td>I go on vacations to have fun with family and/or friends.</td>
</tr>
<tr>
<td>4</td>
<td>Vacations provide relief from regular life.</td>
</tr>
<tr>
<td>5</td>
<td>Vacations are where lasting memories are made.</td>
</tr>
<tr>
<td>6</td>
<td>Vacations remind me of my childhood.</td>
</tr>
<tr>
<td>7</td>
<td>When I travel on vacation, I don’t have strong feelings about where to go</td>
</tr>
<tr>
<td>8</td>
<td>I am amazed at how the vacation attractions work.</td>
</tr>
<tr>
<td>9</td>
<td>I enjoy trying new attractions when I’m on vacation.</td>
</tr>
<tr>
<td>10</td>
<td>I love having a variety of attractions when I’m on vacation.</td>
</tr>
<tr>
<td>11</td>
<td>Florida vacations help me enjoy my family and friends more.</td>
</tr>
<tr>
<td>12</td>
<td>Everyone loves going on a Florida vacation.</td>
</tr>
<tr>
<td>13</td>
<td>None of the above</td>
</tr>
</tbody>
</table>
### [BASE: ALL RESPONDENTS]

**Q9.** Now we would like to explore your attitudes and opinions about Florida vacations generally. Please select all of the statements below that you feel describe you.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Q9.</td>
<td>Now we would like to explore your attitudes and opinions about Florida vacations generally. Please select all of the statements below that you feel describe you.</td>
</tr>
<tr>
<td>[RANDOMIZE]</td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>I only travel for special occasions.</td>
</tr>
<tr>
<td>2.</td>
<td>I travel at every available opportunity.</td>
</tr>
<tr>
<td>3.</td>
<td>I always look for the least expensive vacation options.</td>
</tr>
<tr>
<td>4.</td>
<td>I travel wherever I want, regardless of price.</td>
</tr>
<tr>
<td>5.</td>
<td>I find myself traveling to the same locations every year.</td>
</tr>
<tr>
<td>6.</td>
<td>I find myself going to the same attractions every year.</td>
</tr>
<tr>
<td>7.</td>
<td>I prefer going on long vacations.</td>
</tr>
<tr>
<td>8.</td>
<td>I’d rather go on a shorter, more expensive vacation than a longer, cheaper one.</td>
</tr>
<tr>
<td>9.</td>
<td>I frequently shop for vacation deals.</td>
</tr>
<tr>
<td>10.</td>
<td>I plan my vacation to be as economical as possible.</td>
</tr>
<tr>
<td>11.</td>
<td>I plan vacations that have relatively low risks of danger.</td>
</tr>
<tr>
<td>12.</td>
<td>I avoid highly commercialized vacation sites.</td>
</tr>
<tr>
<td>13.</td>
<td>I prefer to vacation as close to nature as possible (not man-made but natural attractions).</td>
</tr>
<tr>
<td>99.</td>
<td>None of the above [EXCLUSIVE]</td>
</tr>
</tbody>
</table>
[BASE: ALL RESPONDENTS]

Q10. From what you have heard or experienced, please indicate how well each statement describes [INSERT BRAND FROM S8].

[SHOW ONE BRAND AT A TIME FROM THOSE SELECTED IN S8 – RANDOMIZE STATEMENTS FOR EACH NEW SURVEY BUT KEEP IN SAME ORDER FOR EACH RESPONDENT.]

<table>
<thead>
<tr>
<th>Statement</th>
<th>Does not describe at all</th>
<th>Describes extremely well</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Relaxing</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>2 Wholesome</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>3 Fun</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>4 Exciting</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>5 Premium – provides a better quality experience</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>6 Memorable</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>7 Is a one-time only trip</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>8 Good enough to go back for another vacation</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>9 Interesting</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>10 Is more fun that most locations</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>11 Offers a wide variety of entertainment options</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>12 Enjoyable</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>13 Best value for the price</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>14 Authentic</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>15 Safe</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>16 Great for the whole family</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>17 Great for adults</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

[BASE: ALL RESPONDENTS]

Q11. Where do you typically get information about travel and vacation packages? Please select all that apply.

[RANDOMIZE]

☐ TV
☐ Radio
☐ Print newspaper
☐ Print magazine
☐ Online – website
☐ Online – social media
☐ Other (Specify) ____________________________
☐ None of the above [EXCLUSIVE]
[BASE: ALL RESPONDENTS]

Q12. Now we’d like you to think about advertising you may have seen or heard for various theme park hotel brands. For each brand, please indicate about how many advertisements you have seen or heard in the past month in the following types of media

[SHOW A GRID WITH ALL BRANDS SELECTED IN S8 ACROSS THE TOP]
[PRE-FILL RESPONSES WITH 0]  

<table>
<thead>
<tr>
<th>BRAND FROM S8</th>
<th>INSERT BRAND FROM S8</th>
</tr>
</thead>
<tbody>
<tr>
<td>TV</td>
<td>1</td>
</tr>
<tr>
<td>Radio</td>
<td>2</td>
</tr>
<tr>
<td>Print newspaper</td>
<td>3</td>
</tr>
<tr>
<td>Print magazine</td>
<td>4</td>
</tr>
<tr>
<td>Online – website</td>
<td>5</td>
</tr>
<tr>
<td>Online – social media</td>
<td>6</td>
</tr>
<tr>
<td>Other (Specify)</td>
<td>7</td>
</tr>
</tbody>
</table>

[BASE: ALL RESPONDENTS]

Q13. Please record the approximate number of hours you spend each day in the following activity. If you want to record a fraction of an hour, please show this with a decimal (e.g., .5 hrs).

| TV       | ___.____1 |
| Radio    | ___.____2 |
| Print newspaper | ___.____3 |
| Print magazine | ___.____4 |
| Online – website | ___.____5 |
| Online – social media | ___.____6 |
| Other (Specify) | ___.____7 |
| None of the above | [EXCLUSIVE] |
Demographics

[BASE: ALL RESPONDENTS]

D1. Please select your gender.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Male</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
</tr>
</tbody>
</table>

[NOTES]

[BASE: ALL RESPONDENTS]

D2. Into which group does your age fall as of your last birthday?

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Under 18</td>
</tr>
<tr>
<td>2</td>
<td>18-24</td>
</tr>
<tr>
<td>3</td>
<td>25-34</td>
</tr>
<tr>
<td>4</td>
<td>35-44</td>
</tr>
<tr>
<td>5</td>
<td>45-54</td>
</tr>
<tr>
<td>6</td>
<td>55-64</td>
</tr>
<tr>
<td>7</td>
<td>65 or older</td>
</tr>
</tbody>
</table>

[NOTES]

[BASE: ALL RESPONDENTS]

D3. Which of the following best describes your marital status? (Please select one response.)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Single, not living with a domestic partner</td>
</tr>
<tr>
<td>2</td>
<td>Single, living with a domestic partner</td>
</tr>
<tr>
<td>3</td>
<td>Married</td>
</tr>
<tr>
<td>4</td>
<td>Divorced</td>
</tr>
<tr>
<td>5</td>
<td>Separated</td>
</tr>
<tr>
<td>6</td>
<td>Widowed</td>
</tr>
<tr>
<td>7</td>
<td>Prefer not to say</td>
</tr>
</tbody>
</table>

[NOTES]

[BASE: ALL RESPONDENTS]

D4. Are there any children under the age of 18 currently living with you, either on a part-time or full-time basis?

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
</tr>
</tbody>
</table>

[BASE: ALL RESPONDENTS]

D5. How many people, including yourself, live in your household?

Enter number of people: _______  [WHOLE NUMBER ONLY. RANGE: 1-20]
### D6. Which of the following best describes your highest level of education? (Please select one response.)

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High school or less</td>
</tr>
<tr>
<td>2</td>
<td>Trade/technical school</td>
</tr>
<tr>
<td>3</td>
<td>Some college or Associate's degree</td>
</tr>
<tr>
<td>4</td>
<td>Graduated college/Bachelor’s degree</td>
</tr>
<tr>
<td>5</td>
<td>Attended graduate school</td>
</tr>
<tr>
<td>6</td>
<td>Advanced degree (Master’s, Ph.D.)</td>
</tr>
</tbody>
</table>

### D7. Which of the following best describes your current employment status? (Please select one response.)

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Employed full-time (30+ hours)</td>
</tr>
<tr>
<td>2</td>
<td>Employed part-time</td>
</tr>
<tr>
<td>3</td>
<td>Not currently employed</td>
</tr>
<tr>
<td>4</td>
<td>Student</td>
</tr>
<tr>
<td>5</td>
<td>Retired</td>
</tr>
<tr>
<td>6</td>
<td>Homemaker</td>
</tr>
</tbody>
</table>

### D8. Which of the following best describes your racial or ethnic background? (Please select one response.)

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Caucasian</td>
</tr>
<tr>
<td>2</td>
<td>African-American</td>
</tr>
<tr>
<td>3</td>
<td>Hispanic or Latino</td>
</tr>
<tr>
<td>4</td>
<td>Asian/Pacific Islander</td>
</tr>
<tr>
<td>5</td>
<td>Other</td>
</tr>
<tr>
<td>6</td>
<td>Prefer not to say</td>
</tr>
</tbody>
</table>

### D9. Which of the following categories best describes your annual household income before taxes? (Please select one response.)

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Under $50,000</td>
</tr>
<tr>
<td>2</td>
<td>$50,000 to just under $75,000</td>
</tr>
<tr>
<td>3</td>
<td>$75,000 to just under $100,000</td>
</tr>
<tr>
<td>4</td>
<td>$100,000 to just under $125,000</td>
</tr>
<tr>
<td>5</td>
<td>$125,000 to just under $150,000</td>
</tr>
<tr>
<td>6</td>
<td>$150,000 to just under $175,000</td>
</tr>
<tr>
<td>7</td>
<td>$175,000 to just under $200,000</td>
</tr>
<tr>
<td>8</td>
<td>$200,000 or more</td>
</tr>
<tr>
<td>9</td>
<td>Prefer not to say</td>
</tr>
<tr>
<td>10</td>
<td>Don’t know</td>
</tr>
</tbody>
</table>
[BASE: ALL RESPONDENTS]

D10. In what state do you reside?

Enter state: _______ [DROPDOWN BOX]

Closing

Thank you for participating in our survey!
Appendix B: Review of Linear Algebra

Linear Algebra

Multivariate analysis refers to a branch of mathematics involving the analysis of two or more variables. This section is intended to be a brief introduction to matrix algebra, eigenvalues and the maximization of quadratic forms. These topics are used extensively in the analysis of marketing data to understand relationships among variables.

Matrix Notation

A matrix is a two-dimensional array of numbers. These arrays form a convenient “pigeon-hole” system for storing and organizing numbers. For example, suppose we want to regress a dependent variable $y$ on two independent variables, $x_1$ and $x_2$:

<table>
<thead>
<tr>
<th>Obs</th>
<th>$y$</th>
<th>$x_1$</th>
<th>$x_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>

We can call this matrix $A$. $A$ is a $4 \times 3$ matrix, with 4 rows and 3 columns. In general, matrices can have an arbitrary number of rows and columns. The $m \times n$ matrix $A$ can be written:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \text{ or } [a_{ij}]$$
where “i” is the row index and “j” is the column index. Vectors are just special types of matrices. A row vector is a $1 \times m$ matrix, e.g., $[1 \ 2 \ 6]$. A column vector is an $n \times 1$ matrix, e.g., $\begin{bmatrix} 1 \\ 2 \\ 14 \\ 4 \end{bmatrix}$. By convention, all vectors are column vectors:

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

**Other Special Matrices:**

An identity matrix is a special type of square matrix (i.e., equal number of rows and columns):

$$I = \begin{bmatrix} 1 & 0 \\ \vdots & \ddots \\ 0 & \cdots & 1 \end{bmatrix}$$

A scalar matrix:

$$\begin{bmatrix} \sigma & 0 \\ & \ddots \\ 0 & \cdots & \sigma \end{bmatrix}$$

A diagonal matrix:

$$\begin{bmatrix} \lambda_1 & 0 \\ \vdots & \ddots \\ 0 & \cdots & \lambda_n \end{bmatrix}$$

A symmetric matrix is $a_{ij} = a_{ji}$. An example is a correlation matrix:

$$\begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{bmatrix}$$
Matrix Operations

A transpose operation rotates a matrix about the main diagonal:

\[ A' = [a_{ji}] \quad \text{e.g.,} \quad A = \begin{bmatrix} 1 & 2 \\ 3 & 2 \\ 1 & 5 \end{bmatrix} \quad A' = \begin{bmatrix} 1 & 3 & 1 \\ 2 & 2 & 5 \end{bmatrix} \]

The transpose of a column vector is a row vector. Matrix addition is done component-wise, e.g.:

\[ \begin{bmatrix} 1 & 2 \\ 1 & 4 \end{bmatrix} + \begin{bmatrix} 2 & 6 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 8 \\ 4 & 5 \end{bmatrix} \]

If A is an \( m \times n \) matrix and B is a \( p \times q \) matrix, then A+B is only defined if \( m = p \) and \( n = q \).

Matrix multiplication is defined for two matrices A and B as the inner product of the ith row of A and the jth column of B. If A is \( m \times n \) and B is \( n \times p \), then the product AB is of dimension \( m \times p \). In other words, the columns of A must be equal to the rows of B for matrix multiplication to be defined:

\[
AB = \begin{bmatrix}
- & a'_1 & - \\
. & . & . \\
- & a'_m & - \\
\end{bmatrix}
\begin{bmatrix}
| & | & | \\
b_1 & \cdots & b_p \\
| & | & | \\
\end{bmatrix}
= \begin{bmatrix}
a'_1b_1 & \cdots & a'_1b_p \\
. & . & . \\
a'_mb_1 & \cdots & a'_mb_p \\
\end{bmatrix}
\]

For example:

\[
A = \begin{bmatrix} 1 & 2 & 4 \\ 3 & 1 & 2 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 1 \\ 6 & 2 \\ 1 & 2 \end{bmatrix}
\]

\[
AB = \begin{bmatrix}
1 \times 1 + 2 \times 6 + 4 \times 1 & 1 \times 1 + 2 \times 2 + 4 \times 2 \\
3 \times 1 + 1 \times 6 + 2 \times 1 & 3 \times 1 + 2 \times 1 + 2 \times 2 \\
\end{bmatrix}
= \begin{bmatrix} 17 & 13 \\ 11 & 9 \end{bmatrix}
\]

\( A' A \) is always defined because the dimensions will always match. The identity matrix is defined so that \( IA = A \).
The advantages of matrices become clear when you define addition and multiplication. Recall that in the simple regression model we manipulated quantities like $\bar{y}$, $\sum x_i^2$ and $\sum x_i y_i$. These are simple examples of matrix multiplication. If $a$ and $b$ are column vectors, we define the inner product of $a$ and $b$ as:

$$a'b = \sum_{i=1}^{n} a_i b_i$$

Thus,

$$\bar{y} = 1'y/n \quad \sum_{i=1}^{n} x_i^2 = x'x \quad \sum_{i=1}^{n} x_i y_i = x'y$$

where “1” is a vector of ones, $1' = [1, \cdots, 1]$.

**Matrix Inverse**

If $A$ is an $n \times n$ matrix, $A^{-1}$ (if it exists) is a matrix constructed so that:

$$A^{-1}A = I = AA^{-1}$$

For example:

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \quad A^{-1} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/4 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} \quad A^{-1} = \begin{bmatrix} 1 & -1/2 \\ 0 & 1/2 \end{bmatrix}$$

This inverse exists if the columns of the matrix are linearly independent. If:

$$X = \begin{bmatrix} x_1 & \cdots & x_p \end{bmatrix}$$

then an inverse exists if there are no constants, $c_1, \cdots, c_p$ such that $c_1 x_1 + \cdots + c_p x_p = 0$. In other words, no one variable (if $X$ is a regression data matrix) can be written as a linear combination of the other variables.
The Covariance Matrix and the Variance of Linear Combinations

Consider the random vector $x$:

$$
x = \begin{bmatrix}
  x_1 \\
  x_2 \\
  \vdots \\
  x_n
\end{bmatrix}
$$

The distribution of a random vector is often characterized by the mean of each of the components and all possible covariances between the elements of the vector:

$$
E[x] = \mu = \begin{bmatrix}
  E[x_1] \\
  E[x_2] \\
  \vdots \\
  E[x_n]
\end{bmatrix}
$$

Covariances are stored in the variance-covariance matrix:

$$
\Sigma = [\sigma_{ij} = \text{cov}(x_i, x_j) = E[(x_i - \mu_i) (x_j - \mu_j)]]
$$

The variance-covariance matrix can be expressed in matrix form as follows:

$$
\Sigma = E[(x - \mu)(x - \mu)']
$$

For example, if:

$$
x = \begin{bmatrix}
  x_1 \\
  x_2
\end{bmatrix}
$$

Then:

$$
E[(x - \mu)(x - \mu)'] = E\left[\begin{bmatrix}
  x_1 - \mu_1 \\
  x_2 - \mu_2
\end{bmatrix}\begin{bmatrix}
  x_1 - \mu_1 & x_2 - \mu_2 \\
  x_2 - \mu_2 & x_1 - \mu_1
\end{bmatrix}\right]
$$

$$
= E\left[\begin{bmatrix}
  (x_1 - \mu_1)(x_1 - \mu_1) & (x_1 - \mu_1)(x_2 - \mu_2) \\
  (x_2 - \mu_2)(x_1 - \mu_1) & (x_2 - \mu_2)(x_2 - \mu_2)
\end{bmatrix}\right]
$$

$$
= \begin{bmatrix}
  \sigma_{11} & \sigma_{12} \\
  \sigma_{21} & \sigma_{22}
\end{bmatrix}
$$

The procedure generalizes to an arbitrary number of elements. If $x$ is composed of independent and identically distributed (iid) elements, then $\Sigma = \sigma^2 I$ is a scalar matrix.
In the general case, $\Sigma$ is a symmetric matrix. The diagonal elements of the variance-covariance matrix are variances: $\sigma_{ii} = \text{cov}(x_i, x_i) = \sigma_i^2$. In particular:

$$E[X + Y] = E[X] + E[Y] \quad X, Y \text{ random vectors or matrices}$$

$$E[AXB] = AE[X]B \quad X \text{ random and A,B fixed}$$

$$E[AXB + C] = AE[X]B + C \quad X \text{ random and A,B,C fixed}$$

If $x$ is a random vector with mean $\mu$ and covariance matrix $\Sigma$, then:

$$E[Ax + b] = AE[x] + b = A\mu + b$$

and

$$\text{Var}[Ax + b] = E[(Ax + b - A\mu - b)(Ax + b - A\mu - b)']$$

$$= E[A(x - \mu)(x - \mu)'A']$$

$$= A\Sigma A'$$

**Geometric Interpretation of Vectors**

For two vectors $x$ and $y$ of dimension $n$, each can be represented in “$n$ space” originating from the origin, with vector addition carried out on an element-by-element basis. For $x' = [1, 2]$ and $y' = [2, 1]$, then $z' = x' + y' = [3, 3]$

For two vectors $x$ and $y$, the inner product is defined as:

$$x'y = \begin{bmatrix} x_1, \cdots, x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i$$

The squared length of a vector in “$n$ space” is equal to $x'x$ by Pythagorean’s theorem. We will write $|x| = (x'x)^{1/2}$ for the length of $x$.

Two vectors are orthogonal ($\perp$) if and only if $x'y = 0$.

**Proof:**

$$x \perp y \Leftrightarrow |x + y| = |x - y|$$

$$\Leftrightarrow x'x + 2x'y + y'y = x'x - 2x'y + y'y$$

$$\Leftrightarrow 2x'y = 0$$
Consider the vectors $y$ and $x$ where the vector is orthogonal to $x$. We can solve for $c$:

$$
\begin{align*}
(y - cx)'x &= 0 \\
y'x - cx'x &= 0 \\
c &= \frac{(y'x)}{(x'x)}
\end{align*}
$$

And therefore $y_1 = cx = ((y'x)/(x'x)) x$ is the projection of $y$ onto $x$. The length of $cx$ is:

$$
|cx| = \left[ \frac{(y'x)^2}{(x'x)^2} x'x \right]^{1/2} = \frac{(y'x)}{|x|}
$$

Define $\theta$ as the angle between $y$ and $x$. Then:

$$
\cos \theta = \frac{|y_1|}{|y|} = \frac{(y'x)}{|x||y|}
$$
This expression can be interpreted as the correlation between $y$ and $x$. Let $y_d$ and $x_d$ be vectors of mean-centered data: $y_d = y - (1/m)1' y$ and $x_d = x - (1/m)1' x$. Then:

$$\text{Corr}(y, x) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\left(\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2\right)^{1/2}} = \frac{x_d'y_d'}{|x_d||y_d|} = \cos \theta$$

**Matrices as a Linear Transformation**

Consider the equation $y = Xb$ for an arbitrary matrix $X$ of dimension $n \times m$. The vector $b$, of dimension $m$, is being transformed to the matrix $y$ of dimension $n$ via the matrix $X$. The vector $y$ is a weighted combination of the columns of $X$, with the weights supplied by $b$:

$$y = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} b_1 x_1 + b_2 x_2 + \cdots + b_m x_m \end{bmatrix}$$

The matrix $X$ is said to "span" the "vector space" generated by its columns, and can be thought of as providing a "map" from an "$m$ dimensional space" to an "$n$ dimensional space," i.e., $\mathbb{R}^m \rightarrow \mathbb{R}^n$. Therefore, if $y = Xb$, then $y$ must lie in the column space of $X$.

**Determinants**

The determinant of a square matrix $A$ is defined as:

$$\det A = |A| = \sum \text{sgn}(i_1 i_2 \cdots i_n) a_{i_1i_2} \cdots a_{i_n}$$

This is the sum over all possible permutations of the number $(1, 2, \cdots, n)$. A particular permutation is denoted as $(i_1, i_2, \cdots, i_n)$. A permutation is odd (negative sign) if an odd number of interchanges are required to sort the indices into ascending order. For $A$ of dimension $2 \times 2$:

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}$$
Properties of determinants:

i) \( \det cA = c^n \det A \).

ii) \( A \) is not of full rank if and only if \( |A| = 0 \).

iii) \( |AB| = |BA| = |B| |A| \) with the second equality holding if \( A \) and \( B \) are square.

Geometrically, the determinant of a matrix is the “area” of the column space of the matrix, a generalization of the notion of the length of a vector. Consider the identity matrix of dimension 2, where \( |I| = 1 \), which agrees with the standard calculations of the area of a square generated by the two column vectors of \( I \): \( (1,0)^t \) and \( (0,1)^t \). For column vectors that are non-orthogonal, such as \( (1,.5)^t \) and \( (.5,1)^t \), the determinant is \( 1 \times 1 - .5 \times .5 = .75 \), which is the area of a parallelogram generated by the column space.

**Trace**

The trace of a square matrix is the sum of its diagonal elements. A useful property of the trace (tr) is:

\[
tr (AB) = tr (BA)
\]

**Quadratic Forms**

A function that takes vectors and produces a number \( \mathbb{R}^n \rightarrow \mathbb{R}^1 \) of the form \( f(x) = x'Cx \) is called a quadratic form. For example:

\[
f(x) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
= x_1^2 c_{11} + x_1 x_2 c_{12} + x_2 x_1 c_{21} + x_2^2 c_{22}
\]

If \( x'Cx > 0 \) for all \( x \neq 0 \), then \( C \) is called a positive definite matrix. All variance-covariance matrices are positive definite because linear combinations of random variables always have positive variance. Furthermore:
i) \( x'Cx > 0 \) \( C \) is positive definite.

ii) \( x'Cx \geq 0 \) \( C \) is positive semi-definite.

iii) \( x'Cx < 0 \) \( C \) is negative definite.

iv) \( x'Cx \leq 0 \) \( C \) is negative semi-definite.

The Multivariate Normal distribution has a density function involving a quadratic form in the exponent:

\[
p(x|\mu, \Sigma) = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} \exp \left[ -\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right]
\]

where the covariance matrix, \( \Sigma \), is a symmetric positive semi-definite matrix. Symmetry implies that corresponding off-diagonal elements are equal, \( \sigma_{ij} = \sigma_{ji} \). This is reasonable since covariances and correlations are symmetric functions that do not depend on which variable is the first argument of the function and which is second. The positive semi-definite aspect of a covariance matrix implies that all linear combinations of variables also have a zero or positive covariance. That is, for any \( x \) we have:

\[
(x - \mu)' \Sigma^{-1} (x - \mu) \geq 0
\]

**Vector Calculus**

A gradient vector \( \nabla f(x) \) is an \( n \times 1 \) vector of partial derivatives of the function \( f \):

\[
\nabla f = \begin{bmatrix}
\frac{\partial f}{\partial x_1} \\
\vdots \\
\frac{\partial f}{\partial x_n}
\end{bmatrix}
\]

A Hessian matrix is the matrix of second derivatives:

\[
H(x) = \begin{bmatrix}
\frac{\partial^2 f(x)}{\partial x_i \partial x_j}
\end{bmatrix}
\]

Some interesting cases:

i) \( f(x) = a'x \) then \( \nabla f(x) = a \).

ii) \( f(x) = x'Cx \) then \( \nabla f(x) = 2Cx \) for \( C \) symmetric.
Regression Revisited

The standard regression model in long hand is written:

\[ y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_k x_{k,i} + \varepsilon_i \]

and can be rewritten as:

\[ y = X\beta + \varepsilon \]

where:

\[
\begin{bmatrix}
  y_1 \\
  \vdots \\
  y_n 
\end{bmatrix} =
\begin{bmatrix}
  1 & x_{11} & \cdots & x_{1k} \\
  \vdots & \vdots & \ddots & \vdots \\
  1 & x_{n1} & \cdots & x_{nk} 
\end{bmatrix} \begin{bmatrix}
  \beta_0 \\
  \vdots \\
  \beta_k 
\end{bmatrix} +
\begin{bmatrix}
  \varepsilon_1 \\
  \vdots \\
  \varepsilon_n 
\end{bmatrix}
\]

The least squares problem involves choosing the k values of \( \beta \) so as to minimize the residual sum of squares:

\[
\min_{\beta} \varepsilon^T \varepsilon = (y - X\beta)^T (y - X\beta) \\
= y'y - 2\beta'X'y + \beta'X'X\beta
\]

The solution is to take the derivative with respect to \( \beta \), set equal to zero, and solve for \( \beta \):

\[
\frac{\partial (\varepsilon^T \varepsilon)}{\partial \beta} = 0 = 0 - 2X'y + 2X'X\beta
\]

Rearranging we get:

\[ X'X\beta = X'y \]

And the least squares estimator is:

\[ \hat{\beta} = (X'X)^{-1}X'y \]

If we consider \( y \) and the columns of \( X \) as vectors in \( \mathbb{R}^n \), then for any \( \hat{\beta} \), \( \hat{y} = X\hat{\beta} \) is a vector in the column space of \( X \). \( \varepsilon' \varepsilon \) is the squared distance form \( \hat{y} \) to \( y \), and \( \varepsilon \) will be at a right angle to the column space of \( X \) when it is a minimum. The geometric interpretation of regression leads to the following measures of fit. By Pythagorean’s Theorem:

\[
(y'y - \hat{y}'\hat{y}) = y'y - \hat{y}'\hat{y} + \varepsilon'\varepsilon \\
(SST) = (SSR) + (SSE)
\]

One popular measure of goodness of fit is \( R^2 \):

\[ R^2 = \frac{SSR}{SST} \quad \text{or} \quad 1 - \frac{SSE}{SST} \]
Figure 2: Geometry of Regression

\[ \hat{y} = x_1 \beta_1 + x_2 \beta_2 \]

\[ \hat{\beta} \] is that proportion of the variation in \( y \) that can be explained by \( X \). Note that \( 0 \leq R^2 \leq 1 \). \( R^2 \) is closely related to the sample correlation coefficient and is also known as the coefficient of determination:

\[ R^2 = corr(y, \hat{y})^2 \]

\[ = \cos^2(\theta) \text{ for the vectors } y \text{ and } \hat{y} \]

A convenient method of summarizing all of this information is with the “analysis of variance” (ANOVA) table. The analysis of variance breakdown is a direct result of Pythagorean’s Theorem.

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>Mean Squared Error</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression ( \hat{y}' \hat{y} - n\bar{y}^2 )</td>
<td>( k )</td>
<td>MSR</td>
<td>MSR/MSE</td>
</tr>
<tr>
<td>Residual ( \hat{\varepsilon}' \hat{\varepsilon} )</td>
<td>( n-k-1 )</td>
<td>MSE = ( s^2 )</td>
<td></td>
</tr>
<tr>
<td>Total ( y'y - n\bar{y}^2 )</td>
<td>( n-1 )</td>
<td>Var(( y ))</td>
<td></td>
</tr>
</tbody>
</table>
Projections and Idempotent Matrices

In the regression problem \( \hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y \) the matrix \( X(X'X)^{-1}X' = P \) is known as a projection matrix because it projects the vector \( y \) onto the column space of \( X \). The matrix \( P \) has the following properties:

\[ i) \quad PP = P \quad \text{The projection of a projection is the same as the projection.} \]
\[ ii) \quad P' = P \quad \text{Symmetry.} \]
\[ iii) \quad P(I - P) = 0 \quad \text{The column space of } P \text{ is orthogonal to } (I - P). \]

For the regression problem:

\[
y' y = y' Py + y'(I - P)y \\
= y' P'y + y'(I - P)'(I - P)y \\
= (Py)'(Py) + ((I - P)y)'((I - P)y) \\
= \hat{y}'\hat{y} + \hat{\varepsilon}'\hat{\varepsilon}
\]

Multivariate Analysis

Change of Coordinates and Eigenvalue Decomposition

In our analysis of marketing we consider the data as given. That is, we condition on the data and attempt to understand its origin and consequences. The data are what ground our analysis, and our goal is to measure significant associations among variables as well as to propose alternative perspectives on the data that was collected. One aspect of this analysis is to consider whether combinations of the original variables, or subsets of these variables, provide a clearer description of the object of measurement. This task involves what is known as a change of coordinates.

We begin with some definitions. Define \( Q \) as an orthogonal matrix if it is square and:

\[ Q'Q = QQ' = I \quad \text{i.e.,} \quad Q^{-1} = Q' \]

This implies that if the matrix \( Q = [q_1, q_2, \cdots, q_n] \) then where \( q_i'q_j = 1 \) if \( i = j \) and equals zero otherwise. Therefore the columns (or rows) of an orthogonal matrix are what is known as an orthonormal basis for \( \mathbb{R}^n \). Consider, for example, an orthogonal matrix in \( \mathbb{R}^2 \) that corresponds to a rotation of the coordinate system:
APPENDIX B: REVIEW OF LINEAR ALGEBRA

\[ Q = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \]

For an arbitrary vector \( x \) we have:

\[ |x|^2 = x'x = x'Q'Qx = |Qx|^2 \]

implying that \( Q \) is length-preserving. Now, suppose the coordinate system is changed according to \( Q \) with degrees, i.e., a 90-degree rotation of the coordinate system. What are the coordinates of \( x \) in this new coordinate system. Suppose:

\[ x = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad Q = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \]

In the new coordinate system the vector \( x \) will have coordinates:

\[ x_{\text{new}} = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \]

Note that the vector \( x \) has not moved, but the coordinates we use to label it have changed. It is interesting to note that:

\[ x_{\text{new}} = Q^{-1}x_{\text{old}} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \]

In general, if the coordinate system undergoes a rigid rotation according to an orthonormal matrix \( Q \), then new coordinates for a specific point are equal to \( Q^{-1}x = Q'x \) since \( Q^{-1} = Q' \).

Given any real symmetric matrix \( A \) of dimension \( n \) there exists an orthogonal matrix \( Q \) such that:

\[ Q'AQ = L = \begin{bmatrix} \lambda_1 & 0 & \cdots \\ 0 & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \]

or, alternatively, \( A = QLQ' \). The values \( (\lambda_1, \cdots, \lambda_n) \) are the “eigenvalues” of \( A \). Each is associated with an “eigenvector,” or column of \( Q \). Eigenvectors and eigenvalues have the following interesting properties:

1. Consider the transformation of the vector \( x \): \( Ax \). From the above discussion we see that \( Ax = QLQ'x \) where \( Q \) is an orthonormal matrix and \( L \) is diagonal. The transformation \( Ax \) can therefore be decomposed into the following three steps:
MULTIVARIATE ANALYSIS

- Rotate the coordinate system of x: $Q'x$
- Stretch the new coordinate system: $LQ'x$
- Rotate back to the old coordinates: $QLQ'x$

2. If $A$ is a positive definite matrix, all the eigenvalues of $A$ are positive.

3. The eigenvalues of $A$, $[q_1, \cdots, q_n]$, satisfy the equation $Aq_k = \lambda_k q_k$. In other words, the transformation of $q_k$ by $A$ does not change the direction of $q_k$, but just expands or contracts it.

4. The eigenvalues of the square matrix $A$ are defined as the roots of the equation $|A - \lambda I| = 0$.

5. If $\lambda_i \neq \lambda_j$, the corresponding eigenvectors are orthogonal (i.e., $q_i'q_j = 0$). To see this, suppose $Aq_i = \lambda_i q_i$, and $Aq_j = \lambda_j q_j$. Then $q_i' (Aq_i) = \lambda_i q_i' q_i$ and $q_i' (Aq_j) = \lambda_j q_i' q_j$, so for both to be true we must have $q_i' q_j = 0$.

6. For $A$ positive definite symmetric, there exists a matrix $B$ such that $A = BB$ where $B = QL^{1/2}Q'$. $B$ is known as a square root matrix, of which there are many.

**Cauchy-Schwarz Inequality**

The Cauchy-Schwarz Inequality for any two vectors $b$ and $d$ is:

$$(b'd)^2 \leq (b'b) (d'd)$$

with equality if and only if $b = cd$ for some scalar $c$. The proof involves the vector $(b - cd)$ with positive length. Then:

$$0 < (b - cd)' (b - cd) = b'b - 2cb'd + c^2d'd$$
$$= b'b - \frac{(b'd)^2}{d'd} + \left( \frac{(b'd)^2}{d'd} - 2cb'd + c^2d'd \right)$$
$$= b'b - \frac{(b'd)^2}{d'd} + \left( d'd \left( c - \frac{(b'd)}{(d'd)} \right)^2 \right)$$

For the inequality to hold for any $c$ it must be the case that:

$$0 < b'b - \frac{(b'd)^2}{d'd}$$

Or, rearranging, we have the Cauchy-Schwarz Inequality.
The Extended Cauchy-Schwarz Inequality

Let $b$ and $d$ be any two vectors, and let $B$ be a positive definite matrix. Then:

$$(b'd)^2 \leq (b'Bb) (d'B^{-1}d)$$

with equality holding if and only if $d = cBb$ for some arbitrary scalar $c$. The proof makes use of the square root matrix $B^{1/2}$ and its inverse. We have:

$$(b'd)^2 = (b'Id)^2 = (b'B^{1/2}B^{-1/2}d)^2 = \left( (B^{1/2}b)' (B^{-1/2}d) \right)^2$$

and the proof is complete by applying the Cauchy-Schwarz inequality to the vectors $(B^{1/2}b)$ and $(B^{-1/2}d)$.

The Maximization Lemma

This lemma states that for a positive definite matrix $B$ and a vector $d$, then for an arbitrary vector $x$ we have:

$$\max_{x \neq 0} \frac{(x'd)^2}{x'Bx} = d'B^{-1}d \text{ with maximum for } x = cB^{-1}d$$

Proof:

By the Extended Cauchy-Schwarz Inequality, dividing both sides by $x'Bx$ yields the desired result.

Maximization of Quadratic Forms

Let $A$ be a positive definite matrix with eigenvalues $\lambda_1, \cdots, \lambda_n$ and associated eigenvectors $q_1, \cdots, q_n$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Then:

$$\max_{x \neq 0} \frac{x'Ax}{x'x} = \lambda_1 \text{ attained when } x = q_1$$

Proof:

$$\frac{x'Ax}{x'x} = \frac{x'QLQ'x}{x'QQ'x} = \frac{y'Ly}{y'y} = \frac{\sum_{i=1}^{n} \lambda_i y_i^2}{\sum_{i=1}^{n} y_i^2} \leq \lambda_1 \frac{\sum_{i=1}^{n} y_i^2}{\sum_{i=1}^{n} y_i^2} = \lambda_1$$

Which is attained when $y' = [1, 0, \cdots, 0]$, occurring when $x = q_1$. 
Principal Component Analysis

The multivariate analysis technique of principal components seeks to find a low-dimensional representation of variables under study by explaining their correlational structure with a small number of linear combinations of the original variables. Let \( z' = (z_1, z_2, \cdots, z_n) \) be distributed Multivariate Normal with mean \( \mu \) and covariance matrix \( \Sigma \). A special case of the covariance matrix when the variances are all equal to one is the correlation matrix, i.e., \( \Sigma = R \). The covariance matrix can be decomposed into the product of three matrices such that:

\[
\Sigma = QLQ'
\]

with eigenvalues \( \lambda_1, \lambda_2, \cdots, \lambda_n \) and associated eigenvectors \( Q = [q_1, q_2, \cdots, q_n] \).

Consider the linear combinations of the variable \( z \):

\[
y_1 = \ell_1'z = \ell_{11}z_1 + \ell_{21}z_2 + \cdots + \ell_{n1}z_n \\
y_2 = \ell_2'z = \ell_{12}z_1 + \ell_{22}z_2 + \cdots + \ell_{n2}z_n \\
\vdots \\
y_n = \ell_n'z = \ell_{1n}z_1 + \ell_{2n}z_2 + \cdots + \ell_{nn}z_n
\]

Then:

\[
Var (y_i) = \ell_i'\Sigma \ell_i \quad \text{and} \quad Cov (y_i, y_j) = \ell_i \Sigma \ell_j
\]

The principal components are those uncorrelated linear combinations \( y_1, y_2, \cdots, y_n \) whose variances are as large as possible. The first principal component is the linear combination \( \ell_1'x \) that maximizes \( Var (\ell_1'z) \) subject to \( \ell_1'\ell_1 = 1 \). The second principal component is the linear combination \( \ell_2'x \) that maximizes \( Var (\ell_2'x) \) subject to \( \ell_2'\ell_2 = 1 \), etc. The solution to the problem is \( (\ell_1, \ell_2, \cdots, \ell_n) = (q_1, q_2, \cdots, q_n) \), i.e., the eigenvectors, as shown in the Maximization Lemma.

Proof:

\[
\frac{Var (\ell'x)}{|\ell\ell'|} = \frac{\ell'\Sigma \ell}{\ell'\ell} = \frac{\ell'QLQ'\ell}{\ell'QQ'\ell} = \frac{\kappa'\ell \ell \kappa}{\kappa'\kappa} = \frac{\sum_{i=1}^{n} \lambda_i k_i^2}{\sum_{i=1}^{n} k_i^2} \leq \lambda_1
\]

where \( k_i \) are the elements of \( Q \).
and equality is obtained when $k' = (1, 0, \cdots, 0)$, or when $l = q_1$. Similarly, the other combinations of $z$ that satisfy the properties of principal components can be shown to be the other eigenvectors of $R$.

In general, we are interested in reducing the dimension of the original multivariate variables and yet continue to account for the majority of variance. A useful relationship for determining the amount of explained variance is:

$$tr(\Sigma) = tr(QLQ') = tr(Q'QL) = tr(L) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$$

Therefore the proportion of the total population variance due the $k^{th}$ principal component is:

$$\frac{\lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_n}$$

Furthermore, we have the following low-dimensional approximation to the covariance matrix:

$$\Sigma = \lambda_1 q_1 q_1' + \lambda_2 q_2 q_2' + \cdots + \lambda_r q_r q_r' + \cdots + \lambda_n q_n q_n'$$

$$\hat{\Sigma} = \lambda_1 q_1 q_1' + \lambda_2 q_2 q_2' + \cdots + \lambda_r q_r q_r'$$

A pictorial representation of the original data can be obtained with the component scores and component loadings. Component scores are defined as coefficients that describe the original variable, $x$, in terms of the linear composite variable $y$:

$$z_1 = f_{11} y_1 + f_{12} y_2 + \cdots + f_{1r} y_r$$
$$z_2 = f_{21} y_1 + f_{22} y_2 + \cdots + f_{2r} y_r$$
$$\vdots$$
$$z_n = f_{n1} y_1 + f_{n2} y_2 + \cdots + f_{nr} y_r$$

The elements of $\{f_{ij}\}$ can be used to “plot” the original variables in an orthogonal space since $Cov(y_i, y_j) = 0$. Recall that $y = Q'x$ and therefore $x = Qy$ since $Q'Q = QQ' = I$. Therefore the $y$ variables provide a coordinate system with which to plot the original variables, with the matrix of eigenvectors serving as the actual coordinate values. Elements of the vector $z$ that plot close together are highly similar (i.e., correlated).
It is usually desirable to scale each column of $Q$ (the eigenvectors) by the eigenvalues so that distances in the resulting plots are proportional to the amount of explained variance. These scaled eigenvectors are often referred to as factor loadings and the eigenvalues as factor scores. The scaled eigenvectors are related to the original eigenvectors by the transformation:

$$Q^* = QL^{1/2} \quad \text{or} \quad q_i^* = \sqrt{\lambda_i} q_i$$

**Discriminant Analysis**

Discriminant analysis is a multivariate technique that attempts to find a linear combination of variables that maximally discriminates among populations in a sample. The technique assumes we know the population indicator for $\pi_i$ for observations $i$ in a dataset. For example, we may know the brands a person ($i$) uses, and wish to understand how a set of variables measuring their needs best discriminates brand usage. We assume the population variables ($z$) come from one of the populations, which differ in their means but have a common covariance:

$$E[z|\pi_i] = \mu_i$$
$$Cov[z|\pi_i] = \Sigma \quad \text{for all } i$$

We seek to find a vector of discriminating coefficients $\ell$ that keeps the populations as distinct as possible. That is, for $z \in \pi_i$ we consider a linear combination of the original variables where $y = \ell'z$ with:

$$E[y] = \ell' \mu_i = \mu_{iy}$$
$$Cov[y] = \ell' \Sigma \ell = \sigma_i^2$$

that maximizes the variance among populations relative to the variance within populations. For $g$ groups this is expressed as:
\[
\frac{\text{Variance Among}}{\text{Variance Within}} = \frac{\sum_{i=1}^{g} (\mu_{iy} - \bar{\mu}_y)^2}{\sigma_y^2} = \frac{\sum_{i=1}^{g} \ell' (\mu_i - \bar{\mu}) (\mu_i - \bar{\mu})' \ell}{\sigma_\ell^2} = \frac{\ell' \left[ \sum_{i=1}^{g} (\mu_i - \bar{\mu}) (\mu_i - \bar{\mu})' \right] \ell}{\sigma_\ell^2} = \frac{\ell' B \ell}{\ell' \Sigma \ell}
\]

Which is maximized for \( \ell \) equal to the first eigenvector of \( \Sigma^{-1/2} B \).

**Proof:**

Let \( a = \Sigma^{1/2} \ell \). Then:

\[
\frac{\ell' B \ell}{\ell' \Sigma \ell} = \frac{a' \Sigma^{-1/2} B \Sigma^{-1/2} a}{a' a}
\]

By the Maximization Lemma, this expression is maximized for \( a \) equal to the first eigenvector of \( \Sigma^{-1/2} B \Sigma^{-1/2} \), or:

\[
\Sigma^{-1/2} B \Sigma^{-1/2} a = \lambda a
\]

Therefore:

\[
\Sigma^{-1/2} \Sigma^{-1/2} B (\Sigma^{-1/2} a) = \lambda (\Sigma^{-1/2} a) \\
\Sigma^{-1} B (\Sigma^{-1/2} a) = \lambda (\Sigma^{-1/2} a) = \lambda \ell \quad \text{since} \quad \ell = \Sigma^{-1/2} a
\]

In other words, \( \ell \) is the first eigenvector of \( \Sigma^{-1} B \).
Canonical Correlation Analysis

The goal of canonical correlation analysis is to find a linear combination of two sets of variables that has maximal correlation. In marketing, one set of variables may be brand beliefs, and another set of variables might be media consumed. If both are collected using a Pick any/J format, then the model discussed in Chapter 1 on Market Definition could be used to estimate a latent set of variables $z$ that are jointly Multivariate Normal:

$$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \sim \text{Normal} \left( \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$$

The goal is to find a set of weights $\alpha$ and $\beta$ such that the following quantity is maximized:

$$\max_{\alpha, \beta} | \text{corr} (\alpha' z_1, \beta' z_2) | = \max_{\alpha, \beta} \frac{ | \alpha' \Sigma_{12} \beta | }{ \sqrt{ \alpha' \Sigma_{11} \alpha \beta' \Sigma_{22} \beta } }$$

To obtain the solution to this problem we first transform to simplify the denominator:

$$a = \Sigma_1^{1/2} \alpha \quad \text{and} \quad b = \Sigma_2^{1/2} \beta$$

so:

$$\frac{ | \alpha' \Sigma_{12} \beta | }{ \sqrt{ \alpha' \Sigma_{11} \alpha \beta' \Sigma_{22} \beta } } = \frac{ | a' \Sigma_1^{-1/2} \Sigma_{12} \Sigma_2^{-1/2} b | }{ \sqrt{ a' a \beta' \beta } }$$

and we recognize this transformed problem as similar to the maximization of a quadratic form. The next step in the solution is to obtain the singular value decomposition (SVD) of the expression:

$$\Sigma_1^{-1/2} \Sigma_{12} \Sigma_2^{-1/2} = Q \begin{pmatrix} L & 0 \end{pmatrix} P'$$

An SVD is a general form of an eigenvalue decomposition where the matrix need not be square. $Q$ and $P$ are orthogonal, and $L$ contains the eigenvalues (singular values). We recognize that:

$$\left| a' \Sigma_1^{-1/2} \Sigma_{12} \Sigma_2^{-1/2} b \right| = \left| a' Q \begin{pmatrix} L & 0 \end{pmatrix} P' b \right| = \left| c' \begin{pmatrix} L & 0 \end{pmatrix} d \right|$$

Which is maximized if $a = q_1$ and $b = p_1$ since $Q$ and $P$ have orthogonal eigenvectors. Therefore the maximum correlations are the singular values of $\Sigma_1^{-1/2} \Sigma_{12} \Sigma_2^{-1/2}$ and the weights are related to the eigenvectors of $Q$ and $P$: 
\[ \alpha_i = \Sigma_{11}^{-1/2} q_i \quad \text{and} \quad \beta_i = \Sigma_{22}^{-1/2} p_i \]
Appendix C: Review of Statistics

Introduction

The goal of analysis in marketing is to make statements about the likely values of things that are not observed. These things can be model parameters, such as coefficients in a regression model, hypotheses, latent factors and aspects of models that can be expressed as a function of model parameters. We are sometimes interested in predicting the effect of multiple actions, such as a price cut and an accompanying advertisement, or the source of volume of a new product introduction. These more complicated scenarios involve functions of model coefficients.

At the core of statistical modeling and statistical inference is the idea that statistical probability distributions can be used to characterize unknown aspects of models. An example is the error term in a regression model, which is usually assumed to follow a Normal distribution. Error terms are introduced in models to allow the observed data to vary away from a deterministic specification where the exact value of an output, or dependent variable, is predicted. To illustrate, consider again the standard regression model:

\[ y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \ldots + \beta_k x_{k,i} + \varepsilon_i \]

The term \( \varepsilon_i \) is the model error term, and allows the predicted model values (\( \hat{y} \)) to differ from the observed data:

\[ \varepsilon_i = y_i - (\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \ldots + \beta_k x_{k,i}) \]

\[ = y_i - \hat{y}_i \]

A common assumption in regression analysis is that the error terms follow an i.i.d. Normal distribution, denoted as \( \varepsilon_i \sim \text{i.i.d. } N(0, \sigma^2) \) where the symbol “~” means “distributed as” and the expression “i.i.d.” means “independent and identically distributed.”
That is, the errors all come from the same distribution and knowledge of the value of any one error tells us nothing about the value of the others, given that we know the value of $\sigma^2$. Figure 3 provides examples of a regression model with one covariate where $\beta_0 = 1$ and $\beta_1 = 2$. The top left corresponds to data generated from a model where $\sigma^2 = 0.01$, and the top right corresponds to $\sigma^2 = 0.25$. The graphs at the bottom are examples where the error terms are not identically distributed (left side), and where the error terms are not independently distributed.

Figure 3: Data generated from regression models with different error assumptions.

The Normal distribution plays a central role in modern statistical analysis due to the Central Limit Theorem, which states that the mean of a set of random variables tends to a Normal distribution regardless of the distribution of the individual items in the set. The Normal distributions, and all other distributions in statistics, can be thought of as a statement of probable values of something that is not directly observed by the analyst.

We begin our review of statistics by discussing various statistical distributions used in the analysis in this book. We then introduce Bayes theorem and describe why Markov chain Monte Carlo (MCMC) methods, the computational arm of modern Bayesian meth-
ods, is particularly powerful when applied to hierarchical models. We then provide details of model estimation for the models described in this book.

Statistical Distributions

Normal Distribution

The Normal distribution is often used in the analysis of marketing data. The Normal distribution is characterized by two parameters, the mean ($\mu$) and variance ($\sigma^2$). The expression for the probability density function is:

$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{2\sigma^2} (x - \mu)^2 \right]$$

The above expression is for a bell-shaped curve, centered at $\mu$ and with a scale that is related to the square root of the variance, ($\sqrt{\sigma^2}$), also known as the standard deviation ($\sigma$). The Figure below plots the value of the random variable $x$ when the mean is zero and the variance is one. The distribution for $\mu \neq 0$ is centered at $\mu$ rather than at zero, and $\sigma^2 \neq 1$ results in the scale being expanded ($\sigma^2 > 1$) or contracted ($\sigma^2 < 1$) by a factor of $\sigma$. If $\sigma^2 = 0.25$, then the scale on the bottom of the graph would range from (-2,2) rather than the current values of (-4,4).

The multivariate Normal distribution generalizes the univariate Normal distribution to multiple random variables. The multivariate Normal introduces covariance parameters that describe the co-movement, or correlation, of the random variables. The general form of the multivariate Normal distribution for a vector of random variables $x$ of dimension $k$ is:

$$p(x|\mu, \Sigma) = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} \exp \left[ -\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right]$$

For $k = 2$, we have $x' = (x_1, x_2)$, $\mu' = (\mu_1, \mu_2)$ and $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$. The term $\sigma_{12}$ is the covariance between $x_1$ and $x_2$, and is related to the standard deviations of these variables through the expression $\sigma_{12} = \rho_{12}\sigma_1\sigma_2$, where $\rho_{12}$ is the correlation coefficient. Expanding the above expression for $k = 2$ we get:
Figure 4: Density of the standard Normal distribution. The probability of $x$ taking on a value within a specific range of values is computed as the area under the curve.

\[
p(x = 1, \sigma^2 = 1)
\]

\[
-4 -2 0 2 4
\]

\[
x
\]

\[
p(x | \mu_1, \mu_2, \rho, \sigma^2_1, \sigma^2_2)
\]

\[
= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left[ -\frac{1}{2(1-\rho^2)} \left( \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} - 2\rho (x_1 - \mu_1)(x_2 - \mu_2) \sigma_1\sigma_2 \right) \right]
\]

A plot of the contours of the bivariate Normal distribution for $\mu = 0$, $\sigma_1^2 = \sigma_2^2 = 1$ and $\rho = 0.5$ is provided in Figure 5. The probability contours indicate the likely value of random data generated from the distribution, with 99 percent of the data expected to lie within the outer-most concentric ellipse, and 10 percent of the data expected to be concentrated within the inner-most ellipse. The positive correlation coefficient leads to a clock-wise tilt to the data, with positive values of $x_1$ associated with positive values
of $x_2$. A negative correlation would correspond to a counter-clockwise tilt of the data - positive values of $x_1$ would tend to be associated with negative values of $x_2$. As with the univariate Normal distribution, the mean parameter $\mu$ translates the location of the probability density contours, and the diagonal elements of the covariance matrix, $\text{diag}(\Sigma) = (\sigma_1^2, \sigma_2^2)$, rescale the axes.

Figure 5: Density of the bivariate Normal distribution with positive correlation. The contours indicate the probable location of the data mass indicated by each label, with 10% of the data expected to be located within the smallest contour.

A useful property of the multivariate Normal distribution is that its marginal and conditional distributions are also distributed Normal. That is, if the vector $x$ is distributed multivariate Normal, then each element of the vector, $x_i$, is distributed Normal with mean $\mu_i$ and variance $\sigma_i^2$. In addition, all conditional distributions are also distributed Normal:

$$p(x_1|x_2) = \text{Normal} \left( \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \right)$$

where $\Sigma_{12}$ denotes the covariance matrix of the elements of $x_1$ and $x_2$. The above expression implies that if the elements of the vector $x$ are distributed multivariate Normal, then the conditional mean resembles a regression model with $\beta' = \Sigma_{12}\Sigma_{22}^{-1}$. 
Extreme Value (Gumble) Distribution

A statistical distribution encountered often in models of choice data is the Extreme Value distribution, also referred to as the Gumble distribution. Technically, the Extreme Value distribution is used to describe the extreme value of a set of observations, i.e., the minimum and the maximum. It’s original use was in describing extreme behavior over time, such as the maximum height of a river level for the purpose of regional planning, or for the distribution of world-records in a track event. Its application to choice modeling was to stabilize the probability of alternative selection as copies of a choice were introduced. Thus, choice between three alternatives \{coffee, tea, milk\} would be unaffected if the choice set became \{coffee, coffee, tea, tea, milk, milk\} or \{coffee, coffee, coffee, tea, tea, milk, milk, milk\} – i.e., the probability of some coffee being selected from the expanded choice set is the same as coffee being selected from the original choice set with just three elements. We discuss how the Extreme Value distribution is used to model choice in Chapter 4.

The probability density function for the standardized Extreme Value distribution with location set to zero and scale set to one is:

\[ p(x) = \exp[-x] \exp[-e^{-x}] \]

The cumulative density function of the extreme value distribution is

\[ F(x) = \exp[-e^{-x}] \]

and the distribution allows for the introduction of a location (\(\mu\)) and scale (\(\sigma\)) parameter that serves to translate the location of the distribution and rescale the axis: \(F(x|\mu,\sigma) = \exp[-e^{-(x-\mu)/\sigma}]\). The distribution is graphed below, and is seen to have a mode at zero, an abrupt tail to the left, and a tail that slowly decreases on the right. As with all distributions, the probability of an event is computed by calculating the area under the curve within specified regions. The simple expression for the cumulative density function makes the Extreme Value distribution useful for describing these events.

The multivariate generalization of the Extreme Value distribution is known as the Generalized Extreme Value (GEV) distribution. The GEV takes the form:

\[ F(x_1, \ldots, x_k) = \exp \left[ -\sum_{s=1}^{S} \left( \sum_{j \in A_s} \exp[-x_j/\lambda_s] \right)^{\lambda_s} \right] \]

where \(A = A_1 \cup A_2 \cup \cdots \cup A_S = \{1, 2, \ldots, k\}\) represents the partitioning of the \(k\) elements of the vector \(x\) into \(S\) partitions \(A_1, A_2, \cdots, A_S\), and \(\lambda\) a measure of the relative independence of the alternatives within each set. If \(\lambda_s = \lambda = 1\), then the cumulative density function above is equal to the product of standard Extreme Value cumulative density functions. Since the arguments within each partition are exchangeable, the distribution
has an equi-correlated structure within group, and an equi-correlated structure across groups.

**Chi-Square and Inverted Chi-Square Distributions**

The last set of distributions we will use are related to quadratic forms present in the exponent of the Normal distribution. We start with the Chi-Square distribution, which is characterized by a single parameter known as its degree-of-freedom. A Chi-Square random variable with one degree of freedom can be generated from a standard Normal random variable, \( z \sim \text{Normal} (0, 1) \), that is then squared. That is, \( x_1 = z^2 \) is distributed Chi-Square with one degree of freedom for \( z \) standard Normal. The Chi-Square distribution has a recursive property in that a Chi-Square with two degrees of freedom is obtained by simulating two independent realizations of \( x_1 \) and then adding them together. More generally, a Chi-Square with \( k \) degrees of freedom can be simulated by squaring \( k \) independent Normals and summing them together:
\[ x_k = \sum_{i=1}^{k} z_i^2 \sim \text{Chi-Square (k)} \text{ or } \chi_k^2 \]

The probability density function of a Chi-Square distribution with \( k \) degrees of freedom is:

\[ p(x|k) = \frac{1}{2^{k/2} \Gamma(k/2)} x^{k/2 - 1} e^{-x/2} \]

The graph of a Chi-Square distribution with five degrees of freedom is provided below.

Figure 7: Density of a Chi-Square Distribution with 5 d.f.

A related distribution that is directly applicable to analysis with the Normal distribution is the Inverse Chi-Square distribution. The Inverse Chi-Square distribution describes the likely values of the inverse of a variable, \( x \), distributed Chi-Square:

\[ p(x|k) = \frac{1}{2^{k/2} \Gamma(k/2)} x^{-(k/2+1)} e^{-1/(2x)} \]

The Inverse Chi-Square distribution is also defined to be positive. It’s popularity is due to its use in modeling the variance term in the Normal distribution, which appears
in the denominator of the exponential term in the density for the Normal distribution. A graph of the Inverse Chi-Square distribution with five degrees of freedom is provided below:

Figure 8: Density of an Inverse Chi-Square Distribution with 5 d.f.

Scaled versions of the Chi-Square and Inverted Chi-Square distributions that allow for different ranges of the outcome are described by the Gamma and Inverse Gamma distributions. The scale parameters of these and other distributions are best thought of as simply changing the values along the horizontal axis of the figures above, while leaving the shape of the distribution unchanged. The probability density function of the Gamma distribution is:

\[ p(x|m, s) = \frac{s^m}{\Gamma(m)} x^{m-1} \exp(-sx) \]

And the probability density function of the Inverted Gamma distribution is:

\[ p(x|m, s) = \frac{s^m}{\Gamma(m)} x^{-(m+1)} \exp\left(-\frac{s}{x}\right) \]
The Chi-Square distribution is a special case of the Gamma distribution with $k = m/2$ and $s = 1/2$. Similarly, the Inverse Chi-Square distribution is the same as an Inverse Gamma distribution with $k = m/2$ and $s = 1/2$. The popularity of these distributions is due to their use in modeling the variance of the Normal distribution:

$$p(\sigma^2) \propto (\sigma^2)^{-(\nu_0/2 + 1)} \exp\left(-\frac{\nu_0 s_0^2}{2\sigma^2}\right)$$

which is an Inverse Gamma distribution with $m = \nu_0/2$ and $s = \nu_0 s_0^2/2$.

**Wishart and Inverted Wishart Distributions**

The multivariate extension of the Chi-Square and Inverse Chi-Square distributions are known as the Wishart and the Inverse Wishart distributions. The Wishart distribution with one degree of freedom can be simulated from multivariate Normal random variables as follows:

$$S = xx' \sim \text{Wishart}(V, 1) \text{ for } x \sim \text{Normal}(0, V)$$

And, as with the Chi-Square distribution, the summation of Wishart distributed random variables yields a Wishart with degrees of freedom that sum:

$$S(V, m) + S(V, k) = S(V, m + k)$$

The Wishart distribution is used in the statistical analysis of positive definite random matrices, such as the covariance matrix. A related distribution is the Inverse Wishart, defined as:

$$\text{if } S \sim \text{IW}(V, k) \text{ then } S^{-1} \sim \text{W}(V^{-1}, k)$$

The Inverse Wishart distribution is used in the Bayesian analysis of multivariate Normal data as the prior distribution for the covariance matrix.

**Bayesian Analysis**

Statistical models of behavior combine an algebraic structure of regular, perfectly predictable behavior with a part that is not perfectly predictable, or random. The regular part is described in terms of variables observed by the analyst, and the random part contains variables that are not observed by the analyst. In a regression model, the regression equation $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$ is regular, and the error term $\varepsilon \sim \text{Normal}(0, \sigma^2)$.
is random. The two are combined to yield a dependent variable $y$ that is a mixture of
the regular and random components. The goal of statistical analysis is to begin with
the observed data $(y, x)$ and make inferences about unobserved variables $\beta$ and $\sigma^2$ given
the model. The “model” here is the assumption of an additive linear structure for the
regression model, the set of covariates $\{x_i, i = 1, \ldots, k\}$, and the assumption that the
errors are distributed i.i.d. Normal with mean zero.

Bayesian analysis adheres to a principle that says that all information about the model
parameters contained in the data is expressed through the likelihood function. The
likelihood function is a mathematical statement of the data-generating process, which
provides a prescription of how to simulate data for a model. Specific examples are
provided below. The likelihood also provides the means for calculating the probability of
hypothetical (i.e., not yet observed) data given specific parameter values. Of course, in
any particular application, the parameter values are unknown and the goal is to use the
likelihood in reverse, by informing us of likely values of parameters given the observed
data. This reversal is accomplished by Bayes Theorem, which states:

\[ p(\theta|Data) = \frac{p(Data|\theta) p(\theta)}{p(Data)} \]

or:

\[ p(\theta|Data) \propto p(Data|\theta) p(\theta) \]

or:

\[ \text{Posterior} \propto \text{Likelihood} \times \text{Prior} \]

where “\(\propto\)” reads “is proportional to.” The goal of statistical inference is to make state-
ments about unobserved quantities, $\theta$, given the observed data. For the regression model
$\theta = (\{\beta_i\}, \sigma^2)$ and “Data” refers to the dependent variable $y$, but not the explanatory
varibles $\{x_i\}$ that are considered independently generated and are taken as given in the
analysis. Inference in a Bayesian analysis is conducted through the posterior distribution
by specifying a model in terms of the likelihood function and combining it with prior
information about the model parameters.

**Beta-Bernoulli Model**

Consider a simple example involving data that are distributed Bernoulli. A Bernoulli
distribution describes an outcome variable that takes on two values: zero and one. The
distribution is indexed by one parameter, $\theta$, that is restricted to be greater than zero and
smaller or equal to one. The probability that an outcome is equal to the value one is $\theta$
and the probability that the outcome is equal to zero is $1 - \theta$. This distribution could
be applied to a baseball player attempting to get a hit in a baseball game, where $\theta$ could be considered the player’s batting average.

$$p(y_i = 1) = \theta \quad p(y_i = 0) = 1 - \theta$$

Suppose the data comprise $N$ outcomes from a Bernoulli distribution. If we assume that the outcomes are independent from each other given $\theta$, meaning that knowledge of the outcome of the first observation provides no information about the other outcomes, then the likelihood is expressed as:

$$p (\text{Data}|\theta) = \prod_{i=1}^{N} \theta^{y_i} (1 - \theta)^{1-y_i}$$

$$= \theta^n (1 - \theta)^{N-n}$$

where $n = \sum_{i=1}^{N} y_i$ is the number of outcomes with $y_i = 1$ in the $N$ observations.

Bayes theorem states that the posterior distribution is proportional to the likelihood times the prior, so a prior distribution for $\theta$ is needed for a Bayesian analysis to proceed. A convenient distribution is the Beta distribution which has its support restricted to the interval between zero and one:

$$p (\theta) \propto \theta^{\alpha-1} (1 - \theta)^{\beta-1}$$

The Beta distribution is indexed by two parameters, $\alpha$ and $\beta$, which are specified by the analyst. While the Beta distribution is in the same form as the Bernoulli likelihood, it is important to remember that $\theta$ is the unknown variable in the Bernoulli likelihood, it is in the likelihood the data $\{y_i\}$ are unknown with $\theta$ given, or conditioned upon. The mean of the Beta distribution is $\alpha/ (\alpha + \beta)$, and larger values of these parameters lead to a tighter, or more informative distribution of $\theta$. The figure below provides a plot of the Beta distribution for different values of $\alpha$ and $\beta$.

Bayes theorem combines the prior and the likelihood to obtain the posterior distribution:
The posterior distribution for the Beta-Bernoulli model is also a Beta distribution with prior parameters $\alpha$ and $\beta$ modified by the data $n$ and $N$. As the sample size of the data increases, $n$ and $N$ become large relative to the prior parameters, and the posterior mean converges to the sample mean $n/N$. In small samples the prior plays a larger role in determining the posterior mean, which is equal to $(n + \alpha)/(N + \alpha + \beta)$.

Modern Bayesian methods facilitate the analysis of the posterior distribution through a simulation procedure known as Markov chain Monte Carlo (MCMC) methods. The advantage of a simulation-based method of estimation is that it can more easily deal with complicated model structures than the Beta-Bernoulli model, where the posterior distribution is derived analytically. In many marketing applications, the posterior distribution
does not have a known form, and its properties, such as the posterior mean, can only be obtained through simulation.

MCMC is a method for simulating draws from the posterior distribution by creating a Markov chain with a stationary distribution equal to the posterior distribution of a model. An MCMC analysis of the Beta-Bernoulli model would involve generating a Monte Carlo draw of $\alpha$ given $\beta$ and the data, and then a separate draw of $\beta$ given $\alpha$ and the data. We discuss the algorithms for specific models below.

The generation of draws from the full conditional distribution of model parameters is greatly simplified for models written in a hierarchical form. Modern Bayesian methods are often referred to as being “hierarchical Bayes” because it is now so common to employ Bayesian analysis to hierarchical models. Before describing the specific models used in the analysis in this book, it is important to understand why hierarchical models are so amenable to Bayesian analysis.

Hierarchical Bayes Models

Hierarchical Bayes models comprise two distinct parts - i.e., a model that is i) written in a hierarchical form and ii) estimated with Bayes theorem. An example of a hierarchical model is the multivariate probit model for Pick any/J data. The observed data ($x$) are binary but related to a continuous latent variable ($z$) that is not directly observed. The observed variable is thus a censored realization of the latent variable, and the model for the joint data can be written using two equations:

\[
x | z \\
z | \mu, \Sigma
\]

The first equation describes the censoring mechanism, and the model is such that if the latent variable is positive then the censored variable is equal to one, and if the latent variable is zero or negative then the latent variable is equal to zero. Thus the observed data follow a Bernoulli distribution with probability equal to the latent variable being positive:

\[
x = 1 \quad \text{if} \quad z > 0 \\
x = 0 \quad \text{if} \quad z \leq 0
\]

The second equation describes the relationship between the latent variable ($z$) and other model parameters, $\mu$ and $\Sigma$. A simple model for $z$ assumes:
and a slightly more complicated model would be to replace the mean, $\mu$, with a regression model $\mu = a'\beta$.

These equations constitute a hierarchical model because they employ the concept of conditional independence. The hierarchy for this model has censored data ($x$) informing the model parameters ($\mu$, $\Sigma$) through the latent variable ($z$). Given the latent variable ($z$), no additional information is conveyed about the model parameters ($\mu$, $\Sigma$) from the data ($x$). The variables $x$ and ($\mu$, $\Sigma$) are independent of each other given $z$.

Another example of a hierarchical model is a regression model with heterogeneity:

$$
\begin{align*}
\epsilon_{it} &\sim \text{Normal}(0, \sigma_i^2) & t = 1, \ldots, T \\
\beta_i &\sim \bar{\beta} + \zeta_i & i = 1, \ldots, N \\
y_{it} &\sim \text{Normal}(0, \sigma^2_t) \\
\end{align*}
$$

The hierarchy in this model has the dependent variable ($y_{it}$) informing the parameters ($\bar{\beta}$, $\sigma^2_t$, $\zeta_i$, $\sigma^2_i$, $V_{\beta}$) through the individual-level coefficients $\beta_i$. If the set of individual-level coefficients $\{\beta_i\}$ were known, then the parameters of the second equations are not informed any further by $y_{it}$ and $x_{it}$.

A general form for writing hierarchical models is with a bracketed notation, where $[a|b]$ reads the distribution of a given $b$. For the hierarchical linear model described above, the model hierarchy is:

$$
[y_{it}|x_t, \beta_i, \sigma^2_t] \\
[\beta_i|\bar{\beta}, V_{\beta}] \\
$$

Hierarchical models are particularly well suited for estimation by modern Bayesian methods because of the property of conditional independence. As discussed below, Markov chain Monte Carlo (MCMC) methods work by generating draws from the full conditional distribution of model parameters. Consider the following hierarchical model for the distribution of data and model parameters:

$$
[a, b, c, d, \ldots, x, y] = [a|b] [b|c] [c|d] \cdots [x|y]
$$

where the variable “$y$” represents the data. The conditional distribution of the parameters “$c$” given all other parameters is:
The reason for the simplification is that factors such as \([a|b]\) are constant with respect to the integration in the denominator and cancel with the corresponding factor in the numerator. The only factors that do not cancel in the numerator are those that contain the variable \(c\). Moreover, the denominator has integrated out \(c\), i.e., is constant with respect to \(c\), and we can therefore replace the equality with the proportionality symbol. It is important to keep in mind which object is a variable and which is a constant in the above example. For the full conditional distribution of \(c\) given other variables, \(c\) is a variable but the other elements of the expression are constants.

MCMC estimation works well with hierarchical models because the hierarchical property of conditional independence greatly simplifies the full conditional distributions, often to the point of having a known form (e.g., a Normal distribution) so that complicated numerical methods are not needed to generate the draws needed for estimation. For example, in the hierarchical Bayes linear regression model discussed earlier, the conditional distribution of \(\beta\) given all other parameters is a simple linear model. We discuss these and other distributions below.

**MCMC Estimation**

Markov chain Monte Carlo estimation is a simulation-based procedure that employs a Markov chain whose long-run or stationary distribution is the posterior distribution of the model under study. Markov chains are an auto-regressive stochastic process of order one, meaning that the next value of the process depends only on the immediately previous value and not earlier values. The Monte Carlo aspect of the estimation procedure involves the simulation of draws from specific distributions. The MCMC estimator results in draws of all model parameters.

A simple version of the MCMC estimator is known as the Gibbs sampler. The Gibbs sampler is an iterative estimation procedure that simulates draws from the full conditional distribution of all model parameters. Thus, if the model for the data is of the form:

\[
[a, b, c, d, \ldots, x, y] = [a|b] \cdot [b|c] \cdot [c|d] \cdot \\
\]
where, typically, the symbol $a$ reflects observed data and the parameter $y$ denotes a specified parameter of a prior distribution, neither of which needs to be simulated. The Gibbs sampler proceeds by generating random draws from the following distributions:

1. $[b|a, c, \cdots] \propto [a|b] [b|c]$
2. $[c|a, b, \cdots] \propto [b|c] [c|d]$
3. $[d|a, b, \cdots] \propto [c|d] [d|e]$

Repeating this sequence of steps results in Monte Carlo draws from the posterior distribution of the model. The arguments to the right of the vertical bar are the most recent draws of each of the parameters, and as a result the recursion is a Markov chain.

**Metropolis-Hastings Algorithm**

It is often the case that it is either difficult or impossible to simulate directly from the posterior full conditional distributions in steps 1-3 above. This occurs when the prior distribution and the likelihood do not conform to each other, as in discrete choice models where the likelihood is comprised of discrete mass points and the prior is a density. In these cases, the Metropolis-Hastings (MH) algorithm can be used to simulate draws. The simplest form of the MH algorithm uses a random-walk to generate candidate draws which are accepted with probability $\alpha$. If the candidate draw is rejected, then the value of the parameter is not updated and instead retains its current value and no updating occurs. The MH algorithm works by setting $\alpha$ so that the acceptance probability of a new draw makes the Markov chain “time reversible” with respect to the posterior distribution of a model, so that the stationary distribution of the Markov chain is also the posterior distribution. This allows us to use the MH algorithm as a device for simulating from the posterior. The random-walk MH chain proceeds as follows:

1. Generate a candidate value of a parameter $\theta_i^{new}$ using the old value plus a symmetric disturbance: $\theta_i^{new} = \theta_i^{old} + N(0, v^2)$ where $v^2$ is specified by the analyst so that 30-50% of the candidates are accepted.

2. Compute the acceptance probability $\alpha = \min \left\{ 1, \frac{p(\theta_i^{new}|\text{Data})}{p(\theta_i^{old}|\text{Data})} \right\}$

3. Accept the new draw of $\theta_i$ with probability $\alpha$: draw a Uniform(0,1) random variable and if $U < \alpha$ accept the draw of $\theta_i$. Otherwise, retain the old value of $\theta_i$ and proceed to the next draw in the recursion.
To understand why the MH algorithm works it is first necessary to describe a Markov chain more formally and then to establish two facts about them with regard to their stationary long-run distributions and the property of time reversibility. A Markov chain is a stochastic process that describes the evolution of random variables by specifying transition probabilities of moving from one realization to the next. The simplest Markov chain contains just two states, or values, that a variable can assume and has a matrix of transition probabilities:

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$$

where $p_{ij}$ is the probability of moving from state $i$ to state $j$, and the sum of probabilities in each row sum to one, e.g., $p_{1,1} + p_{1,2} = 1.0$. The transition probability $p_{ii}$ is the probability of staying in state $i$. If the probability of being in each of the two states is initially $\pi_0 = (0.7, 0.3)$, then the state probabilities after one iteration of the Markov chain is:

$$\pi_1 = \pi_0 P = \begin{bmatrix} 0.7 & 0.3 \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = [0.7p_{11} + 0.3p_{21} \ 0.7p_{12} + 0.3p_{22}]$$

The transition matrix $P$ is therefore the key component of the Markov chain as it describes how the state probabilities change over time. If

$$P = \begin{bmatrix} 0.50 & 0.50 \\ 0.25 & 0.75 \end{bmatrix}$$

then

$$\pi_1 = \pi_0 P = \begin{bmatrix} 0.7 & 0.3 \end{bmatrix} \begin{bmatrix} 0.50 & 0.50 \\ 0.25 & 0.75 \end{bmatrix} = [0.425 \ 0.575]$$

and we see that the probability of being in second state increases from 0.3 to 0.575. As this chain continues to iterate, the state probabilities will converge to long-run or steady-state probabilities:
\[ \pi_1 = \pi_0 P, \quad \pi_2 = \pi_1 P = \pi_0 PP = \pi_0 P^2, \quad \pi_r = \pi_0 P^r \]

and the effects of the starting distribution \( \pi_0 \) should wear off. The chain will converge to what is known as the stationary distribution, \( \pi \), defined such that:

\[ \pi = \pi P \]

For the transition matrix \( P \) defined above, it can be verified that the long-run stationary distribution is \( \pi = \left[ \frac{1}{3}, \frac{2}{3} \right] \), which is obtained regardless of the initial probabilities \( \pi_0 \).

The goal of the MH algorithm is to construct a Markov chain with stationary distribution equal to the posterior distribution of a specific model. This is accomplished by making the chain time-reversible with respect to the posterior. A time reversible chain is one where the probabilities of moving from state \( i \) to state \( j \) is the same as moving from state \( j \) to state \( i \). At any point in the chain, the probability of seeing an \( i \to j \) transition is \( \pi_i p_{ij} \) and so a chain is time reversible if

\[ \pi_i p_{ij} = \pi_j p_{ji} \]

Furthermore, since the row probabilities in the transition matrix \( P \) sum to one, we have:

\[ \sum_i \pi_i p_{ij} = \sum_i \pi_j p_{ji} = \pi_j \sum_i p_{ji} = \pi_j \]

or

\[ \pi P = \pi \]

In other words, \( \pi \) is the stationary distribution.

The property of time reversibility provides us with an alternative to the complicated task of searching for the transition matrix \( P \) with the stationary distribution we desire, i.e., the posterior distribution of our model. Instead of a direct search for \( P \), we can use the property of time reversibility to modify an arbitrary chain with transition matrix \( Q \) so that it produces the stationary distribution we desire. This is accomplished by modifying the transition probabilities of an arbitrary “candidate-generating” distribution \( q_{ij} \) such that:

\[ p_{ij} = q_{ij} \alpha(i, j) \]

where

\[ \alpha = \min \left\{ 1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}} \right\} \]

That is, a candidate state value is generated according to the transition matrix \( Q \) and accepted with probability \( \alpha \). With probability \( 1 - \alpha \) the candidate value is rejected.
and the old value is retained. This algorithm results in a Markov chain with stationary distribution \( \pi \).

**Proof**

We prove this statement by showing that an \( i \rightarrow j \) transition is equal to a \( j \rightarrow i \) transition with respect to \( \pi \) regardless of the candidate-generating distribution \( Q \).

\[
\pi_i p_{ij} = \pi_j q_{ij} \min \left\{ 1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}} \right\} = \min \{ \pi_i q_{ij}, \pi_j q_{ji} \}
\]

\[
\pi_j p_{ji} = \pi_j q_{ji} \min \left\{ 1, \frac{\pi_i q_{ij}}{\pi_j q_{ji}} \right\} = \min \{ \pi_j q_{ji}, \pi_i q_{ij} \}
\]

The right sides of the above expressions are the same, and therefore \( \pi_i p_{ij} = \pi_j p_{ji} \) and the resulting Markov chain has stationary distribution \( \pi \). If we select \( \pi \) as the posterior distribution of our model and we regard the “states” of the stochastic process as the possible values that our model parameters can assume, then the resulting Markov chain will simulate draws from the posterior distribution \( \pi \). All that is needed is to be able to evaluate the posterior distribution up to the constant of proportionality that cancels from the numerator and denominator of the above expression:

\[
\pi_i \propto p(\text{Data}|\theta_i)p(\theta_i)
\]

The candidate-generating probabilities \( q_{ij} \) and \( q_{ji} \) cancel in the expression for \( \alpha \) for the random-walk MH chain that employs a symmetrical distribution (i.e., the Normal distribution) to generate the candidates. If \( \theta_j = \theta_i + \varepsilon \) with \( \varepsilon \) symmetrical, then the resulting transition probabilities are such that \( q_{ij} = q_{ji} \). Other variants of the MH algorithm generate candidate values of \( \theta \) in other ways, and can result in faster convergence and better mixing properties of the Markov chain. These versions do not result in \( q_{ij} = q_{ji} \) and lead to different values of \( \alpha \). The Gibbs sampler can be shown to be a special case of the MH algorithm with \( \alpha = 1 \). However, regardless of which variation of the MH algorithm that is employed, the result is a general method of employing a Markov chain to simulate draws from the posterior distribution of model parameters.

**Prior Distributions**

Throughout our analysis, we assume proper but diffuse prior distributions \( p(\theta) \) for model parameters:
• Regression coefficients: $p(\beta) = \text{Normal}(\bar{\beta}, A^{-1})$ with mean $\bar{\beta} = 0$ and information matrix $A = .01I$

• Regression error variance: $p(\sigma^2) = \text{Inverted Gamma}(\nu, \text{ssq})$ with $\nu = 3$ and $\text{ssq} = \text{var}(y)$

• Covariance matrix: $p(\Sigma) = \text{Inverted Wishart}(\nu_0, V_0)$ with $\nu_0 = \text{dim}(\Sigma) + 3$ and $V_0 = .01I$

Regression Model

Univariate Regression

The univariate regression model has one dependent variable ($y$) and multiple independent variables $\{x_j\}$:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_k x_{k,i} + \epsilon_i$$

where “i” indexes the observations. The model can be expressed more concisely using matrix algebra with the model for one observation expressed as:

$$y_i = x_i^T \beta + \epsilon_i, \quad \epsilon_i \sim \text{iid } N(0, \sigma^2)$$

Stacking the $n$ observations in the data results in a vector of responses:

$$y \sim N(X\beta, \sigma^2 I_n)$$

with $N(\mu, \Sigma)$ being used as notation for a multivariate Normal distribution with mean $\mu$ and covariance matrix $\Sigma$. The regression model specifies the dependent variable $y_i$ as being normally distributed with a mean that depends on the covariates $x_i$, $\beta$ and a common variance $\sigma^2$ across observations, but with no additional covariance among the $y_i$ values after accounting for the effects of $x_i$. $X$ is a matrix with $i$th row equal to $x_i$. The likelihood for the data is:

$$p(y|X, \beta, \sigma^2) \propto (\sigma^2)^{-n/2} \exp \left[ -\frac{1}{2\sigma^2} (y - X\beta)' (y - X\beta) \right]$$

Bayesian analysis of the univariate regression model requires a prior distribution for $\beta$ and $\sigma^2$. A convenient prior for $\sigma^2$ is:

$$p(\sigma^2) \propto (\sigma^2)^{-((\nu_0)/2) + 1} \exp \left( -\frac{\nu_0 s_0^2}{2\sigma^2} \right)$$
which is an Inverse Gamma distribution with $m = \nu_0/2$ and $s = \nu_0 s_0^2/2$. Also for convenience, we chose a conditional prior for $\beta$, i.e., $p(\beta|\sigma^2)$ so that $p(\beta, \sigma^2) = p(\beta|\sigma^2)p(\sigma^2)$:

$$p(\beta|\sigma^2) \propto (\sigma^2)^{-k} \exp\left[-\frac{1}{2\sigma^2}(\beta - \bar{\beta})'A(\beta - \bar{\beta})\right]$$

By Bayes theorem, we know that the posterior distribution is proportional to the likelihood multiplied by the prior, or:

$$p(\beta, \sigma^2|y, X) \propto p(y|X, \beta, \sigma^2) p(\beta|\sigma^2) p(\sigma^2)$$

$$\propto (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right]$$

$$\times (\sigma^2)^{-k} \exp\left[-\frac{1}{2\sigma^2}(\beta - \bar{\beta})'A(\beta - \bar{\beta})\right]$$

$$\times (\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{\nu_0 s_0^2}{2\sigma^2}\right)$$

and after some algebraic manipulation of these equations we can obtain the following expression for the posterior distribution:

$$p(\beta, \sigma^2|y, X) \propto (\sigma^2)^{-k/2} \exp\left[-\frac{1}{2\sigma^2}(\beta - \bar{\beta})' (X'X + A)(\beta - \bar{\beta})\right]$$

$$\times (\sigma^2)^{-(n+\nu_0)/2} \exp\left(-\frac{\nu_0 s_0^2 + ns^2}{2\sigma^2}\right)$$

with $\bar{\beta} = (X'X + A)^{-1}(X'y + A\bar{\beta})$ and $ns^2 = (y - X\bar{\beta})'(y - X\bar{\beta}) + (\bar{\beta} - \beta)'A(\bar{\beta} - \beta)$. The posterior distribution is therefore expressed as a conditional distribution of $\beta$ given $\sigma^2$:

$$p(\beta|y, X, \sigma^2) = N(\bar{\beta}, \sigma^2(X'X + A)^{-1})$$

and a marginal posterior for $\sigma^2$:

$$p(\sigma^2|y, X) = \text{Inverse Gamma}(m = (\nu_0 + n)/2, s = (\nu_0 s_0^2 + ns^2)/2)$$

Because of the relationship between the Chi-Square, Inverse Chi-Square and Gamma distributions, an equivalent expression for the posterior distribution of $\sigma^2$ is:

$$p(\sigma^2|y, X) \propto \frac{\nu_0 s_0^2 + ns^2}{\chi^2_{\nu_0+n}}$$
where $\chi^2_k$ denotes a Chi-Square distribution with $k$ degrees of freedom.

The results of the above algebra indicate two things. The first is an iterative, simulation-based estimation strategy that involves the following steps:

1. Draw $\sigma^2$ from an Inverse Gamma or Inverse Chi-Square distribution with argument $ns^2 = f(\tilde{\beta}, y, X)$ calculated from the data and prior.

2. Draw $\beta$ given $\sigma^2$ from a multivariate Normal distribution.

3. Repeat.

Secondly, the algebra needed to derive the posterior distribution for even a relatively simple model such as this one is somewhat difficult to follow. The advantage of Markov chain Monte Carlo (MCMC) estimation is that it often does not require math more complicated than that above. This is because the joint distribution of model parameters is simulated recursively through a series of conditional draws, and the hierarchical structure of a model simplifies the expressions for the conditional posterior distributions as above. This will become more evident when discussing models with heterogeneity.

The univariate regression model is implemented in the program “runireg” in the bayesm package in the R statistical language.

### Multivariate Regression

The multivariate regression model is used to relate a vector of $m$ outcome variables ($y_i$) collected across $n$ observations. This model is used extensively in hierarchical models to describe the variation of regression coefficients across respondents as discussed further below. Each observation vector $y_i$ is a row of the matrix $Y$ that is related to a common set of covariates $X$:

$$Y = XB + U, \quad U = [u_i'] , \quad u_i \sim N(0, \Sigma)$$

where $Y$ is of dimension $n \times m$, $X$ is $n \times k$, $B$ is $k \times m$, with each column of $B$ containing the regression coefficients for one of the $m$ elements of $y$.

We assume an Inverted Wishart prior for the error covariance matrix:

$$p(\Sigma) = \text{Inverted Wishart}(\nu_0, V_0)$$

and a multivariate generalization of the conditional Normal prior for the regression coefficients:

$$\beta|\Sigma = \text{vec}(B) | \Sigma \sim N\left(\text{vec} \left( \tilde{B} \right), \Sigma \otimes A^{-1}\right)$$
where $\text{vec}(B)$ is a function that stacks the columns of the matrix $B$ to form the vector $\beta$, and $\otimes$ is the Kronecker (direct) product where if $A = (a_{ij})$ is a $m_1 \times n_1$ matrix and $B$ is a $m_2 \times n_2$ matrix then $A \otimes B$ is of order $m_1 m_2 \times n_1 n_2$ defined as:

$$A \otimes B = \begin{bmatrix}
a_{11}B & a_{12}B & \cdots & a_{1n_1}B \\
a_{21}B & a_{22}B & \cdots & a_{2n_1}B \\
\vdots & \vdots & \ddots & \vdots \\
a_{m_11}B & a_{m_12}B & \cdots & a_{m_1n_1}B
\end{bmatrix}$$

Following Bayes theorem, we have:

$$p(\Sigma, B) \propto p(Y | X, B, \Sigma) \ p(B | \Sigma) \ p(\Sigma)$$

and it can be shown (Rossi, Allenby and McCulloch 2005) that the posterior distribution is of the form:

$$p(\Sigma | Y, X) = \text{Inverted Wishart}(v_0 + n, V_0 + S)$$
$$p(\beta | Y, X, \Sigma) = N(\tilde{\beta}, \Sigma \otimes (X'X + A)^{-1})$$
$$\tilde{\beta} = \text{vec}(\tilde{B}), \quad \tilde{B} = (X'X + A)^{-1}(X'Y + AB)$$
$$S = (Y - X\tilde{B})'(Y - X\tilde{B}) + (\tilde{B} - \bar{B})'A(\tilde{B} - \bar{B})$$

Thus, the multivariate regression model is an extension of the univariate regression model. The posterior distribution is derived from Bayes theorem, and its functional form obtained by multiplying the data likelihood by the prior distribution. The algebraic expression above provides an explicit expression for simulating draws from the posterior distribution.

The Gibbs sampler for the multivariate regression model is implemented in the program “rmultireg” in the bayesm package in the R statistical language.

**Pick any/J Model**

Pick any/J data are a vector of outcomes with each element taking on one of two values: 1 and 0. The data are therefore distributed multivariate binomial, with probabilities that are possibly correlated. We model the data using a hierarchical model:

$$x | z$$
$$z | \mu, R$$
where the first layer of the hierarchy is censoring mechanism that describes how a latent, unobserved continuous variable $z$ gives rise to the observed data $x$:

$$\text{if } x_j = 1 \text{ then } z_j > 0 \text{ and if } x_j = 0 \text{ then } z_j \leq 0$$

The second layer of the hierarchy describes the latent variable $z$ that is assumed to be distributed multivariate Normal:

$$z \sim \text{Normal} \left( \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_J \end{bmatrix}, R = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1J} \\ r_{21} & 1 & \cdots & r_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ r_{J1} & r_{J2} & \cdots & 1 \end{bmatrix} \right)$$

The correlation matrix $R$ appears in our model specification instead of the covariance matrix $\Sigma$ because the covariances are not statistically identified. That is, we can multiply the left and right side of the above inequalities by an arbitrary constant $k$ and obtain the same response probabilities with $k\mu$ and $kR$ as parameters in the multivariate Normal distribution. We therefore need to set the diagonal elements of $\Sigma$ to one so that we can uniquely estimate the mean $\mu$. Larger values of $\mu$ indicate that a greater proportion of $x$ values are observed to be equal to 1, and greater correlations $r_{i,j}$ imply that the co-occurrence of 1’s is greater than that explained by the mean vector $\mu$. However, despite the identification problems that only allow estimation of the correlation matrix $R$, we will estimate the model with the full covariance matrix $(\beta, \Sigma)$ and post-process the MCMC draws to obtain estimates of the identified parameters $(\mu, R)$. This strategy is viable because the posterior distribution of model parameters is always proper as long as the prior distribution is proper. So, even if the likelihood is only informative about a subset of the parameters (i.e., $R$ instead of $\Sigma$), Bayesian analysis of the model can proceed.

Our estimation strategy involves the sequential generation of draws from the following conditional distributions:

1. $z_h|x_h, \beta, \Sigma \sim \text{Truncated Normal}(\beta, \Sigma)$, $h=1, \cdots, H$ respondents and where the region of truncation for each of the J variables is indicated by $x_{hj}$: $z_{hj} > 0$ for $x_{hj} = 1$ and $z_{hj} \leq 0$ for $x_{hj} = 0$.

2. $\beta|\Sigma, \{z_h\} \sim \text{multivariate Normal}(\tilde{\beta}, V)$

$$V = (X'X + A)^{-1}, \quad \tilde{\beta} = V(X'z^* + A\beta)$$

$$\Sigma^{-1} = C'C, \quad X_i = C'I, \quad z_i^* = C'z_i$$
3. $\Sigma|\beta, z \sim \text{Inverted Wishart}(\nu_0 + H, V_0 + S)$ where $S = \sum_h (z_h - \beta_h)(z_h - \beta_h)'$.

We then transform to the identified parameters $(\beta, \Sigma) \rightarrow (\mu, R)$ through the transformation that converts a covariance matrix to a correlation matrix.

$$\mu = \Lambda \beta, \quad R = \Lambda \Sigma \Lambda$$

and

$$\Lambda = \begin{bmatrix} 1/\sqrt{\sigma_{11}} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 1/\sqrt{\sigma_{JJ}} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\\end{bmatrix}$$

This transformation is applied at each iteration.

The Pick any/J model is implemented in the program “rmvpGibbs” in the bayesm package in the R statistical language.

**Cutpoint Model**

The cutpoint model assumes that data collected on a fixed-point rating scale is a censored realization of data that is distributed multivariate Normal. Furthermore, it assumes that respondents vary in their tendency to use different parts of the scale, with some respondents yea-sayers who use the upper part of the scale, and others nay-sayers who use the lower end of the scale. Variation is also assumed to exist in the range of responses that a respondent tends to provide, with some respondents using a wide range of the scale and some using a narrow range. These aspects of response style, coupled with the discreteness of the observed data, are modeled with a two-equation system:

$$x_{i,j} = k \text{ if } c_{k-1} \leq z_{i,j} \leq c_k$$

and

$$z_i \sim \text{Normal} \left( \mu + \tau_i \epsilon, \sigma_i^2 \Sigma \right)$$

where “$i$” denotes the respondent, “$j$” indexes the question, “$k$” is the response and $\epsilon$ is a vector of ones. The parameter $\tau_i$ translates the Normal distribution to reflect yea- and nay-sayers, while the parameter $\sigma_i^2$ expands and contracts the range of responses depending on whether it takes on a value greater or less than one. Once estimates of
τ_1 and σ_1^2 are obtained, a respondent’s latent response z_i can be transformed to make it compatible, or exchangeable, with responses from other respondents:

\[ z_i^* = \frac{z_i - τ_1}{σ_1} \sim \text{Normal}(μ, Σ) \]

The above transformation is similar to directly standardizing the observed data by subtracting the mean and dividing by the standard deviation:

\[ z_i = \frac{x_i - \bar{x}_i}{s_i} \]

except that this simple transformation does not account for the discreteness of the data. The discreteness of the observed data makes it extremely unlikely that responses themselves are Normally distributed. Our model for the data can be expressed as:

\[ x_i | z_i, c \]
\[ z_i | τ_i, σ_i^2, μ, Σ \]

and a Bayesian analysis of this likelihood proceeds by specifying a prior distribution for all the model parameters. We will employ what is known as a hierarchical prior to deal with the fact that the parameters τ_i and σ_i^2 are respondent-specific. Since σ_i^2 must be positive because it reflects a scale adjustment to the covariance matrix Σ, which cannot have negative diagonal elements, we specify the prior in terms of the logarithm of σ_i:

\[ \begin{bmatrix} τ_i \\ \ln σ_i \end{bmatrix} \sim \text{Normal}(ϕ, Λ) \]

and then place an additional prior on the parameters ϕ and Λ. We will revisit the concept of hierarchical priors in our discussion of heterogeneity below. We employ Inverted Wishart priors on the covariance matrices Σ and Λ, and place additional restrictions on ϕ so that the mean of the translation parameter τ_i is zero, and the mode of the prior on σ_i to be equal to one. This is accomplished by the following setting:

\[ ϕ_1 = 0 \quad ϕ_2 = λ_{22} \]

Estimation proceeds by recursively generating draws from the following conditional distributions.

1. \[ z_i | x_i, c, τ_i, σ_i, μ, Σ \sim \text{Truncated Normal}(μ + τ_i, σ_i^2Σ) \] with truncation points defined by the cutoffs c and the observed data x_i.
APPENDIX C: REVIEW OF STATISTICS

2. $\mu, \Sigma|z_i, \{\tau_i, \sigma_i\}, \mu \sim \text{Normal / Inverted Wishart draw.}$

3. $\tau_i, \sigma_i|z_i, \mu, \Sigma, \phi, \Lambda \sim \text{Metropolis-Hastings draw.}$

4. $\phi, \Lambda|\{\tau_i, \sigma_i\} \sim \text{Griddy Gibbs.}$

5. $c|x_i, \{\tau_i, \sigma_i\}, \mu, \Sigma \sim \text{Metropolis-Hastings draw with likelihood equal to the product of probability mass associated with observed data } \{x_i\}.$

“Griddy Gibbs” is a procedure that discretizes the parameter space and considers moves along a grid of points. Exact expressions for all the draws can be found in Rossi, Allenby and McCulloch (2005). The cutpoint model is implemented in the program “rscaleUsage” in the bayesm package in the R statistical language.

**Logit Model**

The logit model is used to describe discrete choices, where only one of the choice options is selected or identified as being the best. Each choice is associated with a probability of selection, and the purpose of using a logit model is to understand what makes a choice more likely. The likelihood for one observation of a discrete choice model takes the form:

$$p(\theta|\text{Data}) \propto p_1^{x_1}p_2^{x_2} \cdots p_k^{x_k}$$

where $x_j$ is a dummy variable equal to one if the $j^{th}$ choice alternative is selected. The choice probabilities, $\{p_i\}$ are required to be positive and sum to one. The binary logit model for two choice options $i$ and $j$ specifies choice probabilities as:

$$p_i = \frac{\exp [x_i'\theta]}{\exp [x_i'\theta] + \exp [x_j'\theta]}$$

and the multinomial logit model specifies choice probabilities as:

$$p_i = \frac{\exp [x_i'\theta]}{\sum_k \exp [x_k'\theta]}$$

We note in the above expressions that the numerator and denominator can be divided by an arbitrary constant without affecting the choice probabilities, and it is customary to statistically identify the logit model by selecting a choice alternative as a standard,
or norm, from which other choice-specific parameters are estimated. If the first choice option is selected as the norm, then the choice probabilities become:

\[
p_i = \frac{\exp \left[ (x_i - x_1)'\theta \right]}{1 + \sum_k \exp \left[ (x_k - x_1)'\theta \right]}
\]

The multinomial logit model is used in product analysis to understand the drivers of utility, where \(x\) are variables describing the choice alternatives and the vector \(\theta\) is referred to as part-worths associated with the levels of product attributes. It is customary in these models to allow for a “none” choice option that reflects the decision to not make a purchase in the product category. The none option is typically used as the norm from which all attribute levels are specified, and the choice model simplifies to the expression:

\[
p_i = \frac{\exp \left[ x_i'\theta \right]}{1 + \sum_k \exp \left[ x_k'\theta \right]}
\]

A Bayesian analysis of the logit model proceeds by assuming a prior distribution for the parameter vector \(\theta\), which we choose to be multivariate Normal:

\[
p(\theta) = \text{Normal}(0, A^{-1})
\]

and by Bayes theorem the posterior is proportional to the likelihood multiplied by the prior:

\[
p(\theta|\text{Data}) \propto \prod_{t=1}^{n} (p_{1,t}^{x_{1,t}} p_{2,t}^{x_{2,t}} \cdots p_{k,t}^{x_{k,t}}) \times p(\theta)
\]

where “\(t\)” indexes the observations.

The Metropolis-Hastings algorithm can be used to estimate the parameters of this model by selecting an initial value of \(\theta\), i.e., \(\theta^{old} = 0\), and repeating the following steps until the accepted values of \(\theta\) converge to a stationary distribution:

1. Generate a candidate value of a parameter \(\theta^{new}\) using the old value plus a symmetric disturbance: \(\theta^{new} = \theta^{old} + N(0, v^2)\) where \(v^2\) is specified by the analyst so that 30-50% of the candidates are accepted.

2. Compute the acceptance probability \(\alpha = \min \left\{ 1, \frac{p(\theta^{new}|\text{Data})}{p(\theta^{old}|\text{Data})} \right\}\) where \(p(\theta|\text{Data})\) is equal to the product of logit choice probabilities multiplied by the ordinate value of \(\theta\) from a multivariate Normal density with mean zero and covariance matrix \(A^{-1}\). The constant of proportionality cancels in the above fraction and does not need to be calculated.
3. Accept the new draw of $\theta_i$ with probability $\alpha$: draw a Uniform(0,1) random variable and if $U < \alpha$ accept the draw of $\theta_i$. Otherwise, retain the old value of $\theta_i$ and go to step 1.

**Continuous Demand Model**

The continuous demand model is a generalization of the logit model where more than one choice option can be selected and each selection can be associated with an arbitrary positive demand quantity. The demand model assumes that respondents make purchase decisions consistent with a model of constrained utility maximization:

$$\text{Max } u(x, z) = \sum_k \frac{\psi_k}{\gamma} \ln(\gamma x_k + 1) + z \text{ subject to } p'x + z \leq E$$

We incorporate the budget constraint by introducing a Lagrangian multiplier $\lambda$:

$$\text{Max } L = \sum_k \frac{\psi_k}{\gamma} \ln(\gamma x_k + 1) + z + \lambda(E - p'x - z)$$

and setting partial derivatives to zero:

$$\frac{\partial L}{\partial x_k} = \frac{\psi_k}{\gamma x_k + 1} - \lambda p_k = 0$$

$$\frac{\partial L}{\partial z} = 1 - \lambda = 0$$

From the second equation we see that $\lambda = 1$, and we can substitute for $\lambda$ in the first equation to obtain:

$$\frac{\psi_k}{\gamma x_k + 1} = p_k$$

By the Kuhn-Tucker conditions, we know that this expression holds whenever demand is positive, or $x_k > 0$. The expression indicates that marginal utility is equal to price for positive demand. When demand is observed to be zero, we have the condition that marginal utility is less than the price, or that the "bang" is less than the "buck." We can rearrange terms to obtain an explicit expression for $x_k$:

$$x_k = \frac{\psi_k - p_k}{\gamma p_k} \text{ for } \psi_k > p_k \text{ else } x_k = 0$$
The parameter $\psi_k$ is the marginal utility of the $k^{th}$ choice option when $x_k$ is zero. As quantity increases, $(x_k > 0)$, marginal utility declines to the point at which it is equal to the price of the offering. The point at which marginal utility is equal to price is the point of optimal demand. We parameterize the baseline utility parameter $\psi_k$ so that it is positive and contains random error to allow for variation in choice quantities across observations:

$$\psi_{k,t} = \exp[a_k^t \beta + \varepsilon_{k,t}]$$

where $a_k$ is a vector of product attributes for the $k^{th}$ alternative. Substituting the expression for the baseline utility into the demand equation and solving for $\varepsilon_{k,t}$ results in the following expression relating model error to observed demand:

$$\varepsilon_{k,t} = g_{k,t} \text{ if } x_{k,t} > 0$$

$$\varepsilon_{k,t} < g_{k,t} \text{ if } x_{k,t} = 0$$

where

$$g_{k,t} = -\alpha_k^t \beta + \ln(\gamma x_{k,t} + 1) + \ln(p_{k,t})$$

The assumption of extreme-value errors, i.e., $\text{EV}(0,\sigma)$, results in a closed-form expression for the probability that $R_t$ of $N$ goods are chosen:

$$\Pr(x_t) = \Pr(x_{n_{1,t}} > 0, x_{n_{2,t}} = 0, n_{1,t} = 1, \ldots, R_t, n_{2,t} = R_t + 1, \ldots, N)$$

$$= |J_{R_t}| \prod_{i=1}^{R_t} \exp\left(-\frac{g_{i,t}}{\sigma}\right) \prod_{j=R_t+1}^{N} \exp\left(-\frac{g_{j,t}}{\sigma}\right)$$

$$= |J_{R_t}| \prod_{i=1}^{R_t} \exp\left(-\frac{g_{i,t}}{\sigma}\right) \prod_{j=1}^{N} \exp\left(-\frac{g_{j,t}}{\sigma}\right)$$

where $f(\cdot)$ is the joint density distribution for $\varepsilon$ and $|J_{R_t}|$ is the Jacobian of the transformation from random-utility error ($\varepsilon$) to the likelihood of the observed data ($x$). For our model, the Jacobian is equal to:

$$|J_{R_t}| = \prod_{i=1}^{R_t} \frac{\gamma}{\gamma x_{i,t} + 1}$$
The expression for the probability of the observed demand vector $x_t$ is seen to be
the product of $R_t$ “logit” expressions multiplied by the Jacobian, where the purchased
quantity $x_{i,t}$ is part of the value (V) of the choice alternative. For the discrete choice
model discussed in Chapter 4, the value of an alternative is just the first and last term
of the expression for $g_{k,t}$ above, and as a result one can think of the standard logit choice
model as reflecting baseline utility parameters only. The continuous demand model
therefore nests the standard logit model specification.

Estimation of the continuous demand model proceeds with the Metropolis-Hastings
algorithm. Similar to the standard logit model, initial values of the parameter vector
$\theta = \{\beta, \gamma, \sigma\}$ are used to obtain the value of the likelihood and prior, and the Markov
chain is iterated until convergence is obtained.

Constraints are made binding within this model by assuming the outside good satiates:

$$u(x, z) = \sum_k \psi_k \ln (\gamma x_k + 1) + \ln(z)$$

leading to an expression for $g_{kt}$ with $z = E - p_t'x_t$:

$$g_{kt} = -a_{kt}'/\beta + \ln(\gamma x_{kt} + 1) + \ln\left(\frac{p_{kt}}{E - p_t'x_t}\right)$$

with the Jacobian equal to:

$$|J_{R_t}| = \det \left[ \frac{\partial R_t}{\partial x_{R_t}} \right] = \det \left[ \begin{array}{cccc}
\gamma x_{kt} + 1 & p_{1t} & \cdots & p_{R_t} \\
p_{1t} & \gamma x_{kt} + 1 & \cdots & p_{R_t} \\
\cdots & \cdots & \ddots & \cdots \\
p_{Rt} & \cdots & \cdots & \gamma x_{kt} + 1 \\
\frac{p_{1t}}{E - p_t'x_t} & \frac{p_{2t}}{E - p_t'x_t} & \cdots & \frac{p_{R_t}}{E - p_t'x_t} \\
\frac{p_{1t}}{E - p_t'x_t} & \frac{p_{2t}}{E - p_t'x_t} & \cdots & \frac{p_{R_t}}{E - p_t'x_t} \\
\cdots & \cdots & \ddots & \cdots \\
\frac{p_{1t}}{E - p_t'x_t} & \cdots & \cdots & \frac{p_{R_t}}{E - p_t'x_t} \\
\end{array} \right] = \prod_{k=1}^{R_t} \left( \frac{\gamma x_{kt} + 1}{\gamma} \right) \left( \sum_{k=1}^{R_t} \frac{\gamma x_{kt} + 1}{\gamma} \cdot \frac{p_{kt}}{E - p_t'x_t} + 1 \right)$$

The off-diagonal elements of the Jacobian are non-zero because of the right-most term
in the expression for $g_{kt}$. The expenditure allotment $E$ is treated as a parameter and
is statistically identified through the Kuhn Tucker condition associated with positive
demand that result in the equality restrictions $\varepsilon_{kt} = g_{kt}$. 
Heterogeneity

Marketing data often exist in a panel structure where there are multiple observations for each unit-of-analysis. When analyzing respondent-level data, the presence of multiple observations provide information for estimating unique parameters for each respondent. More often than not these estimates are informed by only a handful of responses, and some form of pooling is desirable while still allowing for respondent differences. Bayesian models are well-suited for these broad but shallow datasets where there may be thousands of respondents, each with a limited number of responses to questionnaire items or shopping occasions.

The presence of respondent heterogeneity is easily accommodated in hierarchical models by adding another layer to the model hierarchy. Consider, for example, the regression model in which regression coefficients are heterogeneous.

\[
y_{it} = x_{it}'\beta_i + \varepsilon_{it}; \quad \varepsilon_{it} \sim \text{Normal}(0, \sigma_i^2) \quad t = 1, \cdots, T
\]

\[
\beta_i = \bar{\beta} + \zeta_i; \quad \zeta_i \sim \text{Normal}(0, V_{\beta}) \quad i = 1, \cdots, N
\]

In this model there are T observations per respondent, and N different respondents. The second equation above describes the variation of the regression coefficients among the respondents. Other forms of heterogeneity may include covariates for the mean function, as in a regression model, or autocorrelated errors. The analysis of this model proceeds as follows:

1. Obtain a draw of \( \beta_i, \ i = 1, \cdots, T \) for each respondent, one at a time. For these draws, the prior distribution is the distribution of heterogeneity, i.e., \( \pi(\beta) = \text{Normal}(\bar{\beta}, V_{\beta}) \).

2. Given the \( \{\beta_i\} \), the data are no longer informative about the hyper-parameters \( \bar{\beta} \) and \( V_{\beta} \). Draws of these hyper-parameters follow a standard multivariate regression model with one regressor – an intercept – with \( X = \iota \), a column of ones. When regressors are present, \( X \) contains additional columns that describe observable aspects of the respondents.

3. Repeat.

As discussed earlier, the simplicity of these draws stems from the presence of conditional independence in hierarchical models.
Bibliography


