The Pennsylvania State University

The Graduate School

Department of Industrial Engineering

AN INVESTIGATION ON THE IMPORTANCE OF DESIGN FORM AND FUNCTION: MARKET SUCCESS AND CONSUMER PREFERENCES

A Thesis in

Industrial Engineering

by

Taehyun Kim

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2009

The thesis of Taehyun Kim was reviewed and approved* by the following:

Gül E. Okudan Kremer
Associate Professor of Engineering Design and Industrial Engineering
Thesis Adviser

Andris Freivalds
Professor of Industrial and Manufacturing Engineering

M. Jeya Chandra
Professor of Industrial and Manufacturing Engineering
Professor in Charge of Academic Programs & Graduate Program Coordinator

*Signatures are on file in the Graduate School

ABSTRACT

The goal of product design is to obtain maximum functionality and aesthetic beauty at a minimum cost. Additionally consumers are attracted to designs that reflect their use, behaviors, and psychological responses more so than designs which simply alter the visual appearance of a product. When product functions and qualities are similar across products, customers make their purchasing decision upon aesthetic form. Form is a significant factor that improves the value of a product and gives the manufacturer a competitive advantage in business. Overall, the purpose of this study is to examine the relative importance of product form and function as design factors by investigating their impact on the market share trends of companies and consumer preferences. The study uses product characteristics for 1,028 mobile phones released between 2003 and 2008 as a case study.

Multiple linear regression analysis is used to select highly correlated design factors that influence the market share, and Mallow's Cp method is used to determine the best-fitting model. The Partial Regression Coefficients are used to evaluate the relative importance of design factors. The nine mobile phone design features that affect the market share were identified, and block form style was determined as the most important design factor. During this investigation, I have utilized actual market share information, and hence refer to it as historical data mining.

As a complementary analysis, I have investigated consumer preferences as future oriented data mining via a survey study. With the survey, I investigated the relationship between differences in age, race, and gender and examined how these differences affected consumer preferences regarding design factors in mobile phones. The Nominal Logistic Regression method was used to develop a predictive model that describes the relationship between predictors and responses variables.

Using these approaches (historical and future oriented data mining), this study demonstrates how design investments should be directed while making critical decisions during the product development processes.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge and extend my heartfelt gratitude to the following persons who have made the completion of this thesis possible:

My advisor, Dr. Gül E. Okudan Kremer, Department of Industrial & Manufacturing Engineering and School of Engineering Design for her vital encouragement and assistance and inspiration and motivation she extended for this thesis. I am very thankful for her sincere understanding and continuous support from the beginning till the end of my graduate study.

Dr. Andris Freivalds, Department of Industrial & Manufacturing Engineering for all the insightful advice and grateful assistance.

All The Harold & Inge Marcus, Department of Industrial & Manufacturing Engineering faculty members, staff and my lab mates. My parents and family who always support me with unconditional love and prayers from 7000 miles away.

Most especially for my girlfriend, Ji Young Park, a psychology student at the University of Michigan for assisting in the collection of the data, analysis, and becoming my greatest emotional support. I wish to thank all those who helped me. Without them, I could not have completed my thesis.

Chapter 1

Introduction

1.1 Motivation

Every year, manufacturers invest vast sums of money in upgrading their products. However, these development practices remain focused heavily on increasing the number of functions regardless of consumers' interests. These careless manufacturing decisions which produce excessive functionality only add to end users' confusion and frustration. Often consumers are not able to fully enjoy their brand new products because they do not know how to use its newly added functions, and many of their first experiences with the product reflect the slow learning curve associated with feature-dense technology. Indeed, many people only use their products for its primary application. For example, many consumers go no further than using their cell phones for making phone calls and mp3 players for listening to music. For those people, many complicated functions are rarely or never used. Consequently, consumers have difficulty making purchasing decisions based on aspects of the product that they do not intend to use; moreover, end users pay more money for such unwanted functions.

To improve consumer satisfaction, manufacturers should investigate consumers' interests and develop their products accordingly. They also must realize that increasing the number of functions does not always improve profit in a competitive market, especially if these corporations are not able to satisfy customers. Therefore, manufacturers should establish valuable evaluation methods that can screen out unimportant aesthetic product forms and functions.

Furthermore, manufacturers must not forget to consider diversity and subjectivity during the product design process. Consumers' preferences for product forms differ by age, race, gender, and cultural backgrounds. These differences affect consumers' perceptions in product forms differently. Many gender studies have found that there are gender specific colors: generally, blue for male and pink for female (Andree et al., 1990). In fact, some color assignment studies have also found that color preferences differ quite significantly for various age groups (Dittmar, 2001). With advancing age, the preference for blue decreased steadily, whereas the popularity of green and red increased. In addition to changing attitudes toward color, it stands to reason that age differences might also impact preferences for shapes. In particular, older generations who have presbyopia may prefer bigger keypads while younger people may prefer easy-to-carry slim phones. In modern product design and development, therefore, manufacturers need to more carefully target certain populations.

Both form and function are important design factors that motivate consumers' purchases. However, manufacturers sometimes overlook the needs of consumers. Instead, they operate under the philosophy that “newer” and “more” make a better product. But product forms and functions must first meet customers' needs and preferences. Hence, manufacturers should develop appropriate investment strategies for developing their products and strengthen communication with consumers to save manufacturing time and cost.

1.2 Overview of the Research Methodology

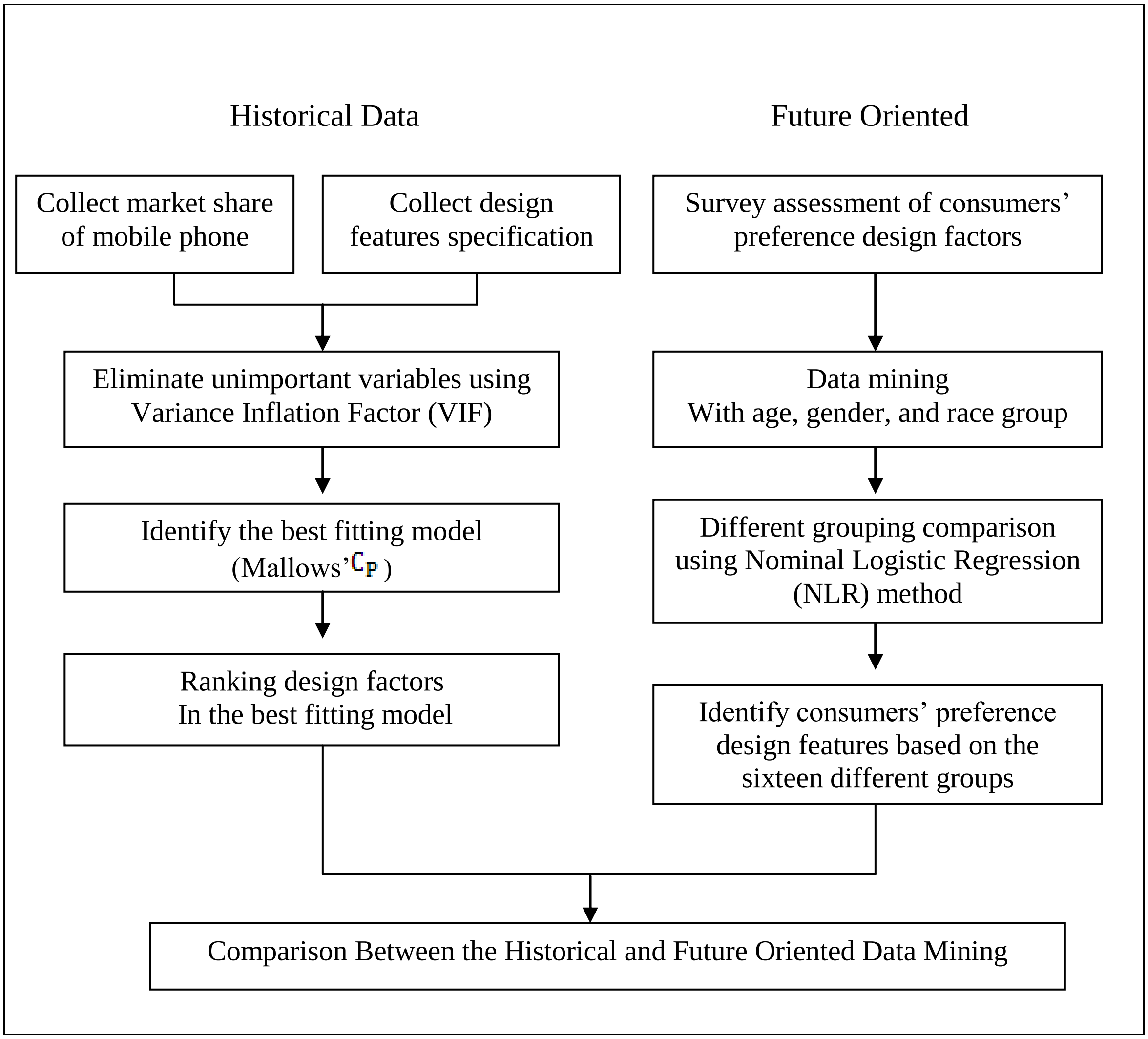

Figure 1-1: Research Methodology

In this study, I have adopted the research flow presented in Figure 1-1. Through historical data mining, this study investigates individual criteria of 1028 mobile phones released between 2003 and 2008 as a case study. Through future oriented data mining, survey assessments were developed and conducted. With historical and future oriented data, multiple linear regression analysis and nominal logistic regression method were applied. By using these two methods, consumers' preferences for design features in the past and their preferences at the present moment were compared. Ultimately, this research framework outlines the most important product design factors for market success, and it suggests factors that should be taken into consideration when striving for innovative product designs in an unpredictable future market.

In the historical data shown in Figure 1-1, the research uses multiple linear regression analysis with respect to manufacturers. First, real data for the released mobile phones and market share for nine consecutive years for manufacturers are investigated. Second, the variance inflation factor (VIF) method is applied to remove unimportant variables and to discern the amount of standard error. Third, Mallows' C_P method was applied to identify the most fitting regression model with chosen variables under VIF detection. Finally, I rank design factors in the best fitting regression model with partial regression coefficients.

With regard to the future oriented data shown in Figure 1-1, the research uses Nominal Logistic Regression (NLR) with respect to consumers. The proposed research framework also represents a survey assessment conducted to more accurately gauge current preferences. First, the survey data was collected from randomly chosen people who reside in the United States. Second, collected data was divided by age group (14-17, 18-22, 23-29, 30-39), gender (male, female), and race (South Koreans residing in the USA, American who resides in USA). Third, NLR method was applied to compare preference factors of the product design by different age groups. Finally, I identify product feature preferences for the sixteen different groups based on age, gender, and race in an attempt to predict emerging trends in product functions and forms.

By then comparing the historical past data and present data, I examined likely shifts in product design processes and functions that consumers prefer. By adopting this approach, companies might better establish more effective product design strategies and product research and development (R &D). In the design decision making process, designers and engineers can determine proper design features to satisfy with company's profit and consumers' expectation.

One of the primary goals of this study is to determine which elements of a product correlate with its impact on a market, ultimately attempting to dissect the "dream product."

1.3 Thesis Organization

In Chapter 2, I examine the literature which considers innovation as it pertains to product form and function. This provides general information about which designs are considered to be good and how design management is related to market success. The literature review also contains methods for classifying and forecasting product forms and functions for improving corporate market share. Chapter 2 also investigates Kansei engineering to understand users' perceptions of product form, to deduce consumers' psychological feelings about technological devices, and to introduce innovative approaches to design methods.

Chapter 3 gives detailed explanations of the research methodology. Chapter 4 then presents the results of this analysis and compares them with historical data mining and future oriented data mining through a survey. For this reason I looked at mobile phones as a case study. For historical data mining, the variance inflation factor (VIF) method was used to remove insignificant variables, and Mallows' C_P method was introduced to determine the best fitting regression model. A partial regression coefficient to rank product design factors is then described in detail. Also, consumers' preference design factors are evaluated through examining survey data. In order to complete future oriented data mining through a survey, the nominal logistic regression model was used for predicting statistical association between two response variables.

The final conclusion of the thesis is presented in Chapter 5, and recommendations for future directions in research on this topic are also discussed.

Chapter 2

Literature Review

In this chapter, a review of the literature on developing product designs for market success is provided. The review is divided into four sections. Section 2.1 provides definitions of "good" design, and what good design might constitute for market success. Section 2.2 highlights how design management relates to market success. Section 2.3 provides methods for forecasting potentially successful product forms and functions from a set of design criteria. Finally, section 2.4 discusses innovative design methods.

2.1. Definitions of Good Design

What is good design? Even though this seems to be a simple question, it is not. In fact, good design can be interpreted variously by designers, engineers, producers, and consumers. However, good design can be defined generally as one that is comfortable to use, looks beautiful, or makes consumers happy (Water, 2007). Some consumers may prefer utility over appearance, while other consumers may prioritize price over utility. Even if consumers' preferences differ in determining good designs, in a practical sense the best selling product must constitute an influential indicator of design quality because of its success in the market. Overall, a well-designed product motivates consumers with aesthetic beauty and gives them pleasure with efficient functional devices (Water, 2007).

In addition to beauty, a good design also carries out its function(s) efficiently (Rawsthorn, 2009). Even if a product design looks cool or unique, aesthetic beauty can be degraded by poor product performance. Resultantly, an inefficient product design makes consumers uncomfortable when they handle a product (Rawsthorn, 2009). As the consumers are confronted with the difficulty of using product functions, they will become dissatisfied and the product overall will fail in the market.

Good designs have beautiful aesthetic forms (Peters, 2007). Aesthetic form can be defined as the culmination of a variety of visual elements influencing consumers' psychological emotions (Veryzer, 1995). Form can be classified into several categories such as motion, size, scale, color, symmetric balance, texture, pattern, rhythm, repetition, and depth (Peters, 2007). These aspects of physical form influence consumers' first impression and influence whether or not consumers will purchase the new product (Marielle and Jan, 2005). Beautiful aesthetic forms are, in general, visual pleasers, and—when they are particularly well designed–they contribute to the effectiveness of the user interface (Marielle and Jan, 2005). However, judging the quality of a design, or design evaluation, is not only concerned with how a product looks or how its functional features fair against competition; it also involves several other factors such as social, economic, environmental, and political dimensions.

Crafting highly influential formal designs requires not only aesthetics and functions, but also designers' insights, as they explore latent human desires and behaviors (Water, 2007). Visual forms and functional devices are important factors to increase consumers' satisfaction, though designers' perceptual standpoint might provide better alternatives (Bloch, 1995). As an example, digital cameras were evaluated for experiment of visual comfort appreciation in a product form (Chang, 2008). Moreover, while designers usually improve product quality based on form and function, consumers' preferences can be shaped by other factors, such as an emergent economic crisis, or socially adopted environmental protection policies. Further, design trends can be contextualized differently within various circumstances such as the surrounding environment, or the historical moment (Bloch, 1995). Thus, good design can be controversial, but cannot be easily estimated by designers, engineers, or consumers.

Traditionally, most companies first increase functionality when they release new products (Lojacono and Zaccai, 2005). However, as the technology of product functions becomes widespread, each product becomes more and more similar to one another. Customers are likely to be motivated more by simple or beautiful looking products than by products with new functional features (Bloch, 1995). By considering aesthetic forms as well as functional efficiency, manufacturers can provide products in various styles to consumers (Jiao and Tseng, 2004; Du et al., 2006). Aesthetic forms, then, can play a significant factor during the product decision making process (Marielle and Jan, 2005). Overall, seeking the ideal physical form necessitates an understanding of consumers' cultural, social, and innate preferences. This search can also provide successful optimization of products in the global marketplace (Bloch, 1995). As a sample method for seeking the ideal physical form, Mcdonagh et al. (2002) proposed product personality profiling (PPP) as a technique to represent the emotional relationship between users and products to be used in evaluating visual product images. While the PPP technique aids user-centered design through discussions of important product appearances between designers and users, it has the limitation of user's highly subjective interpretations (Mcdonagh et al., 2002).

Accordingly, achieving a good design can be seen related to both functionality and aesthetic beauty, and these are two important factors for achieving successful market presence. Therefore, it is very important to understand both consumers' and users' design desires in order to create a successful product in such a competitive marketplace (Water, 2007). Section 2.2 provides explanations on design management for market success.

2.2. Design Management for Market Success

Most people think that design refers to physical appearance and the related aesthetic beauty of a product. However, a design cannot be simply defined through abstract images and aesthetic forms, rather it also implicates underlying technological and engineering decisions which culminate in the functionality of the product. In this section (2.2), I introduce the holistic view of the design process, through which a change in the business is inevitable. In general, design management is important for market success (Lojacono and Zaccai, 2005).

The role of design has evolved for consumers from the past to the present moment. Before the era of mass production, form design was valuable to the few people who purchased works of art and decoration in the eighteenth century, and design's importance was only limited to wealthy consumers with higher social status (Mozota, 2003). In the early days of mass production during the early twentieth century, appearance was still not an important factor when consumers purchased products, but product function was considered the significant factor as, in general, functions brought efficiency (Mozota, 2003). However, by the middle of twentieth century, consumers' income levels and standard of living increased, and accordingly product manufactures began competing with multiple product designs. From 1975 to 1990, companies such as Sony, Philips, and Swatch regarded design as enterprises' effective power for market success (Ravasi and Lojacono, 2005). Since the 1990s, as technological differences within various types of enterprise waned, innovative and creative product designs, which motive consumers' emotions, have increasingly led to market success (Mozota, 2003).

Companies that adopt distinct and innovative design strategies excel in product development when compared to others (Ravasi and Lojacono, 2005). Development of competitive strategies strictly based on cost is no longer sufficient to survive in the long term. Marielle and Jan (2005) indicate that design value will have a significant role, perhaps even more than quality and cost, in the evolving marketplace. For example, in 2003, the Dyson Company introduced the cyclone vacuum cleaner by using a volleyball size ball instead of a small two-wheel system for easy mobility, and in less than two months Dyson's ball DC-15 model has become one of the ten bestselling models of vacuum cleaners (Nancy et al., 2006). Moreover, the Apple I-Pod Nano, which is lighter and smaller than the original I-Pod, but has a larger flash memory size than its older counterpart, increased the company's sales from 4.5 million to 14 million in a single year (Nancy et al., 2006). Desbarats (2005) indicates that the innovative thermal imaging tool, the Argus 3, which is used by firefighters to see through smoke and darkness, is useful because it allows users to grip it with both hands, helping rescuers to capture images more easily in urgent conditions.

Success with distinct and innovative designs is not easily accomplished though. Dubberly (2008) illustrates how the Motorola Razr phone exemplifies a successful case of innovative design strategy in the short term. However, the Motorola Company seems largely to have failed in the long term regarding its design strategy. Motorola's slim and light design style was innovative for consumers who were tired of thick forms and complex functions (McCullagh, 2006). While initially Motorola's market share increased in 2004, the company rapidly lost its market share due to the difficulty of the Razr's user interface (Dubberly, 2008). In general, while Motorola's product design quality keeps up with its rivals, its awkward interface design has made long-term success elusive (Dubberly, 2008).

Companies that more deliberately invest in the early stages of the design process broaden their products' value because it enables them to produce new products by redesigning a previous design form. Furthermore, redesigning a product, as opposed to creating entirely new ones, decreases market risks and design costs (Bangle, 2001). Upgrading a previous model motivates consumers' nostalgia about the past. Consumers are much more likely to purchase a newer version of the same model or series if they have been satisfied with a previous model. The BMW mini cooper model, for example, showed that sales volumes of the redesigned version of the mini cooper surpassed those of the previous model (Bangle, 2001).

By thinking systematically about the design process, companies generate human-centered innovative ideas for market success (Clark and Smith, 2008). Generally, most people think that a creative and talented designer alone can improve a company's profits. However, good ideas and creative design processes also develop by sharing subjective or objective opinions with various experts such as engineers, designers, marketers, psychologists, and anthropologists as well (Lojacono and Zaccai, 2005). Designer's user experiences provide the foundation of new and creative design thinking (Desbarts, 2005). In particular, the IDEO design consulting company shows how design processes based on human beings means direct observation of what people do every day, including an investigation of what they want, why they want it, when their problems occur, and how they think (Brown, 2008). They then implement their ideas by telling a story or making sketches with their own frame (Brown, 2008).

2.3 Design of Products for Successful Market Shares

Manufacturing companies should investigate appropriate product design elements -- product forms and functions --, in order to support their vision, direction, and improved market share. Barone et al. (2007) proposed the Conjoint Analysis (CA) method, which is useful for measuring how consumers' preferences and perceptions change, and also for analyzing the relative importance that each feature plays in determining product purchasing decisions (Barone et al., 2007). By using this method, it becomes possible to more effectively assemble the most highly valued combination of features before launching a product on the market.

The Conjoint Analysis (CA) method is a valuable tool in Kansei engineering (Nagamachi, 1995). Kansei engineering is characterized by close attention to the emotional responses of consumers to product forms and functions, and this type of engineering fosters understanding of consumers' preference elements by evaluating their psychological interactions with product features and appearances (Nagamachi, 1995). Within this framework, rigorous analysis of design elements helps understand customer's latent preferences and predict consumers' consumption in yet unknown cultural environments (Veryzer, 1993). Even though designers' thoughts sometimes provide enough impetus in the product design process, conjoint analysis is a worthwhile approach in that it reduces statistical errors and improves innovative designs.

The Conjoint Analysis (CA) method is used to classify consumers' desire for market research (Pullman et al., 2002), and, in the context of market research, it is useful for analyzing consumers' preference factors in the product design decision making process (Du et al., 2005). However, conjoint analyses do not always provide valid results when testing new products because respondents tend to overestimate their preferences toward less important factors (Barone et al., 2007). Accordingly, through the analysis of current design elements, designers can predict the aesthetic forms and functions in the near future. In this study, design factors are divided into two classes: form and function. Methods that are used to forecast design functions are introduced in Section 2.3.1.

2.3.1 Methods to Use to Forecast Design Functions

A function refers to a particular part of an artifact or an important characteristic related to its operation (Veryzer, 1995). In terms of the general product, a function can be defined by either the simple tasks performed by the device or features used for operating the product (Veryzer, 1995). The meaning of a product function also includes the particular quality and ability which makes possible connections from one device to another (Marielle and Jan, 2005).

An optimal design, then, selects the best combination of features among the extant possible functions. However, designers and engineers have experienced difficulties in determining which functions increase consumers' satisfaction. Previous researchers have developed several methods to deal with this exact issue.

Yan and Li (2008) proposed group technology based on an effective function extraction methodology for data mining. Data mining eliminates unimportant functions from a large volume of functional candidates and selects a subset of predictors. According to Yan and Li (2008), this data mining technique is applied to extract functions of non- stationary signals. Fu and Geng (2007) suggest the data mining model for determining product preference functions for target customers.

Fu and Geng (2008) also discussed the classification method of predicting product functions, in which optimal product functional parameters are selected by identifying specific target consumers. In Fu and Geng's case study (2008) of the decision tree classification method, they analyzed mobile phones with different functions to forecast target consumers. This classification method incorporates five distinct methods: Decision Tree (DT), Artificial Neural Network (ANN), Support Vector Machine (SVM), Bayesian Network (BN), and Ensemble Learning (Li et al., 2008). Furthermore, this approach helps to classify a few important parameters (two or three), but if many parameters are considered, the classification method becomes too unwieldy to find feasible solutions.

The cluster method categorizes similar features by splitting a larger cluster (Kim and Ding, 2005). Kim and Ding (2005) proposed a clustering method in order to help find the optimal design functions in the assembly process of an SUV frame. Within each cluster, this approach helps to minimize the measure of dissimilarity in the functional criteria of unknown samples (Anderberg, 1973). However, cluster algorithms may not have the validity because it is easy to put the preferred data to certain clusters (Anderberg, 1973).

In the product design process, several methods are introduced to screen out product design criteria. Beyond identifying whether or not product forms and functions are important for the user, understanding the relative importance of a set of characteristics might be more meaningful. Yun et al. (2003) proposed the stepwise selection method as the most common procedure for selecting significant characteristics. However, for screening design characteristics, stepwise methods do not show the correct p-value (significance value) because there is not a clear correlation among independent variables. Design decision markers may lead biased results due to their subjective judgments.

In order to eliminate bias from the experts' judgments, several authors suggested the use of methods such as partial least squares (PLS) (Geraldi and Kowalski, 1986), principle component analysis (PCA) (Jolliffe, 1986), cluster analysis (CA) (Anderberg, 1973), and genetic algorithm-based partial least-squares approach (GA-based PLS) (Han and Yang, 2004). The PLS method withdraws the latent variables using a high number of design factors by maximizing their correlation. The PCA method is an attractive technique for analyzing data and determines significant factors, but selected variables might have obscure correlation. In the cluster analysis, variables which have high pairwise correlations are classified into the same cluster, while variables which have low pairwise correlations are assigned into unrelated clusters. GA-based PLS is regarded as a valid model of the best variable combination when compared to previous methods. This method is useful for the product designers to screen out a number of variables in the design process.

Conceptual design methods such as quality function deployment (QFD) (Pullman, 2002), fuzzy sets (Wang, 2002), Analytics Hierarchy Process (AHP) (Saaty, 1980), and Pugh's method (Pugh, 1991) are proposed to screen out unimportant functional concepts among multi-functional conceptual features. These methods are used to evaluate functional priority which satisfies with consumers' requirements and desires.

With QFD, designers gather information from customers' requirement, and this information pushes designers to fulfill the functional requirement as well as to develop more creative and efficient functional designs that save time (Pullman, 2002). When users are faced with vague linguistic answers such as "slightly better" and "much better", the fuzzy set method is used to determine the best design concept. From this approach, ambiguous linguistic terms can be represented by arithmetic operations for evaluating qualified design functions (Wang, 2002).

Saaty (1980) suggested the AHP method, which is performed by comparing the weights of relative candidate concepts in order to lead to a rational decision that reflects several people's decisions. When designers are faced with complex decision making problems, the AHP method provides the most suitable alternatives, and the decision makers systematically evaluate relative importance of elements by comparing various criteria (Saaty, 1980). However, the AHP method does not always screen out functional constraints that consumers want through decision makers' judgments about important criteria. Later, Wang et al. (2006) proposed the evolutionary multi-objective optimization algorithm based on preference. This method is useful for determining the optimal selection about multi-objective and multi-constraint problems toward feasible domain.

When using Pugh's method, if team members have different opinions in terms of selecting conceptual design criteria, they tend to spend too much time to achieve general agreement among the members. Furthermore, if one design team member has weak knowledge, the probability of selecting the right concept will decrease (Pugh, 1991). On the other hand, Pugh's method helps to quickly choose the best conceptual design when each designer independently selects criteria for comparison (Pugh, 1991). Improving the functional quality of a product is regarded as indispensable for creating a successful enterprise. Customers' satisfaction generally increases when quality improves through suitable decisions about functional concepts.

Moreover, Isiklar and Buyukozkan (2007) have proposed a multi-criteria decision making (MCDM) method, which is used to identify features and to control consumers' preferences. In their research the mobile phone industry is also taken as a case study. The MCDM method helps individuals or groups of people select a better potential solution by comparing multi criteria choices (Isiklar and Buyukozkan, 2007). However, this approach may not completely take into consideration the relative importance of criteria if decision makers already investigate or recognize consumers' preference information in choices about product functions. Methods that are used to forecast design form are introduced in Section 2.3.2.

2.3.2 Methods to Use to Forecast Design Forms

A form can be defined as shape, appearance, or style (Bloch, 1995). A form represents the appearance of its outside edges or surfaces, for example whether it is round, square, curved, or rectangular (Marielle and Jan, 2005). A form refers product's appearance. The aesthetic of a form is used to talk about consumers' appreciation of product beauty (Veryzer, 1995).

In the competitive market, designers struggle to understand consumers' preferences about product forms, and it is a challenge to grasp consumers' latent desires (Chang et al., 2005). Designers still have trouble selecting effective design forms because consumers consider such dimensions of a product from a different standpoint based upon perception and emotion (Chang, 2008).

Previously researchers have introduced several approaches to reflect consumers' latent desires for product forms. For example, Chang (2008) explored the relationship between form features and perceptions of visual comfort with product forms, and applied these concepts to digital cameras as a case study. Moreover, Chang (2008) constructed a Kansei hierarchical framework for visual comfort toward product form, and this framework inspired important elements for new digital camera designs. The Kansei hierarchy framework helps designers determine factors influencing visual comfort and appreciation of product forms.

Traditionally, analyses of users' perceptions of form have employed market research techniques by randomly collecting a limited set of participants' evaluation data. However, this strategy sometimes has data errors and thus drawbacks. Nagamachi (1995) proposed the Kansei Engineering method to measure consumers' psychological feeling and impression for product forms, and used it to evaluate different emotional responses. Emotional response is a way to measure stimulus such as sight, smell, hearing, touch, and taste. Through consideration of consumers' subjective judgments and emotion, the Kansei engineering technique selects important forms among potential design elements of the next generation, and uses that information to create new product designs by statistical analysis. Kansei engineering has been applied to many fields including doors (Matsubara and Nagamachi, 1997), car interiors (Nagamachi, 1995), and Mp3 players (Wang and Chen, 2007).

Kansei engineering analyzes the image of a new product and tries to discern consumers' psychological feelings about the product design elements (Nagamachi, 1995). For example, Wang and Chen (2007) point out that product form design using the ANFIS-KANSEI model is able to distinguish the consumers' preferred image of products. For example, Mp3 players have been used as a case study to show various form types. This approach is used to analyze human's psychological response to new product designs in order to improve future designs. Before launching a new product in the market, consumers' preference criteria should be analyzed. Accordingly, research must investigate the types of images that have considerable influence in consumers' psychology (Demirbilek and Sener, 2003).

Kansei engineering method (Nagamachi, 1995) leads to consumer's satisfaction by analyzing their psychological and physical condition. In terms of predicting the design/form combination, a grey relational analysis (GRA) (Lai et al., 2005), Fuzzy support vector machine (fuzzy SVM) (Shieh and Yang, 2007), and fuzzy logic methods (Lin et al., 2007) were applied to mobile phones as a case study. These approaches contribute to understanding consumers' perceptions based on Kansei engineering, and determine the best design elements of mobile phones. The Kansei engineering method helps determine proper product forms for consumers, but the result is based on the subjective responses and insights of consumers. Thus, these methods, based on Kansei engineering, may not reflect all of consumers' latent desires which should be included product design.

Users' perception of design forms has played an important role in recent design trends, and it has been a subject of study with increasing importance. To understand users' perception and product form, the semantic differential (SD) method has been useful to measure connotative meaning (Alcantara et al., 2005). In this method, users' perceptions of a product image are evaluated by using meaning words, and a scale is constructed of contrasting adjective pairs like good-bad.

The semantic differential (SD) method is used to evaluate preference toward form, style, and color in order to elicit personal attitudes towards a design. For example, Alcantara et al. (2005) proposed the differential semantic method for assessing users' form perception of shoes. Hsu et al. (2000) investigated cognitive preferences between users and designers, and users' preferred telephone samples were used to find optimal design factors. Likewise, Zhu et al. (2006) conducted a study on mobile phones to examine designers' color perception. A color image scale approach helped to guide influential color factors in the product market. Designers can make appropriate color selection for mobile phones of different styles, and thereby save time through the use of color image scales. Chuang et al. (2001) evaluated user preference perception in form designs. In this case mobile phones stood as the example selected to identify design trends based on important design components. However, one must consider the influence of the number of subjects when measuring users' perception of form with the semantic method.

McLachlan (2004) explains that discriminate analysis can be used to eliminate the excessive number of factors which play into consumers' different preferences. Indeed, it is useful to classify relevant preference factors among several groups with respect to product image (McLachlan, 2004). However, this method is used to select consumers' perception features in two dimensional space.

Petiot and Grognet (2002) suggest that using the Multidimensional Scaling (MDS) method assesses consumers' subjective impression and feeling and identifies their preferences with respect to new product designs. Whereas the semantic differential method (Alcantara et al., 2005) compared designers' and users' form preferences towards given images or objects, Multidimensional Scaling (MDS) method finds similarities or dissimilarities by identifying users' subjective attributes in their perceptual space in order to arrive at an optimal product design (Petiot and Grognet, 2002). Moreover, MDS provides not only design criteria lists in the perceptual dimension by depending on designers' subjective judgments—like discriminate analysis (McLachlan, 2004) and conjoint analysis (Barone et al., 2007)—but also does so by incorporating consumers' subjective judgments about physical appearance of the product.

The product's appearance, then, is an important evaluation criterion that can improve the product design quality and reflect customers' requirements. The different roles that product appearance occupies in consumers' decision making progress can be identified as communication of aesthetic, symbolic, functional, and ergonomic information, as well as the characteristics of attention drawing, and categorization (Marielle and Jan, 2005). Kelly and Papalambros (2007) indicate that preference mapping (PREMAP) is applied to elicit the end user's preference toward product shape and explores design attributes regarding product quality and visual aesthetic form as they operate in the market. This method helps to determine customer's preference toward shape and their overall perception by using subjective rating scales. In addition, Kelly (2006) points out that Interactive Genetic algorithms (IGA) help effectively evaluate ideal design factors when users make a decision about visual aesthetic preferences in the market place; in this case a cola bottle was used to assess user's shape preferences. IGA generates forms of several alternatives by gathering user's preference data, then designers can take inspiration from attractive designs by understanding consumers' aesthetic preferences (Kelly, 2006). Zhang et al. (2006) specify that the aesthetic evaluation method based on the artificial neural network (ANN) offers image modeling as a way to bring about new aesthetic product forms and reflect human intelligence. It is used to establish desired image forms through the relationship between product images and an aesthetic feeling score. Hsiao and Tsai (2005) propose a hybrid approach based on a fuzzy neural network and genetic algorithm, which enables an automatic product form search or product image evaluation at the conceptual design stage. Parameter based on fuzzy neural network and gray theory is applied to electronic door lock design to search design alternatives and find the best product form image (Hsiao and Tsai, 2005). These approaches help designers to generate various ideas with different points of view and search for the best product form by controlling parameter variables such as form and target image values. Section 2.4 discusses innovative design approach methods.

2.4 Innovative Design Approach Methods

In the past, manufacturers were focused on the benefits of mass production, regardless of the consequences for aesthetic design (Mozota, 2003). However, domination of the market has since changed from producers to customers. Mass customization can satisfy various requirements and expectations of customers, and suggests the possibility to optimize consumers' order (Jiao and Tseng, 2004; Du et al., 2006). Furthermore, as the relative importance of form design grows, Interactive Grammar Based Design System (IGBDS) helps users to develop previous ideas through user interactions (Lee and Tang, 2006). Ultimately, an innovative design approach based on both form and function leads to an increased ability to satisfy consumer's desires in the global market. The moments of change are the ones researchers should study and designers should emulate in the creation of new products (Lojacono and Zaccai, 2005).

Innovative design is based on human beings (Water, 2007). Human beings feel attracted to creative, fresh, and unexpected designs. For example, Apple has recently introduced several innovative products such as the iPhone, iPod, and Macbook. These products changed previous design paradigms by de-emphasizing complex functions and emphasizing aesthetic form. Apple enterprise can predict unknown customer's needs and generate new ideas (Thomke and Feinberg, 2009). Apple's success is based on simple product designs that are intuitive for consumers who don't want to spend much time learning to handle a new product (Yoffie and Slind, 2008). In the end, the design strategy of Apple enterprise strives toward convenience and familiarity for consumers, who find pleasure and gratification in their innovative products (Thomke and Feinberg, 2009).

Innovative design is not only accepted and recognized as a new experience and idea for potential customers, but also includes design factors such as beauty, durability, usefulness, and simplicity. Innovative design is based on understanding the increasingly mobile nature of our changing society, culture, and market. Francis and Tay (2003) explain that an evolutionary design method is used to develop previous information about product form and function. By applying a creativity-based design process, designers can effectively read the changing environment, anticipate future design trends and create redesigned innovative products (Hsiao and Chou, 2004). To generate competitive new designs that satisfy consumer's psychological requirements in the market place, the shape grammar approach has been applied to examples such as a chair (Wang and Chen, 2007), a motorcycle (Pugliese, 2002), a vehicle (Orsborn et al., 2008), and coffee maker (Agarwal et al., 1999). In summary, Lee and Tang (2006) show that Interactive Grammar Based Design Systems (IGBDS) can be applied to the design of a digital camera as yet another example. This system generates a number of design forms by using shape grammar until the users are satisfied with new design forms. To measure successful factors of new products, Horn and Salvendy (2006) display Creative Product Semantic Scale (CPSS) and Consensual Assessment Technique (CAT). Even though these two methods have the weakness of biased judgment and subjective criteria in product evaluation, these measurement tools help predict consumer's perception about innovative products.

Through previous methods, functional characteristics can be measured directly and objectively, but many designers and engineers fail to clearly evaluate product forms due to consumers' subjective input. Part of the problem is that consumers' preferences vary: it is hard to define universal aesthetic product forms that affect market share. Many manufacturers are still not able to use valuable evaluation methods that can screen out unnecessary product forms, despite the fact that aesthetic aspects of products occupy a significant part of consumers' purchasing decision. The method employed in this study of combining both historical data mining and future oriented data mining enable manufacturers to evaluate both form and function more efficiently.

Chapter 3

Research Methodology

In this chapter, a detailed step by step explanation of the proposed methodological framework is discussed. Statistical conclusions are obtained from data analysis, and conclusions can be inherently biased due to poor datasets. Therefore, before carrying out confirmatory data analysis, the exploratory research method must be examined in order to assess the data's integrity. Exploratory research helps gather unclear information in order to illuminate difficult aspects of a problem. This process is most useful during the preliminary stages of an investigation, during which time it provides significant insight for determining the best research methodology (Emory and Cooper, 1991). Exploratory research helps clearly define problematic issues. Moreover it can inspire new ideas and generate creativity. By detecting incomplete, incorrect, or irrelevant records, exploratory research gives researchers greater confidence in the conclusions they draw from analysis. In addition, exploratory research supports the selection of statistical models or tools, assess assumptions of the statistical models, and suggest hypotheses to test. Overall, it leads to well thought out research directions for manufacturers, because it saves time and money in the long run. Exploratory research is a preliminary step in analyzing data.

In the methodological framework for this study (represented in Figure 3-1), historical data mining shows the market share data for the top five mobile phone manufactures and the individual criteria of 1,028 mobile phones, released between 2003 and 2008. Next, unimportant variables were eliminated by using the Variance Inflation Factor (VIF). This process continues until all variables in the model reach a VIF score of less than 10. With Variables removed by VIF score, eight possible regression models are generated to determine the best fitting model.

Mallow's Cp method then functions as a stopping rule in selecting a reduced model. The best fitting regression model is identified by choosing the model with the lowest Cp value. In the best fitting model, important design variables are ranked by using partial regression coefficients.

The methodology framework (represented in Figure 3-1) also includes the future oriented data mining, which uses a survey to evaluate consumers' preferences regarding design factors. Employing a survey as a research tool requires representative sample of individuals, and when used correctly it investigates behavioral patterns and attitudes for a particular target group, and finds out what people think through a direct means (Emory and Cooper, 1991). The objective of a survey is to collect data using either questionnaires or interviews. The questionnaire aspect of the survey is provided in Appendix A. For higher response rates and validities, these direct techniques are utilized because they are more efficient and therefore suggest a better solution than gathering data through e-mail, telephone, and surface mail (Emory and Cooper, 1991). Surveys provide a systematic yet inexpensive way to collect information from very large samples (Emory & Cooper, 1991). With survey data, sixteen different groups based on age, gender, and national origin were classified, and these groups were used to investigate consumers' preferences toward design factors. Moreover, I compared these individuals' preference for various design factors by using the Nominal Logistic Regression (NLR) method.

In this chapter, section 3.1 shows the multiple linear regression method for ascertaining design factors that affect the market share. Section 3.2 reports the process of the nominal logistic regression method to study a relationship between predictor variables and a categorical response variable. By comparing the historical data mining and future oriented data mining, it is possible to forecast consumers' preferred design factors for the next generation of users.

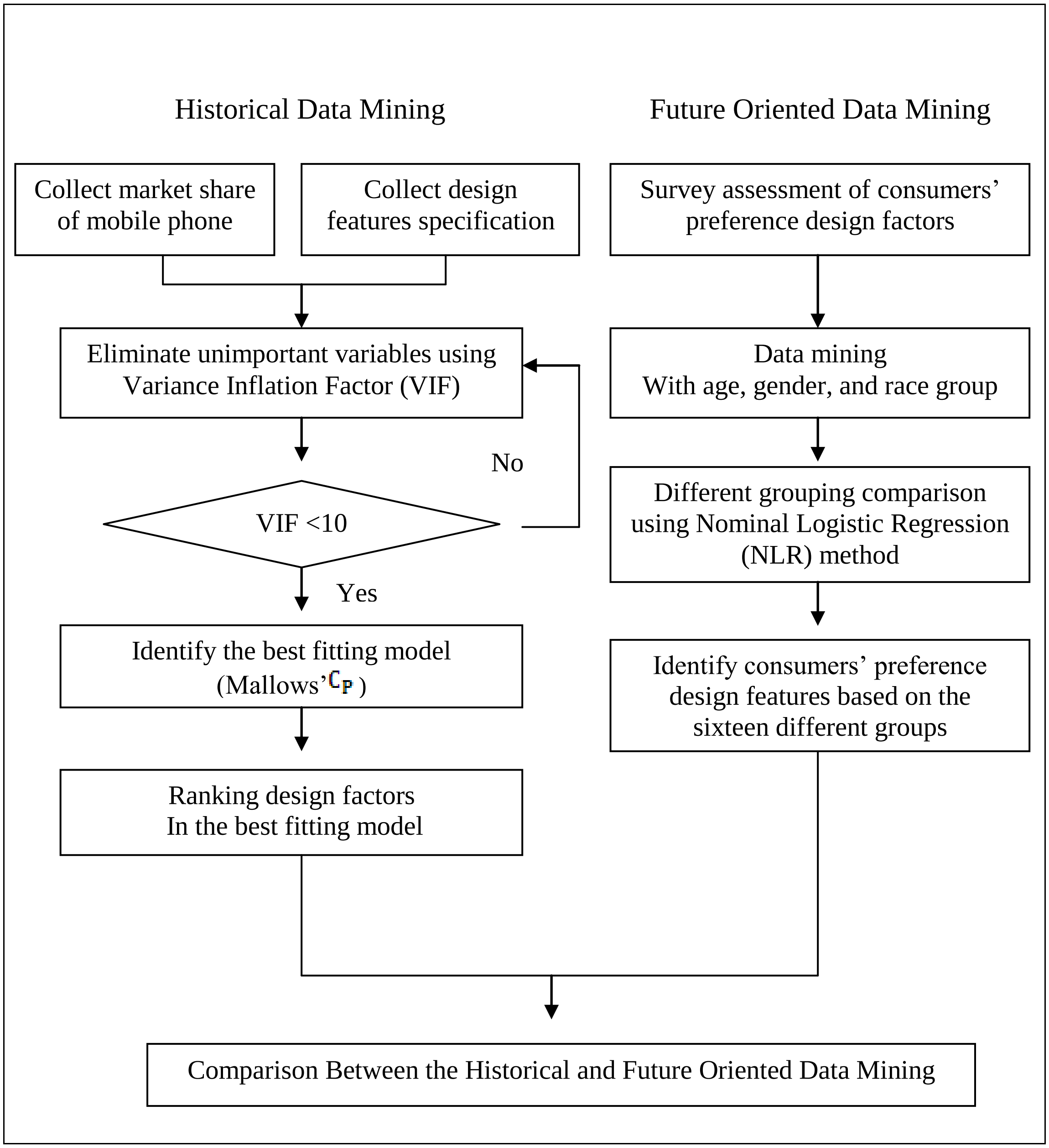

Figure 3-1: Methodology Framework

3.1 Multiple Linear Regression Model

Montgomery and Peck (1982) suggest that multiple linear regression modeling is an appropriate method to model the linear relationship between a dependent variable and one or more independent variables. Independent variables are called predictors, and dependent variables often are called response variables. This model of multiple linear regressions means a linear function of the regression parameters (Neter et al., 1996). This can be written as presented in Equation 3.1:

$$Y= A + \beta_1 \mathbf{X_1} + \beta_2 \mathbf{X_2} + \beta_3 \mathbf{X_3} + \beta_4 \mathbf{X_4} + \ldots + \beta_k \mathbf{X_k} + e \qquad (3.1)$$

Where,

A : Regression Constant

$\mathbf{X_k}$: Independent variables values

β_k : Coefficients of independent predictors

Y : Dependent variables

e : Random errors

However, the multiple linear regression model shows that predictors, as independent variables, can be highly intercorrelated even though predictors are statistically independent from each other (Wilks, 1995). This situation is called multicolinearity (Neter et al., 1996). Multicolinearity can be written as represented in Equation 3.2.

$$\lambda_1 \mathbf{X_1} + \lambda_2 \mathbf{X_2} + \lambda_3 \mathbf{X_3} + \lambda_4 \mathbf{X_4} + \lambda_5 \mathbf{X_5} + \ldots\ldots\ldots + \lambda_i \mathbf{X_k} + \nu_i = 0 \qquad (3.2)$$

Where,

λ_i : Constants

$\mathbf{X_k}$: Explanatory Variables

ν_i : Error terms

In this case, λ_i represents constants and $\mathbf{X_k}$ stands for explanatory variables. When multicollinearity is present, the coefficient estimates that results from regression analysis may change erratically in response to small changes in the model or the data. As a result, it becomes impossible to interpret and rank the relative importance of individual predictors from the estimated coefficient (Wilks, 1995). That is, regression may not give valid results about any individual predictor that exhibits high multicollinearity. Accordingly, Montgomery and Peck (1982) propose that the Variance Inflation Factor (VIF) be used to identify multicolinearity on the variance of estimated regression. VIF formulation will be explained later, in chapter 4 alongside the study's results.

3.2 Nominal Logistic Regression Model

A nominal logistic regression allows us to develop a predictive model that describes the relationship between one or more predictors, such as gender, age, and origin, and categorical response variables, such as responses to a multiple-choice survey (Hosmer, 2000). Logistic regression functions well with regard to a response with three or more possible values that have no natural order (Hosmer, 2000). In logistic regression, each logit estimates the probability of an event in relation to a reference event, using the Equation 3.3:

$$\text{Log (Prob. of event/Prob. of reference event)} = ß_0 + ß_1X_1 + ß_2 X_2 + ...+ ß_KX_K \qquad (3.3)$$

In this equation X_1, X_2, and X_K are the predictors in the model, and both the event and the reference event are pairs of specific responses that are meant to be compared. The coefficient of a predictor is the estimated change in the log of P (response event) / P (reference event) from Equation 3.3 for each unit change in the predictor (Hosmer, 2000). This outcome assumes the other predictors remains constant.

The coefficient can also be used to calculate the odds ratio. By using the model to determine the odds ratio of choosing one response (event) versus another response (reference event) given the value of each of the predictors, an odds ratio compares the odds of two events, where the odds of an event equals the probability of the event's occurrence divided by the probability that it does not occur (Hosmer, 2000). For example, let us see the fourth question in the Appendix A. I want to compare both the Asian group and American groups who prefer easy user interfaces. The response event of each group is the easy user interface. The odds ratio is constructed from Equations 3.4, 3.5 and 3.6 and compares the odds of each level of a categorical response variable to quantify how each predictor affect the probabilities of each response level (Hosmer, 2000).

Odds A = p (Asian choosing easy user interface) / 1- p (choosing easy user interface) (3.4)

Odds B = p (American choosing easy user interface) / 1- p (choosing easy user interface) (3.5)

Odds ratio = Odds A / Odds B (3.6)

If this odds ratio equals 5.0, I conclude that the odds of choosing the easy use interface are five times greater for Asians than Americans.

Chapter 4

Results and Discussion

Section 4.1 explains how this study identifies design factors that affect the market share by using multiple linear regression models; it analyzes relationships between new model productions and market share gained in a year for a mobile phone manufacturer. In section 4.2, the study examines the relationship between different age, race, and gender groups and their most preferred design features for mobile phones, as ascertained through survey data. Section 4.3 employs the chi-square test to determine whether a relationship exists between two categorical variables. In section 4.4, the odds ratio of choosing one response (event) versus another response (reference event) given age, race, and gender values were analyzed. Section 4.5 details which product features might contribute to an innovative product design by comparing historical data mining and future oriented data mining.

4.1 Historical Data Mining

4.1.1 Data Procedures

I collected market share information for mobile phone manufacturers between 2000 and 2008. The top mobile phone manufacturers were chosen based on market share rates over that time period. As shown in Table 4-1, these market shares changed over the years between 2000 and 2008. A few points need mentioning with respect to the data in Table 4-1, *Siemens* data set does not include the values for 2005 to 2008. The market share percentage for *Siemens* is combined with the market share percentages of the category entitled "*other*." While *LG* has been one of the top mobile phone companies for the recent years, there is no market share data between

2000 and 2002. In order to limit data bias, the market share data between 2000 and 2002 for this company was disregarded.

Table 4-1: Market Share Information Over Nine Years (Source: Gartner Dataquest)

MarketShare (%) / Company	2000	2001	2002	2003	2004	2005	2006	2007	2008
Nokia	30.1	35.0	35.8	34.8	30.7	32.5	34.8	37.8	39.5
Motorola	14.6	14.8	15.3	14.5	15.4	17.7	21.1	14.3	10.0
Samsung	5.0	7.1	9.8	10.5	12.5	12.7	11.8	13.4	15.2
Siemens	6.5	7.4	8.2	8.4	7.2	-	-	-	-
Sony Ericsson	10.0	6.7	5.5	5.1	6.2	6.3	7.4	8.8	7.5
LG	-	-	-	5.0	6.3	6.7	6.3	6.8	8.8
Other	33.2	29.0	25.5	21.7	21.6	24.1	18.6	18.9	19.0
Total	100	100	100	100	100	100	100	100	100

In order to underscore which design factors were most important, data for 1,028 different mobile phone characteristics was considered. Then design characteristics were classified into 25 individual criteria as shown in the Table 4-2. Upon examination of the characteristics, it is possible to see that 10 of these are form related, while the other 15 relate to functions. Overall shape of the phone is categorized into seven types, clamshell, block, slider, swivel, flip, dual face or watch.

Table 4-2: Design Characteristics of Mobile Phone

Form Variables	a. Form Style (Clamshell, Block, Slider, Swivel, Flip, Dual face, Watch) b. Weight (g) c. Volume-Length(mm), Width(mm), and Thickness(mm) d. Display Size (Pixels)
Function Variables	f. Network (2G or 3G) g. SAR Value(w/kg) h. Sensor (touch screen, QWERTY key pad, or scroll wheel) i. Contact numbers j. Received call numbers k. Outgoing call numbers l. Shared memory (Mb) m. Photo call n. MP3 player o. Instant Messenger (IM) p. Data Speed (kbps) q. Bluetooth r. Camera Pixels s. Battery standby time(Hour) t. Battery talk time(Hour)

After collecting the market share data for manufacturers and summarizing and classifying the design features of mobile phone models, the variable levels to include in the study were determined. Table 4-3 provides 17 variables along with short descriptions. Mobile phone features can be classified into two groups as per their data type. One group contains discrete variables for which I recorded the total number in the sample; the other includes continuous variables, for which averages were determined. For variables that were classified as discreet, the value from the discrete model data was calculated. For example, if a mobile phone has the Bluetooth feature, it will be added to the total count of mobile phones with the Bluetooth feature. Hence, ‘Bluetooth’ means the total number of models released by a given manufacturer and year that include the Bluetooth feature. Variables that represent the average value were calculated from continuous model data. For instance, ‘Weight’ signifies the average weight value for all mobile phones released by each manufacturer and year.

Table 4-3: Mobile Phone Features Defined as Discrete and Continuous Variables

Features		Descriptions
Discrete Variables	Networks	Most of cell phones utilize 2G networks, and 3G network also supports 2G network. Number of phones which support a 3G or 2G network for a given manufacturer and year.
	Camera	Number of phones that include a camera for a given manufacturer and year
	MP3Player	Number of phones that include a MP3 player for a given manufacturer and year
	Instant Messaging	Number of phones that include instant messaging capabilities for a given manufacturer and year
	Bluetooth	Number of phones that include Bluetooth capabilities for a given manufacturer and year
Continuous Variables	SAR (w/kg)	Average measurement of radio frequency in order to communicate with the network
	Weight	The average weight of mobile phone
	Volume	The average volume of mobile phone based on length, width, and thickness
	Contact	To measure the average contact numbers that can be stored to the mobile phone
	Received	To measure average received calls that can be stored to the mobile phone
	Outgoing	To measure average outgoing calls that can be stored to the mobile phone
	Shared Memory	Average memory to share features such as multimedia messages, images, text, phone book, ring tone, and games
	Data Speed	The average speed of transferring data supported by mobile phone
	Standby time	Without recharging the battery, average standby time for operating mobile phone.
	Talk time	Without recharging the battery, average talk time to communicate on mobile phone.
	Display Pixels	The average size of the mobile phone screen
	Camera Pixels	The average resolution of the camera

After an initial assignment of product characteristics as either “form” or “functional”, I further verified my assignments by conducting a patent search for all these characteristics. Not surprisingly, for form characteristics all the patents found were design patents. Further, most of the patents for the functional characteristics were utility patents. Table 4-4,Table 4-5, and Table 4-6 contain the information reached.

Table 4-4: Patents I for Function Characteristics

Feature		Patent No	Patent Date	Description
Function	Camera	US D500,992 S US 7,228,151 B2	Jan. 18, 2005 Jun. 5, 2007	The function that can capture either photographs or motion video
	Rotatable Camera	US D528,093 S US 0088310 A1	Sep. 12, 2006 Apr. 27, 2006	The function that rotates camera mobile phone without users' movement.
	QWERTY Keypad	US D561,723 S US 7,107,018 B2	Feb. 12, 2008 Sep. 12, 2006	QWERTY Compact for on-hand typing using three fingers that are placed over the keyboard having such small, portable size
	Navigation Trackball	US 0188471 A1	Aug. 16, 2007	The device that points out designated information on a display screen as navigation tool device.
	Touch sensitive scroll wheel	US 0134578 A1	Jun. 23, 2005	The function that uses Synaptic' knack for navigation by making scroll wheel clickable
	Touch screen	US D558,756 S US 0052422 A1	Jan. 1, 2008 Feb. 28, 2008	The function that touch or contact to the display of device by finger or hand
	MP3 Player	US 0027385 A1	Feb. 3, 2005	The function that plays mp3 file, digital audio file compressed using a standard defined by Motion Picture Experts Group (MPEG)
	Instant Messaging	US 7,200,634 B2	Apr. 3, 2007	The function that sends real time messages to another mobile phone user. It is much faster and simpler way to communicate than using e-mail
	Wireless Banking Service	US 7,258,267 B2	Aug. 21, 2007	M-banking Service used for performing balance checks, account transactions, and payment. The function that can use the mobile banking service anywhere if you have your own PIN
	GPS	US 0065326 A1	Mar. 13, 2008	The Global Positioning System (GPS) is a satellite-based navigation system to help accurately determine their locations world- wide.

Table 4-5: Patents II for Function Characteristics

Feature		Patent No	Patent Date	Description
Function	Battery Lifetime	US 6,463,305 B1	Oct. 8, 2002	Time that can operate mobile phone without recharging the battery. Battery power is provided in a form which specifies available talk time and available standby times.
	2G Network	US 6,594,242 B1	Jul. 15, 2003	CDMA-based personal communication service that allows for the introduction of digital data services, such as short messaging service (SMS) and e-mail.
	3G Network	US 7,206,604 B2	Apr. 17, 2007	Third generation that represent an international standard for W-CDMA (Wideband Code Division Multiple Access). It results in faster data transmission for advanced multimedia services and large network capacities.
	Multimedia	US 0005767 A1	Jan. 3, 2008	The function that can experience various mobile internet services such as shopping, game, and video with your phone browser.
	Call Log Memory	US 6,920,208 B1	Jul. 19, 2005	The memory capacity that can save received, dialed, and missed call number, and the approximate length of your calls in the network service.
	Personalization	US 7,248,835 B2	Jul. 24, 2007	The function that adjusts various phone settings for different events and environments and customizes the profile you want to change
	Shared Memory	US 6,681,287 B2	Jan. 20, 2004	The amount of memory shared other features such as phone book, text and multimedia messages, images and ringing tones in gallery, calendar, to-do notes, and Java games and applications
	Bluetooth	US 7,242,970 B2	Jul. 10, 2007	A wireless hand free kit that uses low power radio communications over short distances from fixed and mobile
	One Touch Dialing	US 7,190,975 B2	Mar. 13, 2007	The function that assigns a phone number to one of the speed dialing keys from 1 to 500 in the address book.

Table 4-6: Patents for Form Characteristics

Styles		Patent No	Patent Date	Description
Form	Clamshell	US D566,670 S	Apr. 15 2008	When the clamshell is open, the device is ready for use. The interface components are kept inside the clamshell, which offers more surface area than the device is closed.
	Folder	US D571,344 S	Jun. 17, 2008	A folder form has a main body, and is in two or more parts that fold via a hinge
	Slide	US D570,815 S	Jun. 10, 2008	A slide form made of two or more parts that slide against each other
	Swivel	US D541,248 S	Apr. 24, 2007	A swivel form made multiple segments that swivel past each other around a point.
	Block	US D552,058 S	Oct. 2, 2007	A block form is a rigid rectangular cuboids, and it resembles a candy bar or slab in size and shape.
	Watch	US D543,192 S	May 22, 2007	A watch form is constructed in a manner such that a pair of housing facing each other can rotate relative to each other

In Table 4-7, visual form variables are classified into seven categorical types. The seven different form styles are defined as discrete variables. Instead of calculating the average value for continuous variables, discrete variables were computed according to the number of the mobile phone’s form style that manufacturers released for given years. Since over 95 % of all mobile phones are blocks, clamshells, or sliders, these three main variables are represented as the manufacture’s preferred form.

Table 4-7: Visual Form style

Form		Description	Count
Discrete Variables	Block	# of block phones that manufacturers released for given years	401
	Clamshell	# of Clamshell phones that manufacturers released for given years	394
	Slider	# of Slider phones that manufacturers released for given years	187
	Flip	# of flip phones that manufacturers released for given years	26
	Swivel	# of swivel phones that manufacturers released for given years	17
	Dual Face	# of dual face phones that manufacturers released for given years	2
	Watch	# of watch phones that manufacturers released for given years	1
Total			1028

The Variance Inflation factor was used to measure the magnitude of multicollinearity. Multicollinearity refers to a situation in which three or more variables are highly correlated in a multiple regression model. The variance inflation factor (VIF) is defined from Equation 4.1

$$\mathbf{VIF}_j = \frac{\mathbf{1}}{\mathbf{1 - R^2}_j} \quad (4.1)$$

In this case, $\mathbf{R^2}_j$ is the measure of the squared coefficient of the multiple correlations. If the VIF is bigger than 10, it indicates the existence of serious multicollinearity. As shown in Table 4-8, the variables with a high VIF were identified and deleted from the regression model. After removing the variables with high VIF values, a new regression model was generated without the removed variables. This process continued until all variables in the model had a VIF score of less than 10. As a result, three variables, ‘DReceived’, ‘DVolume’, and ‘DSlider’, were deleted. Table 4-7 presents these deleted variables. For example, ‘DReceived’ means the change in ‘Received’ from one year to the next year. Because I am interested in modifications or improvements of manufacturers that affected the market share, I calculate the difference of each of these variables. After all, potential variables remained as shown in Table 4-9. Variables which remained after accounting for VIF scores generated all possible regression models.

Table 4-8: Removed Variables Based on VIF scores

Step	Removed Variable	VIF	Model R-square	F-value	ρ -value
1	DReceived	149.63	95.3	3.19	0.184
2	DVolume	65.78	94.5	3.84	0.101
3	DSlider	32.39	94.5	5.07	0.041

In this study, the relative importance of nine design factors was determined through the partial regression coefficients method. Based on the ranking of design factors, customers prefer the block and clamshell types as design form. Furthermore, the existence of a digital camera is still a necessary function of mobile phones. Also, as manufacturers increase the average number of contacts, the average data transfer speed, and the average talking time, the market share of mobile phone can increase. However, there are some interesting conclusions. As manufacturers increase average share memory and Bluetooth-enabled phones, their market share is decreased.

4.1.2 Mallows' Cp and Multiple Linear Regressions

Mallows' C_p was used for identifying the best fitting model among several choices, which were all possible subsets of the independent variables (Mallows, 1973). This procedure provides all candidate regression models and determines the best-fitting models in multiple regressions. Table 4-9 indicates that the top eight best fitting models are selected in terms of C_p. Based on the maximum R-sq (adjusted) and the minimum C_p, the best subset model is chosen. It consists of nine independent variables, and the value of the minimum C_p is calculated as 4.9, which is calculated using Equation 4.2

$$C_p = \frac{RSS_p}{S^2} - N + 2P \tag{4.2}$$

Where:

C_p : Subset criterion to be used in selecting a reduced model

N: the number of observations

P: the number of variables in the regression

$RS{S_p}$: the residual sum of square error using p variables

S^2 : The residual mean of square error

The best fitting model which is selected in terms of C_p represents the following multiple regression model from Equation 4.3

$$Y = A + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 \tag{4.3}$$

Where,

A: The constant term

X_k : Independent variables for design factors of mobile phone

β_k : Coefficients of independent design factors

Y: Dependent variables for market share of mobile phone

Table 4-9: The Top 8 Best-Fitting Models in Terms of C_p

#Variables	R^2	R^2 (adj)	Cp	s	DWeight	DContacts	DOutgoing	DShared	DDataSpeed	DBattery	DTalk	DDisplayPixel	DCameraPixel	DBlock	DClamshell	D3G	DCamera	DMP3	DIM	DBluetooth
9	92.8	87.8	4.9	0.65	x	x		x	x		x			x	x		x			x
8	90.2	84.6	5.7	0.73		x		x	x		x			x	x		x			x
9	91.7	85.9	6.1	0.70	x	x		x	x					x	x		x	x		x
10	93.5	88.2	6.1	0.64	x	x		x	x		x			x	x		x		x	x
10	93.1	87.4	6.5	0.66	x	x		x	x		x			x	x		x	x		x
11	93.9	87.8	7.7	0.65	x	x		x	x		x			x	x		x	x	x	x
11	93.7	87.5	7.8	0.66	x	x		x	x	x	x			x	x		x		x	x
7	85.1	78.1	9.3	0.87		x		x	x					x	x		x			x

β_k represents the change in Y corresponding to a unit increase in X_k, holding all other variables constant. ‘DWeight’, ‘Dshared’, and ‘DBluetooth’ have negative coefficients. Negative coefficients signify a decrease in the market share as predictors increase. Oppositely, ‘DContacts’, ‘DDataSpeed’, ‘Dtalk’, and ‘DCamera’ all had positive coefficients. As average contact numbers, average data transfer speeds, average talk time, or camera phones improved, the manufacturers’ market share increased. In addition, both ‘DBlock’ and ‘Dclamshell’ have positive coefficients. This may not tell us much, though, because block and clamshell styles were the predominant forms of mobile phone overall.

4.1.3 Design Factors that Affect the Market Share Trends

Partial regression coefficients were calculated to identify regression weights of independent variables that affect the response variable (Pedhazur, 1997). Partial regression coefficients were used to rank design factors that affect the market share trends. In Table 4-10, all p value are less than 0.05, and all design factors in the best fitting model have a score of less than 10. This suggests that design variables of mobile phones are moderately correlated. The best fitting model reduces multicollinearity by removing unimportant independent predictors. Among remaining design factors of the best fitting model, the most important design factors can be calculated. As a result, block form style of mobile phone is regarded as the most significant factor of design. Bluetooth enabled mobile phone is determined as the second significant factor of design in function part. By using Equation 4.4, the coefficients of β_k can be interpreted as shown Table 4-9.

$$\beta_k = \frac{S_k}{S_y} \times b_k \tag{4.4}$$

Where,

β_k : The standardized partial regression coefficient or the β -Weight

b_k : Coefficients of the Kth independent variable

S_k : The standard deviation of the Kth independent variable

S_y : The standard deviation of the dependent variable

Table 4-10: Weight of Design Factor for Mobile Phones

Attribute / Predictor	$\boldsymbol{\beta}$k	Coefficient	Sk	Sy	VIF	T	p	$\boldsymbol{\beta}$-Weight	Ranking
Dweight	$\boldsymbol{\beta}1$	-0.059	0.52	.648	.21	2.15	.05	0.50	9
Dcontacts	$\boldsymbol{\beta}2$	0.004	11.4	.648	.67	0.16	.00	1.89	4
Dshared	$\boldsymbol{\beta}3$	-0.002	18.9	.648	.23	5.35	.00	1.22	6
DdataSpeed	$\boldsymbol{\beta}4$	0.001	3.0	.648	.62	0.12	.00	1.12	7
DTalk	$\boldsymbol{\beta}5$	1.148	0.39	.648	.51	0.65	.02	0.68	8
DBlock	$\boldsymbol{\beta}6$	0.310	0.15	.648	.79	0.53	.00	2.94	1
Dclamshell	$\boldsymbol{\beta}7$	0.112	0.13	.648	.95	0.26	.00	1.58	5
Dcamera	$\boldsymbol{\beta}8$	0.184	0.94	.648	.20	0.66	.00	2.54	3
Dbluetooth	$\boldsymbol{\beta}9$	-0.144	2.69	.648	.16	6.22	.00	2.82	2

In this study, the relative importance of nine design factors was determined by the partial regression coefficients method. Based on the ranking of design factors, customers prefer the block and clamshell types as design form. Also, as manufacturers increase the average number of contacts, the average data transfer speed, and the average talking time, the market share of mobile phone can increase. However, there are some interesting conclusions. As manufacturers increase average share memory and Bluetooth-enabled phones, their market share is decreased.

4.1.4 The Relationship between New Model Productions and Market Share

Beyond the above mentioned results, in order to understand the mobile phone market, it is necessary to plot market share data and new model productions for each mobile phone manufacturer. There did not always appear to be a constant trend of market share for some manufacturers. Only, Samsung and LG appeared to have significantly increasing market share trends. Furthermore, in order to get some sense of the relationship between new model productions of mobile phone and market share, I overlaid the number of new models that each manufacturer produced per year with the market share data for those companies

The data for Nokia is shown in Figure 4-1.1. Between 2003 and 2008, Nokia produced an average of 33 new mobile phone models each year. New model productions did not have an overall trend in increment or diminishment every year. The market share of Nokia has averaged 35% since 2003. It increased constantly at an average rate of 2% per year since 2004. Compared to 2005, Nokia produced 6 less new phone models in 2006. Thus, new models productions of mobile phone did not affect market share trend.

Motorola's information is shown in Figure 4-1.2. Between 2003 and 2008, Motorola produced an average of 28 new models of mobile phone each year. During same period, their average market share was 15%. However, the year 2005 was a turning point for Motorola, because it released 42 new models. In 2006, though, Motorola only produced 12 new models. Motorola produced 30 fewer new models in 2006 than in 2005. On the other hand, their market share increased by 3.4 % during same period. Also, compared to 2006, Motorola released 13 more new models of mobile phone in 2007, whereas its market share decreased by 6.8 % overall. Remarkably, the data for Motorola suggests that market share and new model productions of mobile phone had an inverse relationship, meaning that decreasing the output of new products improved the company's market share.

Samsung's data is represented in Figure 4-1.3. It produced an average of 65 new models per year between 2003 and 2008. New models of mobile phones continuously increased every year except 2008. In 2008, Samsung produced six fewer new models than 2007. Beginning in 2005, Samsung released twice as new models as they did in 2004. Samsung produced an average of 88 new models per year between 2005 and 2008. The average market share showed 13% between 2003 and 2008. In addition, their market share constantly increased and averaged 14% for last three years. From 2005 to 2006, new model productions increased, but the market share decreased by 0.9%. The Samsung manufacturer showed that market share and new model productions of mobile phone did not have any relationship.

Sony Ericsson manufacturer is shown in Figure 4-1.4. It produced an average of 21 new models per year between 2003 and 2008. Since 2005, Sony Ericsson had produced an average 27 new models. Their market share had continuously increased until 2007, but it decreased by 0.9 % in 2008. Sony Ericsson had occupied an average 7 % market share since 2003. Sony Ericsson rapidly introduced 15 more new models in 2005 than 2004, but market share during this time stayed the same rates. Therefore, new model increases of mobile phones did not appear to have an important effect on market share.

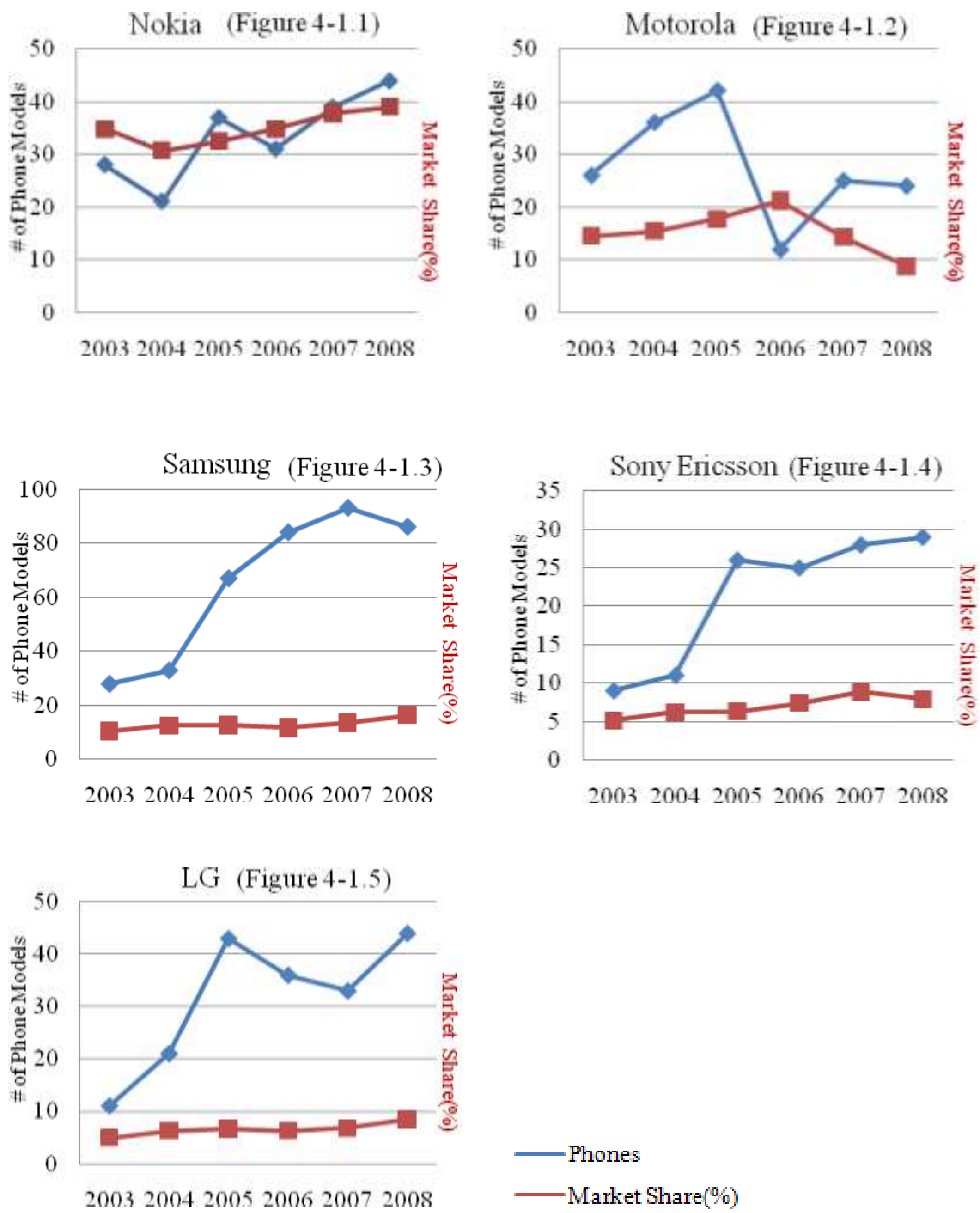

Figure 4-1: Market Share vs. New Models Productions of Cell phones

Information for the manufacturer LG is depicted in Figure 4-1.5. From 2003 to 2005, LG released an average of 31 new models per year. From 2003 and 2005, the new model productions of mobile phone swiftly increased, but they largely decreased between 2005 and 2007. LG released 11 more new models in 2008 than they did in 2007. The average 39 new models were produced for the last four years. Moreover, LG's market share constantly increased between 2003 and 2008, regardless of new model productions. During the same period, their average market share was 7 %. The LG manufacturer demonstrated that the market share trend and new model productions of mobile phone did not show any important correlation.

4.1.5 Market Share vs. New Models Productions of Cell Phones (With Lag)

Secondly, I investigated the degree to which new model productions of mobile phone in one year can affect the manufacturer's market share in the next year. New model productions of mobile phone replace the number of new models in previous year. Accordingly, each manufacturer started with the missing value given the first year. Then, it is necessary to plot both market share and new model productions of mobile phone by each year, between 2003 and 2008.

Data for Nokia is shown in Figure 4-2.1. Between 2004 and 2008, Nokia produced an average of 31 new models per year. Between 2004 and 2005, new model productions of mobile phone decreased from 28 to 21, whereas the market share increased from 30.7% to 32.5 %. Nokia produced 6 less new models in 2007 than in 2006. Although in 2006, Nokia only represented 34.8% of mobile phone market share; by 2007 it achieved a 3% increase in their market share, ending the year with 37.8% of the market share. Thus, new model productions of mobile phone in a year did not affect market share trend in the next year.

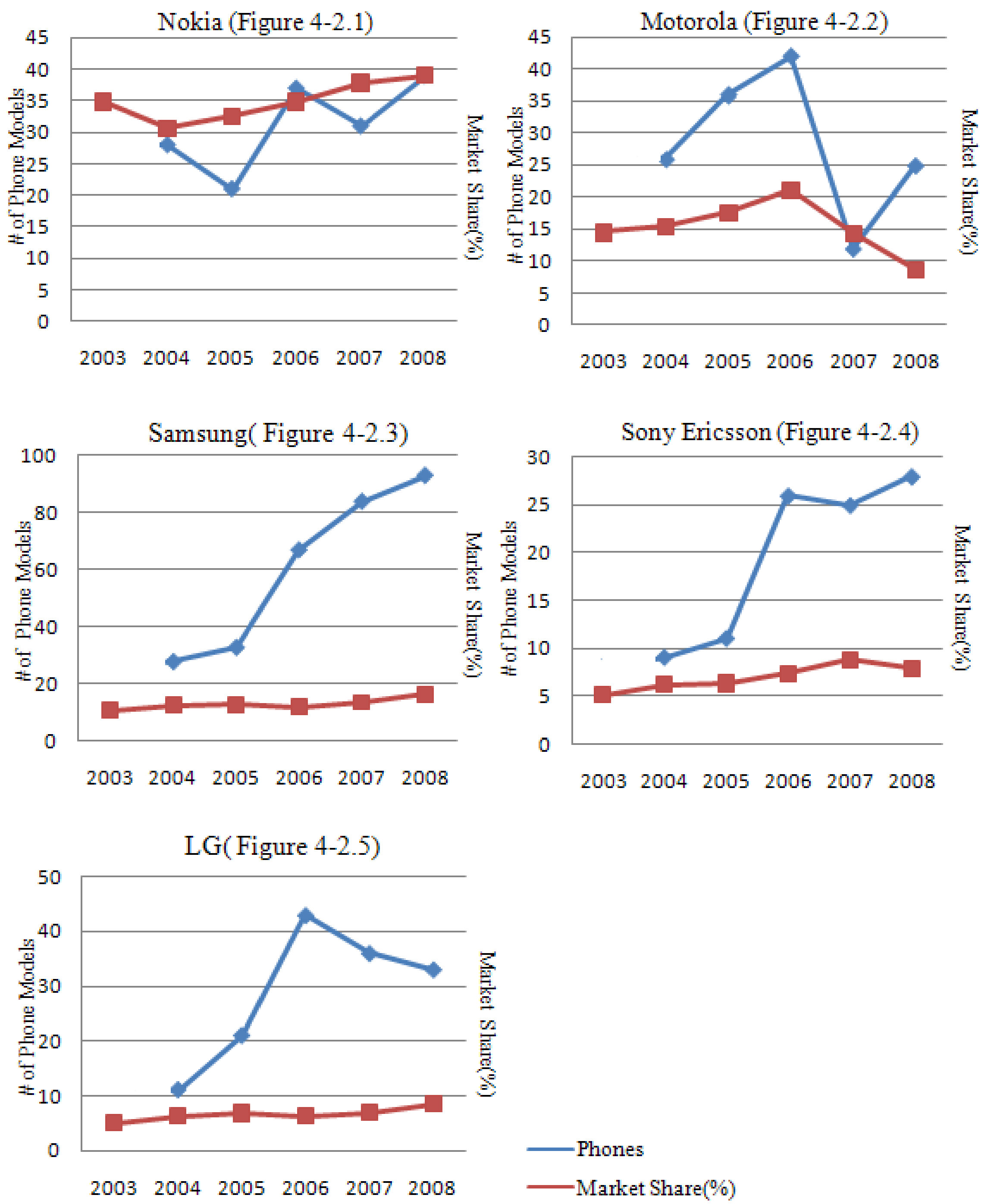

Figure 4-2: Market Share vs. New Models Productions of Cell Phones (With Lag)

Motorola's information is shown in Figure 4-2.2. Motorola produced an average of 28 new models per year from 2004 and 2008. Although in 2007, Motorola only released 12 new models; by 2008 it achieved 13 more new models, ending the year with 25 new models. However, from 2007 to 2008, market share decreased by 5.6 %. Motorola manufacturer appeared that market share trend in the following year and new model productions of mobile phone in a year did not agree very well.

Samsung's figure is shown in Figure 4-2.3. Between 2004 and 2008, Samsung produced an average of 61 new models each year. In the recent three years, Samsung produced an average of 81 new models. From 2005 to 2006, production of new models of mobile phone rapidly increased, but the market share conversely decreased. Thus, new model productions of mobile phone in a year did not affect the market share in the next year.

Information about Sony Ericsson is shown in Figure 4-2.4. Between 2004 and 2008, Sony Ericsson produced an average of 20 new models each year. Sony Ericsson produced an average of 26 new models per year for most recent three years. Although in 2007, Sony Ericsson only released 25 new models; by 2008 it achieved 3 more new models, ending the year with 28 new models. But, from 2007 to 2008, the market share decreased by 0.9 %. Therefore, new model productions in a year did not appear to have an important effect on the market share trend in the next year.

LG manufacturer is shown in Figure 4-2.5. From 2004 and 2008, LG released an average of 29 new models per year. LG produced 22 more new models in 2006 than in 2005, whereas the market share decreased by 0.4 %. Between 2006 and 2008, the new model productions of mobile phone rapidly decreased, but the market share constantly increased. LG demonstrated that the market share trend in next year and new model productions of mobile phone in a year did not show any relative correlation.

4.1.6 Fitted Line Plot between the Mean New Models Productions and the Mean Market Shares

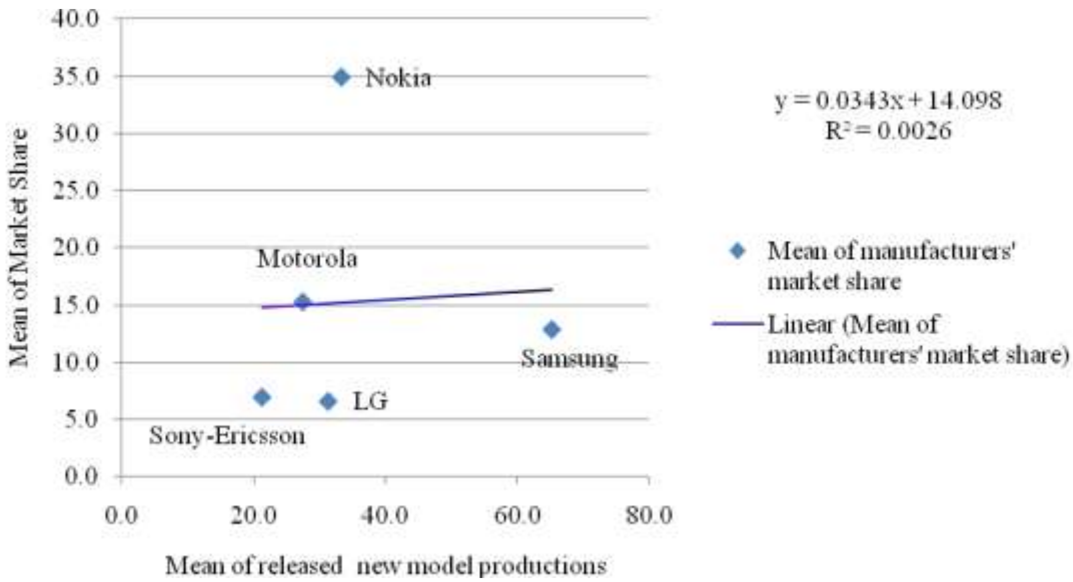

Figure 4-3: Fitted Line Plot

After analyzing the data, Figure 4-1 and Figure 4-2 did not show a consistent trend of market share when compared to new model productions of mobile phones each year. Thus, it is necessary to investigate the potential linear relationship between the mean new model productions and the mean market share if statistical data fits the regression model. In Figure 4-3, I calculated the mean new model productions released by each mobile phone manufacturer between 2003 and 2008, and computed the mean market share.

Figure 4-3 indicated the average market share between 2003 and 2008. Nokia had an average of 34.9 %, placing them at the top of the market share. Motorola's share of the market was second, with an average of 15.3 %. Samsung held onto its third place position in the rankings, with an average of 12.9%. Sony Ericsson is in fourth position in the rankings, and its market share occupied an average of 7.0 %. LG's share of market was last, and it kept an average of 6.6 %.

Figure 4-3 shows that each mobile phone manufacturer differed from the average new model productions per year between 2003 and 2008. Samsung was the largest producer, and it released an average of 65 new models per year. As the second position of new model production rankings, Nokia produced an average of 33 new models per year. LG is third in the production rankings, and its new model productions averaged 31 per year. Motorola produced an average of 28 new models per year. As the last position, Sony Ericsson produced an average of 21 new models per year.

The linear regression model attempted to explain the relationship with a straight line fit to the data. The linear regression model assumes that Y= α X +β b +e, where residual e is a random variable that is normally distributed with a mean zero and having variance. X represents the average new model productions of mobile phone per year as the predictor. Y signifies the average market share of mobile phone manufacturers per year as the response. Y follows a normal distribution. The coefficients α and β were estimated in regression by making the sum of the square residuals as small as possible. The coefficient α confirmed to 0.034 values, and the coefficient β showed 14.09 as a value. The estimated regression equation and goodness of fit statistics were shown at Figure 4-3.

There did not seem to be any discernable relationship between mean new model productions per year and the mean market share per year in Figure 4-3. The very low $\mathbf{R}^2$ value represented the nearly flat fitted line. The scatter point of each mobile phone manufacturer was not close to the linear regression model except Motorola. Normally, I would check the residuals as an assumption of the linear regression. The residual represented the difference between the true value and the predicted value. It should be normally distributed with mean zero and consistent variance. If not, it may have been necessary to consider a transformation. However, I omitted the residual check because of the low $\mathbf{R}^2$ value and the small number of scatter points.

I assumed that the mean market share could be affected by other factors such as years and manufacturers in the dataset. I investigated the effects of years and manufacturers on the market shares and analyzed the relationship between the mean new model production of mobile phone and the mean market share. The fourth linear regression models could be made using the Analysis of Covariance (ANOCA) as provided in Figure 4-4.

Montgomery (1991) proposes that ANCOVA is the best general linear model with one continuous response variable and one or more factors. This model helps to check certain factors to affect on response variable, and it removes the effects from other independent variables, which are known as covariates. Generally, the inclusion of covariates, which accounts for very effective variance in the response variable, can increase statistical confidence, while a covariate which explains for little variance in the response variable can decrease the accuracy of data analysis (Montgomery, 1991). Careful choice of covariates is a significant factor that can increase statistical power.

In Figure 4-4, there was the term *Manufacturer*phones*. This is the interaction term. In the statistical model, an interaction is added when the effect of two or more variables is not addition. Such a term reflects that the effect of one variable depends on the value of one or two more variables. For example, an interaction between adding sugar to coffee and stirring the coffee is important because neither adding sugar, nor stirring coffee has much effect on sweetness but both adding sugar and stirring the coffee, when done simultaneously, makes the coffee sweeter. In this case, if the interaction of manufacturer*phones is significant, new model productions of mobile phone can lead to different market share for each manufacturers. It means that the new model production of mobile phone may increase market share for one manufacturer, while having no effect for another producer. If the interaction term is meaningful, then each factor included in this interaction is also significant. However, if it is not, the importance for each factor included in the interaction can be assessed by looking at the p-value.

4.1.7 General Linear Model – ANCOVA Models

Factor	Type	Levels	Values
Manufacturer	random	5	LG, Motorola, Nokia, Samsung, Sony-Ericsson
Year	random	6	2003, 2004, 2005, 2006, 2007, 2008

Analysis of Variance for Market Share (%), using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Phones	1	26.51	20.57	20.57	2.80	0.115
Manufacturer	4	3216.03	123.90	30.98	4.21	0.018
Year	5	10.23	7.16	1.43	0.19	0.960
*Manufacturer*Phones*	4	26.56	26.56	6.64	0.90	0.487
Error	15	110.39	110.39	7.36		
Total	29	3389.71				

S = 2.71284 R-Sq = 96.74% R-Sq(adj) = 93.70%

Term	Coef	SE Coef	T	P
Constant	10.875	2.611	4.17	0.001
Phones	0.13600	0.08135	1.67	0.115
*Phones*Manufacturer*				
LG	-0.02881	0.09037	-0.32	0.754
Motorola	-0.1665	0.1431	-1.16	0.263
Nokia	0.2386	0.1365	1.75	0.101
Samsung	-0.08230	0.06895	-1.19	0.251

Unusual Observations for Market Share (%)

Obs	Market Share (%)	Fit	SE Fit	Residual	St Resid
9	17.0	14.0783	2.2858	3.6217	2.48 R
10	21.1	16.3624	2.2534	4.7376	3.14 R
12	8.70	14.5426	1.8107	-5.8426	-2.89 R

Figure 4-4: ANCOVA with Year, Manufacturer and Manufacturer*Phones

As shown in Figure 4-4, a significant relationship could not be found between the mean market shares and mean new model productions of mobile phones. In order to examine if both the 5 level categorical variables of *manufacturer* and the 6 level categorical variables of *Year* had a relationship with continuous variable of *market share,* ANOVA was done. Also, *Phones* as the covariate added in the ANCOVA model. R^2 , the coefficient of determination, indicates 96.74%. This means that the model is very useful because a value close to 1. However, Figure 4-4 shows that interaction of *Manufacturer*Phones* is not statistically significant (F (4, 25) = 0.90, ρ = 0.487). From this analysis, it can be concluded that *Manufacturer and Year* as random factors did not have an effect on mean market share.

Factor	Type	Levels	Values
Manufacturer	random	5	LG, Motorola, Nokia, Samsung, Sony-Ericsson
Year	random	6	2003, 2004, 2005, 2006, 2007, 2008

Analysis of Variance for Market Share (%), using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Phones	1	26.51	3.70	3.70	0.51	0.482
Manufacturer	4	3216.03	3214.31	803.58	111.48	0.000
Year	5	10.23	10.23	2.05	0.28	0.916
Error	19	136.95	136.95	7.21		
Total	29	3389.71				

S = 2.68479 R-Sq = 95.96% R-Sq(adj) = 93.83%

Term	Coef	SE Coef	T	P
Constant	14.114	1.757	8.03	0.000
Phones	0.03384	0.04721	0.72	0.482

Unusual Observations for Market Share (%)

Obs	Market Share (%)	Fit	SE Fit	Residual	St Resid
10	21.1000	15.6524	1.7535	5.4476	2.68 R
12	8.7000	15.5745	1.6700	-6.8745	-3.27 R

Figure 4-5: ANCOVA with Manufacturer and Year

As shown in Figure 4-5, a strong relationship was not found between the mean market shares and mean new model productions of mobile phone. In order to examine if both the 5 level categorical variables of *manufacturer* and the 6 level categorical variables of *Year* had a relationship with continuous variable of *market share*, ANOVA was done. Also, *Phones* as the covariate added in the ANCOVA model. R^2, the coefficient of determination, indicates 95.96%. It means that the model is very useful because a value close to 1. However, Figure 4-5 shows that *Phones* is not statistically significant (F (1, 28) = 0.51, ρ = 0.482). From this analysis, it can be concluded that *Manufacturer and Year* as random factors are not statistically significant.

Factor	Type	Levels	Values
Manufacturer	random	5	LG, Motorola, Nokia, Samsung, Sony-Ericsson
YearInterval	random	5	2003-2004, 2004-2005, 2005-2006, 2006-2007, 2007-2008

Analysis of Variance for DMkt Share, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
DPhones	1	2.663	2.371	2.371	0.30	0.595
Manufacturer	4	22.113	27.000	6.750	0.85	0.520
YearInterval	4	16.021	10.835	2.709	0.34	0.843
*Manufacturer*DPhones*	4	24.895	24.895	6.224	0.79	0.556
Error	11	86.863	86.863	7.897		
Total	24	152.554				

S = 2.81009 R-Sq = 43.06% R-Sq(adj) = 0.00%

Term	Coef	SE Coef	T	P
Constant	0.7355	0.7540	0.98	0.350
DPhones	-0.0549	0.1002	-0.55	0.595
*DPhones*Manufacturer*				
LG	0.0588	0.1180	0.50	0.628
Motorola	-0.0835	0.1058	-0.79	0.447
Nokia	0.1879	0.1334	1.41	0.187
Samsung	-0.0783	0.1081	-0.72	0.484

Unusual Observations for DMkt Share

Obs	DMkt Share	Fit	SE Fit	Residual	St Resid
11	-4.10000	-0.36780	2.21533	-3.73220	-2.16 R

Figure 4-6: ANCOVA with Manufacturer and YearInterval

As shown in Figure 4-6, I did not see a strong relationship between the mean market shares and mean new model productions of mobile phones. In order to examine if both the 5 level categorical variables of *manufacturer* and the 5 level categorical variables of *YearInterval* had a relationship with continuous variable of *market share,* ANOVA was done. Also, *Phones* as the covariate added in the ANCOVA model. $\mathbf{R}^2$, the coefficient of determination, indicates 43.06%. Figure 4-6 shows that *Manufacturer*DPhones* is not statistically significant (F (4, 20) = 0.79, ρ = 0.556). From this analysis, it can be concluded that *Manufacturer and YearInterval* as random factors are not statistically significant.

Factor	Type	Levels	Values
Manufacturer	random	5	LG, Motorola, Nokia, Samsung, Sony-Ericsson

Analysis of Variance for DMkt Share, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
DPhones	1	2.663	0.233	0.233	0.04	0.853
Manufacturer	4	22.113	21.001	5.250	0.81	0.540
*Manufacturer**DPhones*	4	30.080	30.080	7.520	1.15	0.369
Error	15	97.697	97.697	6.513		
Total	24	152.554				

S = 2.55209   R-Sq = 35.96%   R-Sq(adj) = 0.00%

Term	Coef	SE Coef	T	P
Constant	0.4649	0.5935	0.78	0.446
DPhones	-0.01084	0.05737	-0.19	0.853
*DPhones**Manufacturer*				
LG	0.0443	0.1023	0.43	0.671
Motorola	-0.12550	0.08083	-1.55	0.141
Nokia	0.1699	0.1167	1.46	0.166
Samsung	-0.06647	0.08664	-0.77	0.455

Figure 4-7: ANCOVA with Manufacturer

As shown in Figure 4-7, I did not see a strong relationship between the mean market shares and mean new model productions of mobile phone. In order to examine whether the 5 level categorical variables of *manufacturer* had a relationship with continuous variable of *market share*, ANOVA was done. Also, *Phones* as the covariate added in the ANCOVA model. $\mathbf{R^2}$, the coefficient of determination, indicates 35.96%. It means that the model is not very useful because it produces a value close to 0. However, Figure 4-7 shows that *Manufacturer*DPhones* is not statistically significant (F (4, 20) = 1.15, ρ = 0.369). From this analysis, it can be concluded that *Manufacturer* as random factor is not statistically significant.

4.2 Future Oriented Data Mining through a Survey

4.2.1 Subjects and Procedure

A total of 527 people (274 U.S citizens, 253 non-U.S citizens) participated in the survey. Subjects' ages ranged from 14 to 39. A total of 274 U.S citizens consisted of 142 males and 132 females, whereas 253 non-U.S citizens (125 males and 128 females) were primarily Asian and live in the United States. Asians who responded the survey were predominantly Korean.

Between March 15 and April 15, 2009, a survey was conducted both online and in-person to determine which were the most important factors and design characteristics for users of mobile phones. The survey respondents varied by age, gender, education level, and ethnicity. The objective of the study was to determine which mobile phone factors were the most important for each group of respondents and whether there were patterns in the general characteristics and design issues that were preferred by various groups or subgroups.

A survey with 11 questions was developed to ascertain user preferences for various mobile phone factors and design characteristics. The survey was posted online at the Korean Student Association's web site of, and responses from members of that association were received primarily from Penn State University, University of Michigan, University of Wisconsin, University of Minnesota, University of Illinois, and Ohio State University. Also, responses of high school students were collected mainly from Pioneer High School in Ann Arbor, Michigan. In addition, volunteers were recruited at these college campuses to "randomly" hand out the survey in person to various passersby, and were gathered from a Korean and American Church in Pennsylvania. Subjects completed the survey study individually.

The survey respondents were categorized by age, race, gender, and educational level, based on four survey questions, as follows:

Ethnicity: U.S citizens, non-U.S. citizens (Asian-predominately Korean)

Age: 14-17, 18-22, 23-29, 30-39

Gender: Male, Female

Education: High school student, Only High school diploma, Undergraduate student, Graduate students, Bachelor degree, Master degree, Ph.D. Degree

These are the independent categorical variables in the study. The first three variables (Ethnicity, age, and gender) result in 16 distinct demographic subgroups among all the survey respondents.

Survey respondents answered seven questions related to what factors and design characteristics of mobile phones they deemed to be most important. This study is interested in focusing on responses to only 3 of these questions, which resulted in the following dependent categorical response variables:

What factor do you consider as the most important when purchasing a mobile phone?

Purchasing factors: Design (1), Price (2), Functions (3), Brand (4), Easy user interface (5)

What is the most significant design issue in mobile phones in your opinion?

Design Issues: Weight (1), body color (2), keypad design (3), screen size (4), form (5)

What mobile phone form do you prefer?

Form: Clamshell folder (1), block (2), slider (3), flip (4)

For the first two categories (purchasing factors and design issues) questions, each survey respondent indicated the answer which he or she felt was the most important features for a mobile phone from the given choices in that category. For the third category question, “Form”, user indicated which form they preferred. For example, if a respondent felt that price was the most important of five given factors when purchasing a mobile phone, the respondent chose (2) as the answer for the factor question; if the respondent felt that the weight was the most important design issue of the five given design issues, the respondent chose (1) as the answer for the design issues question. This survey study is to evaluate which mobile phone general factors and design features are most important for the 16 different demographic subgroups defined by origin, gender, and age. Then patterns of preference toward forms can be identified within each subgroup.

4.2.2 Factors to Affect Purchasing Pattern

Figure 4-8 shows that design, price, function, and easy user interface are all significant considerations for every American age group. Americans in all age groups are more likely to be influenced by functions than by other factors. Americans, 23-29 years old, are more likely to be motivated by functions and easy user interface. This might reflect the different ways in which adults and younger teenagers adjust to a changing technological environment and follow modern trends.

As shown on Figure 4-8, design, price, function, brand, and easy user interface are important motivations for all Asian age group that reside in the US. The price of mobile phones currently occupies a considerable importance for all Asian age groups. Function and price are bigger priorities for Asians 30-39 age groups. Asian females, 14-29 age groups, are more likely to be motivated by design, while Asian males, 18-39 age groups, are more likely to be influenced by functions. Asian males, 23-29 years old, prefer to choose mobile phones based on brand. Through data analysis, I can identify that Asian males and females have different preferences for selecting their mobile phones.

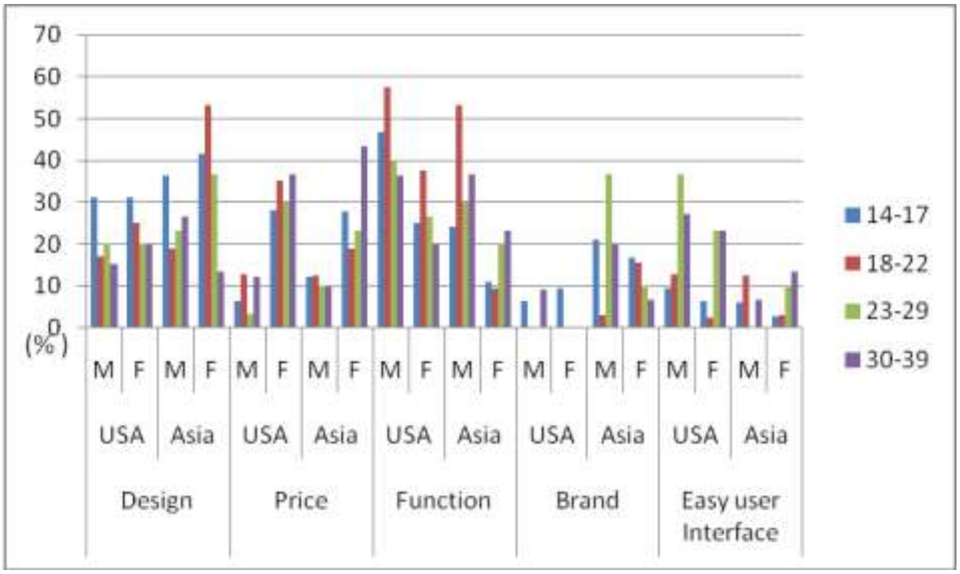

Figure 4-8: Factors to Affect Purchasing Pattern of Mobile Phone

4.2.3 Design Issue

Figure 4-9 indicates that a mobile phone's weight, keypad, screen size, and form are important design issues for all American age groups. Body color and texture are not considered as significant design issues for Americans. For all American age groups, respondents have a greater interest in keypad design. Americans 14-22 years old are more likely to be motivated by design issues such as keypad, weight, form style, while keypad, screen size, and form style are larger consideration for 23-39 age groups. Keypad and form design occupies a significant proportion of all age groups.

According to Figure 4-9, all Asian age groups are concerned with weight, body color, screen size, and form when deciding on their mobile phone purchase. Asians from age 14-29 years old are more likely to be influenced by form styles that indicate their personal identities. Asian females, 30-39 years old, are more likely to be motivated by keypad design than form design. Asians females, 18-39 age groups put body color as their priority as compared with Asian males. Younger Asian age groups tend to concentrate on light weight as the most important design issue. Asian males from age 30 and 39 do not highly consider light weight of mobile phones, while they do focus on bigger screen size and form types. These data results explain that Asians from 14-29 years old are the mostly likely to consider form styles.

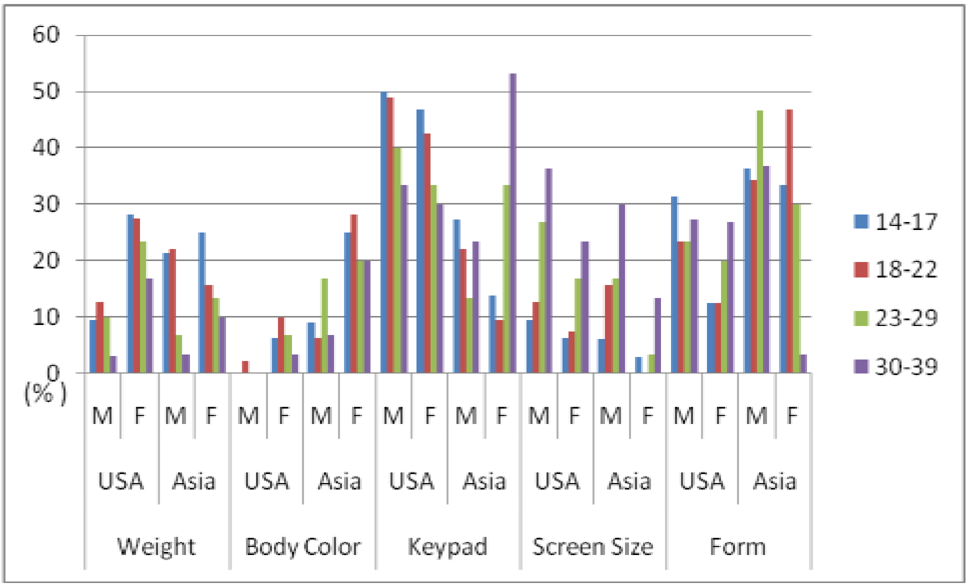

Figure 4-9: Design Issue of Mobile Phone

4.2.4 Form Style

According Figure 4-10, preference for form style is divided by different American age groups. American females, 14-29 years old, are more likely to purchase a block style phone among various form styles. American males, 14-17 and 23-29 age groups, are more likely to be motivated by slider style in purchasing a mobile phone. Americans from 30-39 years old are more likely to be influenced by folder style. Due to the emerging Smartphone industry, consumers' preferences toward form styles reflect the variety of options.

Figure 4-10 signifies that slider form is a significant preference style in consumers' decision making process for all Asian age groups. On the other hand, Asian males, 14-17 age groups are more likely to be influenced by clamshell folder and slider form. The preference of folder style is decreased for this demographic. Slider form prevents common malfunctions, especially unintentional button pushing, and Asian consumers are used to using various slider phones which are mainly produced by Asian manufacturers such as LG and Samsung. Unlike folder and block style, Asians prefer the slider form.

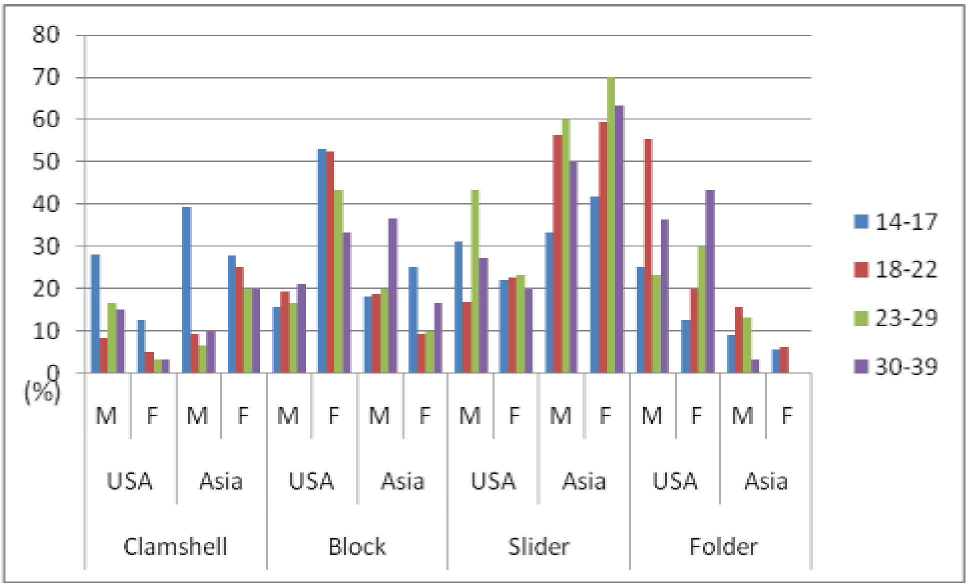

Figure 4-10: Form Style of Mobile Phone

4.2.5 Essential Functions

Table 4-11 displays the most essential five functions of mobile phone which have been identified by American teens and adults. About 65.6 % of the 14-17 American age groups consider text messaging as one of the most important mobile phone functions. Approximately 68.3 % of Americans between the ages of 30 and 39 regard battery life as an essential function of mobile phones. Besides, all American age groups consider camera abilities to be one of the important functions. This indicates that mobile phone manufacturers should think about what the primary function is preferred by different age groups.

Table 4-11 indicates that the most essential five functions of a mobile phone are identified by Asian teenagers and adults. About 63.8 % of Asians between the ages of 14 and 17 claim that that text messaging is primary function, while approximately 58.3 % of 30-39 Asian age groups consider battery life as the dominant function. It is interesting that 23-29 Asian age groups do not consider text messaging as one of essential functions. I can therefore distinguish differences in essential functions within Asians and Americans.

Table 4-11: Essential Functions of Mobile Phone

Functions Ranking \ age	14-17		18-22		23-29		30-39	
	USA (%)	Asia (%)	USA (%)	Asia (%)	USA (%)	Asia (%)	USA (%)	Asia (%)
1	Text (65.6)	Text (63.8)	Text (56.3)	Text (45.3)	Battery life (45.0)	Battery Life (43.3)	Battery life (68.3)	Battery Life (58.3)
2	Camera (25.0)	Camera (30.4)	Battery life (40.2)	Battery Life (29.7)	Text (23.3)	Camera (21.7)	Text (20.6)	Camera (31.7)
3	Battery Life (25.0)	Mp3 player (26.1)	Camera (24.1)	Camera (20.3)	Camera (18.3)	Mp3 player (20.0)	Camera (19.0)	Sensitive sensor (18.3)
4	Sensitive sensor (23.4)	Game (18.8)	Memory Capacity (20.7)	Mp3 player (18.8)	Sensitive sensor (15.0)	Sensitive sensor (18.3)	Memory Capacity (19.0)	Text (16.7)
5	Mp3 player (18.8)	Battery Life (17.4)	Game (19.5)	Memory Capacity (14.1)	Memory Capacity (15.0)	Memory Capacity (16.7)	Call log memory (15.9)	Memory Capacity (16.7)

4.2.6 Consumers' Function Preferences for Future Purchase

Mobile phone manufacturers focus on emphasizing specific functions based on consumers' preferences in the near future. Figure 4-11 suggests one way of identifying functions that consumers would like to purchase the most for the future. American teenagers and adults, of all age groups, are the most likely to purchase mobile phone which emphasizes sensitive functions such as touch screen or a QWERTY keypad. Improved gaming function is not an important factor for any American or Asian age groups. Especially American teenagers from age 14 and 17 want to a purchase mobile phone that develop instant messenger, cam messenger, picture messaging, or text messaging. These advanced functions reflect communications tools which may enlarge social networking with their friends. On the other hand, Americans of 30-39 age groups are more likely to purchase user interface phone that can handle menu organization easily. This reflects that they do not want to spend much time learning to use their new advanced technology. American males of 14-29 age groups prefer an advanced Mp3 player function more than American females. Moreover, American females of all age groups are more likely to purchase mobile phone that includes a digital camera with high resolution than American males.

According to Figure 4-11, Asians tend to indicate similar preferences for functions as Americans when considering a future purchase. However, Asian adults of 18-39 years old are more likely to purchase mobile phone that has sensitive touch screen or QWERTY keypad than instant messenger, cam messenger, picture messaging, or text messaging. Asian females of all age groups prefer to include digital camera function of high quality. The relative importance of easy user interface menu grows bigger for Asian adults of 23-39 years old.

There is a relative difference between age groups 14-17 and 18-39 for Asians. Even though many of the teenagers were born in Asia, they have spent their childhood in United States of America. They experience American culture, environment, and social network in a different way than the older age groups. Oppositely, Asian adults of 18-39 age groups have been raised in their own county, and they may have stronger ties to their Asian cultural background, habits, and attitudes. Some do not have the opportunity to network socially with American friends because they came to study abroad. If Asian people who live in their own country were investigated, the data results regarding consumer preferences may be changed. Data results demonstrate that future’s preference functions represents the difference by race, culture, background environment.

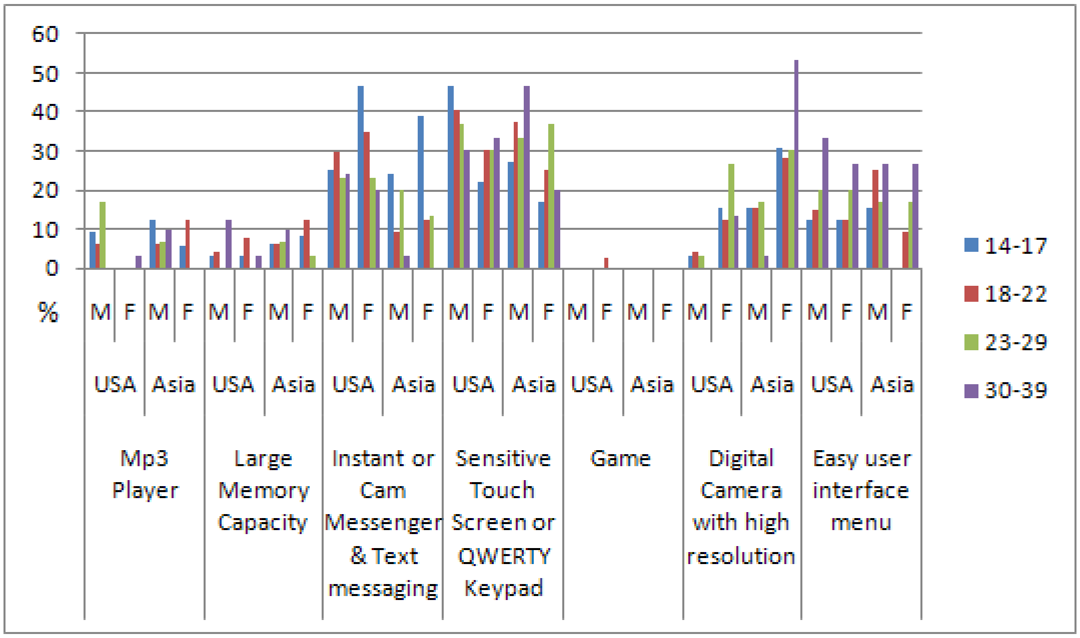

Figure 4-11: Consumers' Function Preferences for Future Purchase

4.2.7 Built in Functions that Consumers Use the Most

According to Figure 4-12, functions that consumers use the most are identified by all of American age groups. About 92.2 %, 14-17 American age groups are most likely to use text messaging. Approximately 79.3%, college age groups rank texting as their first priority. On the other side, about 61.9 %, 30-39 American age groups are most likely to primarily use the phone calling function. None of the American age groups are mainly using their mobile phones as Mp3 players, games, or cameras. However, approximately 16 %, American males from 23-29 age groups are mostly using e-mail function for business purpose. Ultimately, text messaging is the most frequently used function for all of American age groups.

In Figure 4-12, the most used functions are identified by all of Asian age groups. 14-22 Asian age groups are the most likely to use text messaging. On the other hand, Asians aged 23-39 are the most likely to use phone calls. About 93.3 %, Asians from age 30 and 39 are mainly using phone call. Asian respondents are mostly students, and they prefer to send E-mail by using computer. The E-mail is not a frequently used function for all of Asian age groups. Mobile phone manufacturers need to do more research investments about both phone calls and text messaging.

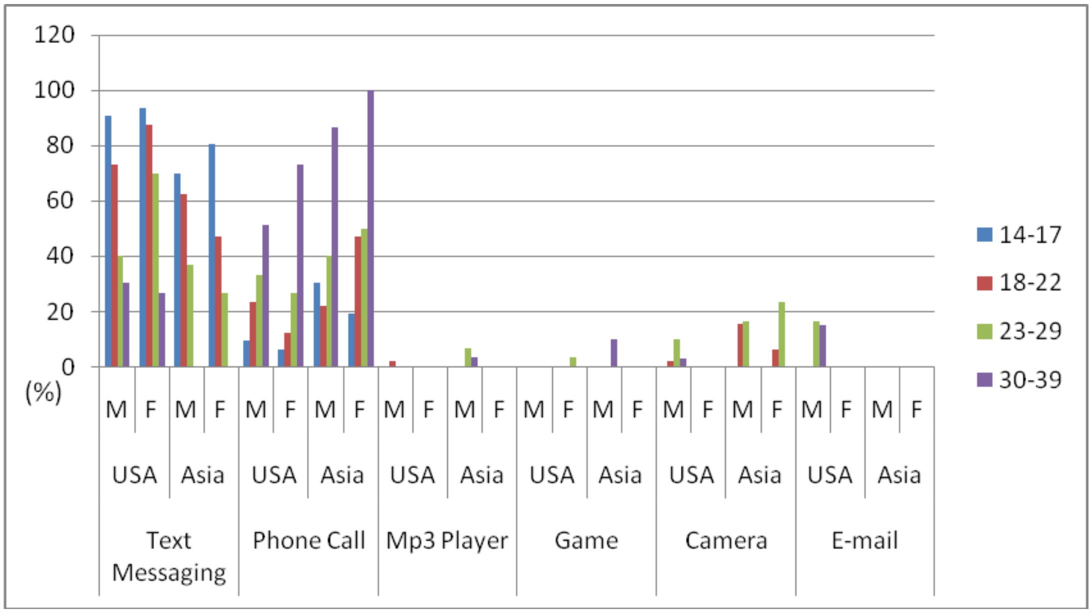

Figure 4-12: The Most Used Function

4.2.8 Frequencies of Each Function Use

Text messaging usage frequency per day for the 14-17 age groups is higher than other age groups. In Figure 4-13, 72 % Americans and 65.2 % Asians between 14-17 are using text messaging over 20 times per day. Americans and Asians from age 18 and 22 still show high text messaging frequency. On the other hand, 66.7 % Americans and 81.7 % Asians of 30-39 age groups are using text messaging less than 6 times per day. If mobile phone manufacturer focus on the teenagers as a main target, they should develop efficient functions related to text messaging usage.

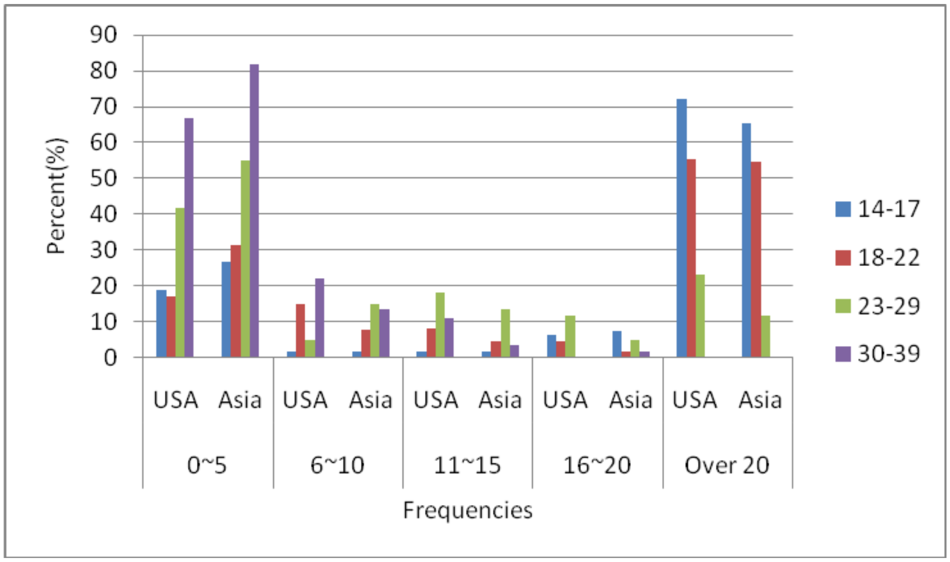

Figure 4-13: Text Messaging Usage Frequency per Day

Phone call frequency per day is different for Americans and Asians. In Figure 4-14, American and Asian teens are mainly using phone calls about 6-10 times every day. About 60.3 %, 30-39 American age groups are mainly using the phone to call less than 6 times every day, while Asians are mostly using the phone to call about 6-10 times per day. Besides, approximately 20.3 %, Asian college age groups are using the phone to call over 20 times every day, while Americans prefer text messaging. Asian phone call frequency per day is higher than American frequency

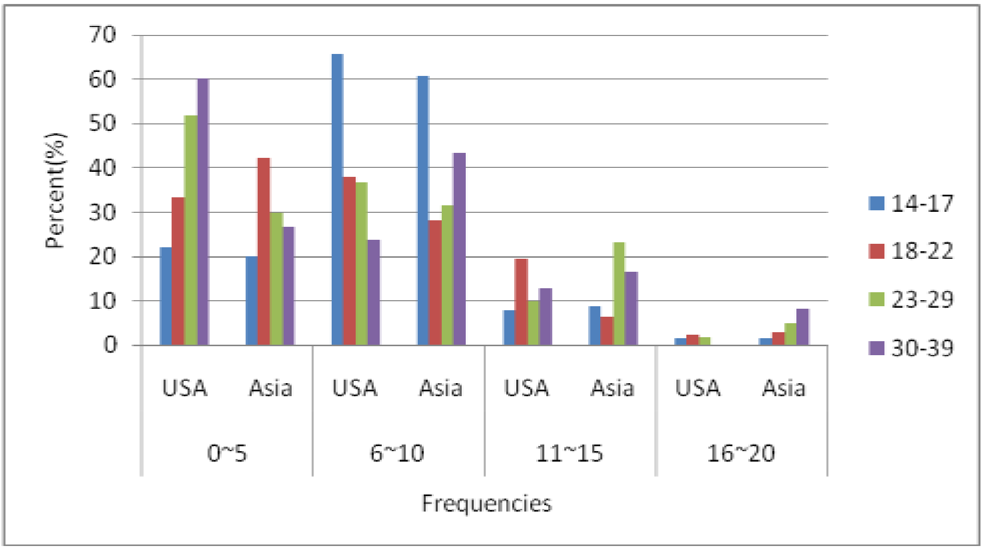

Figure 4-14: Phone Call Using Frequency per Day

Instant messenger using frequency per day is lower for Americans and Asians. Figure 4-15 indicates that all American and Asian age groups are rarely using instant messenger about 90 %. About 10.9 %, 18-22 Asian age groups are only using instant messenger about 6-9 time per day. In addition, about 9.3 %, 23-29 American age groups are simply using instant messenger about 6-9 times per day. American and Asian are familiar with accessing instant messenger by using a computer.

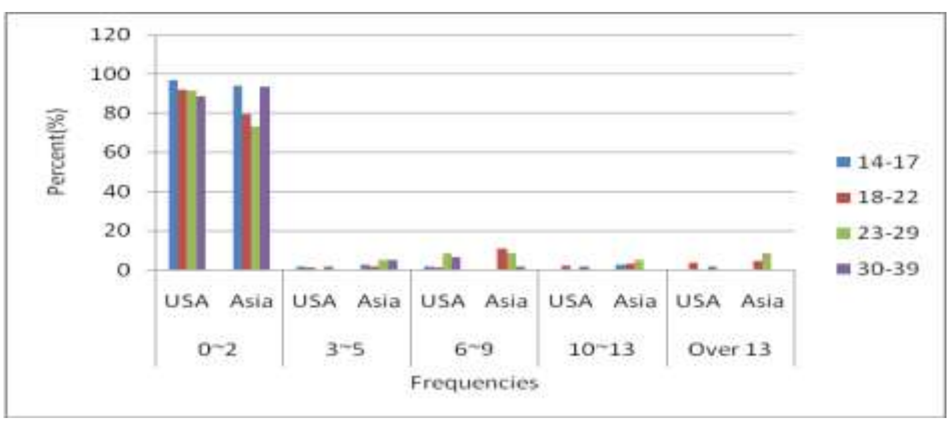

Figure 4-15: Instant Messenger Using Frequency per Day

Camera using frequency in Americans is lower. Figure 4-16 shows that approximately 70 % Americans are not using camera function of mobile phone, and they prefer to purchase their own digital camera. About 24 %, American teenagers and college age groups are using camera function about 3-5 times per day. On the other hand, all Asian age groups are familiar with camera function of mobile phone. About 26.7 %, 23-29 Asian age groups are using camera function about 6-9 times every day. From these data, Asian people are more likely to be influenced by mobile phones that include the camera function.

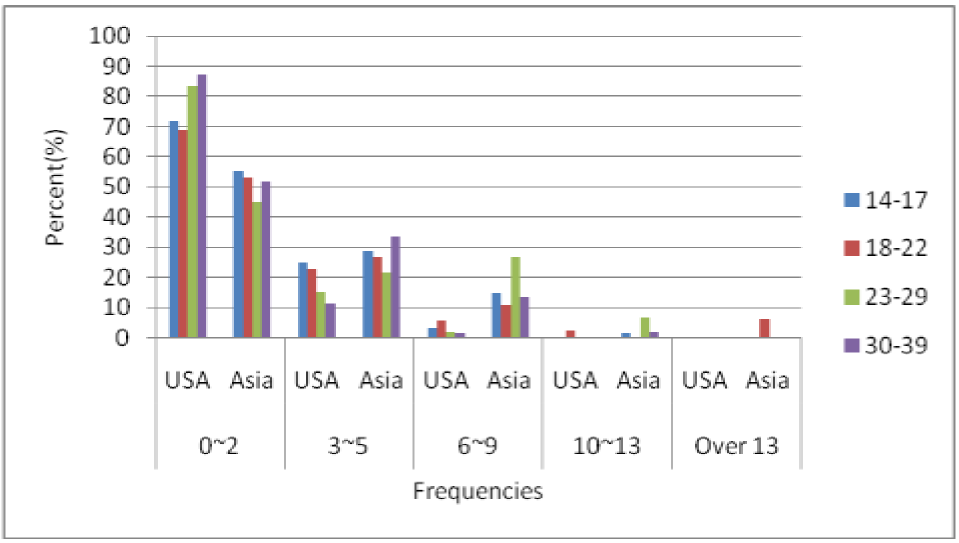

Figure 4-16: Camera Using Frequency per Day

Mp3 player usage frequency per day is lower for Americans. Figure 4-17 indicates that 14.1 % of 14-17 American age groups listen to music by using an Mp3 player function about 6-9 times every day. About 11.7 %, 23-29 American age groups are using an Mp3 player function about 3-5 times per day, while 30 % of 23-29 Asian age groups are using that. Americans from age 14 and 22 prefer to have the Mp3 function than Asians. Asians, 23-29 years old, are frequently using Mp3 player function, more than Americans.

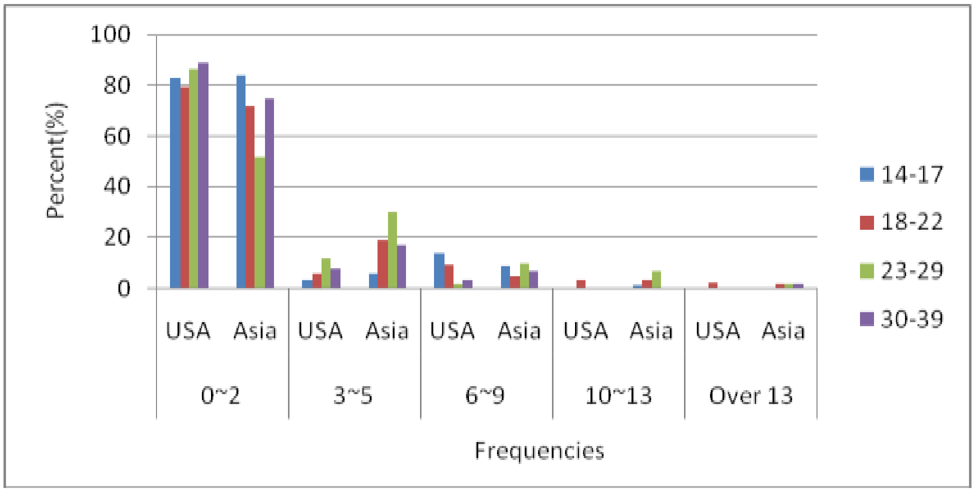

Figure 4-17: MP3 Player Usage Frequency per Day

Game usage frequency per day is different between various age groups and races. According to the Figure 4-18, 35.9 % of 14-17 American age groups are enjoying game function of mobile phone about 3-5 times every day. However, 92.1 % of 30-39 American age groups are using a game function less than 3 times per day as Americans growing older might tend to focus on their job, and hence business related usage. On the other hand, game usage frequency of 18-22 Asian age group varies. Through survey results, I can identify that Asians are generally using mobile phone game.

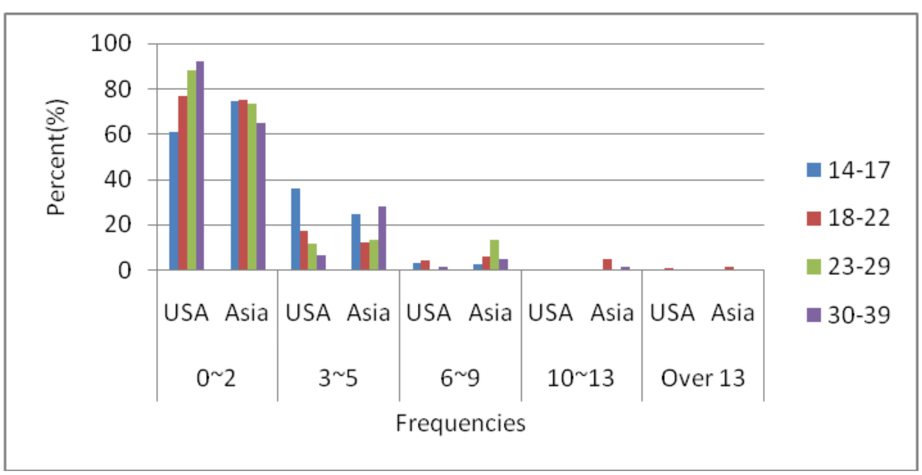

Figure 4-18: Game Usage Frequency per Day

23-29 American and Asian age groups show higher E-mail usage frequency per day than 14-17 age groups. According to the Figure 4-19, 15 % of 23-29 American age groups are using E-mail over 13 times every day. Moreover, 12.7 % of 30-39American age groups are using E-mail about 6-9 times per day. Based on these data, very few Americans and Asians are using the E-mail function for business and networking purpose.

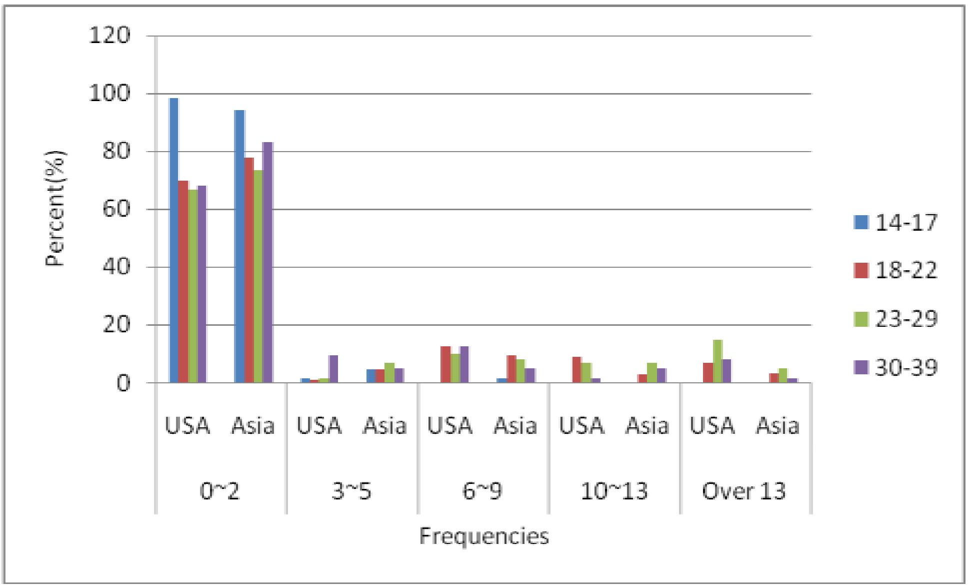

Figure 4-19: E-mail Usage Frequency per Day

In Figure 4-20, multimedia usage frequency per day is lower for Americans and Asians. More than 80 %, all American age groups does not play multimedia function. Only a few 14-17 American age groups and college age groups are most likely to use multimedia function about 3-5 times per day. Also, 18-29 Asian age group indicates various frequency distributions about using the multimedia function. I identify that younger Americans and Asians have better opportunity to access mobile multimedia information.

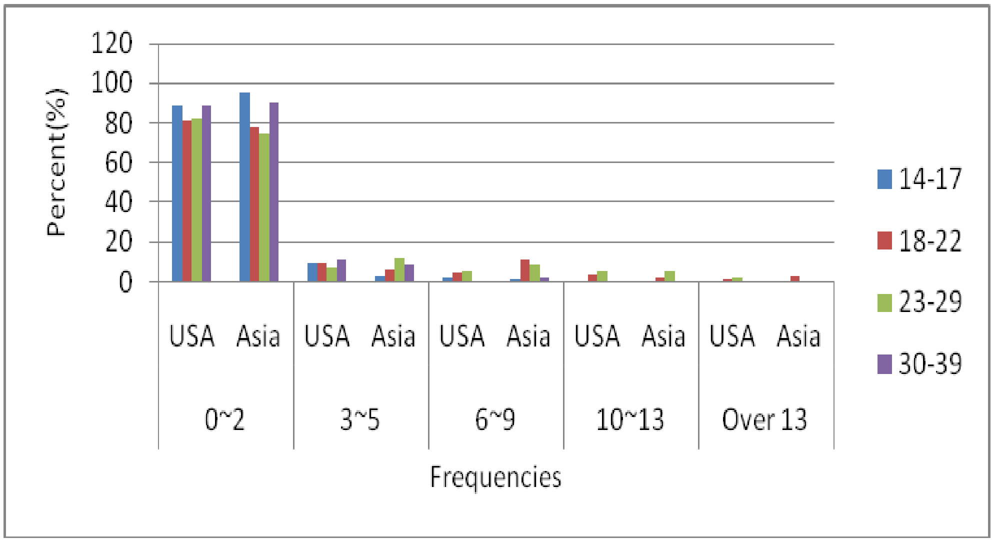

Figure 4-20: Multimedia Usage Frequency per Day

4.2.9 Consumers' Brand Preference

In Figure 4-21, Americans and Asians have different brand preference. Americans and Asians from ages 14 and 22 are more likely to purchase LG mobile phone brand. 23-39 Americans and Asians are more likely to be influenced by Samsung brand. Besides, even if market share of Nokia manufacture is the first ranking in the worldwide, Americans and Asians who live in the United State do not prefer to Nokia brand.

In the United State, Apple and Blackberry brand preferences are just behind Samsung, LG, and Motorola grows quickly because of its sensitivity, simple forms, and functions such as virtual or QWERTY keypad. Blackberry has a relatively large screen and easy keyboard. As a result, it helps to send emails simply. Blackberry brand shows high appeal to business professionals. Based on survey data, consumers tend toward specific brand preference about mobile phone manufacturers.

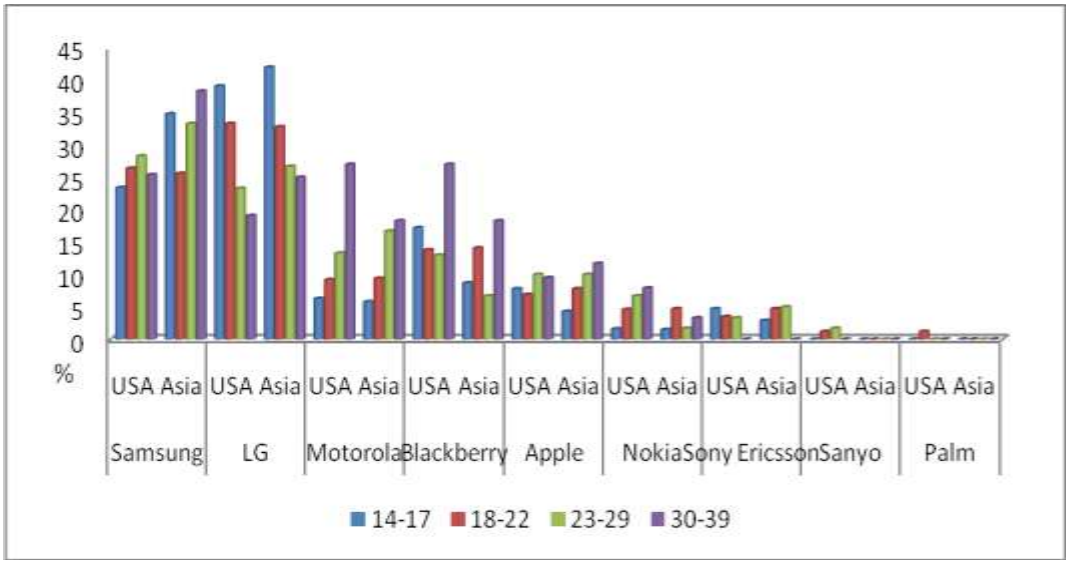

Figure 4-21: Brand Preference

4.2.10 Consumers' Wireless Preference

The top mobile phone manufacturers have a partnership with wireless companies such as Verizon, AT&T, Sprint, and T-mobile. In Figure 4-22, about 53 %, 30-39 Asian age groups are contracting with AT&T wireless. More than 35 %, all Asian age groups are mostly using AT&T wireless service, while all Americans are mainly using AT&T and Verizon wireless service. Moreover, all Asian age groups consider leaning toward T-mobile wireless company more than Americans. In the end, consumers' wireless preferences will be changed by determining their preference mobile phone. For example, Apple I-phone has a partnership with AT&T wireless company. Accordingly, if Apple I-phones' sales are increased, consumers using AT&T wireless service will be increased.

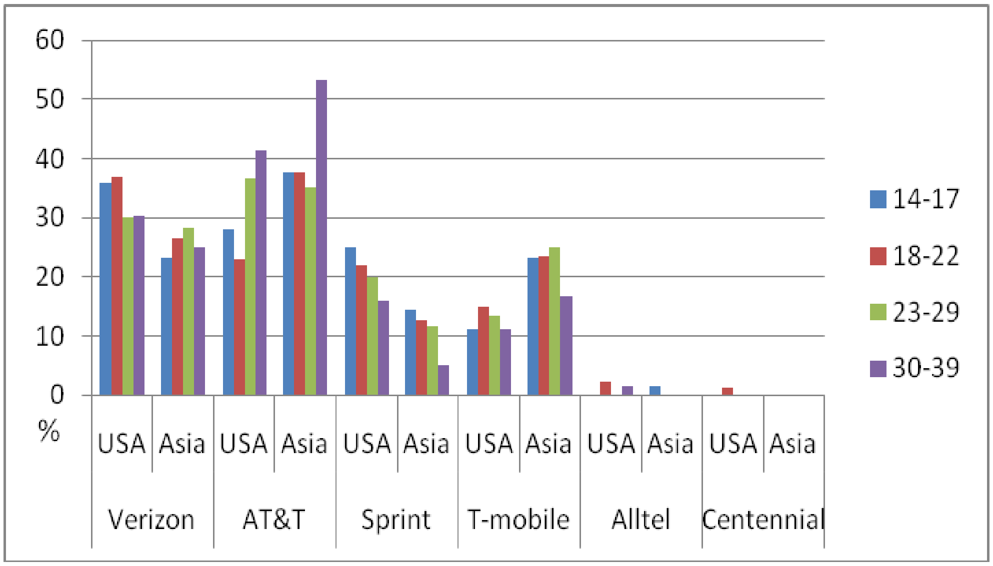

Figure 4-22: Wireless Company Preference

4.3 Exploratory Data Research

4.3.1 Independent Variables

Before evaluating the response variables in a survey study, it is critical to evaluate the breakdown of independent variables that define the survey respondents. In this first study, understanding the distribution of the origin, gender, and age variables is important for properly interpreting the frequency of the response and for determining whether the data can be accurately evaluated using nominal logistic regression.

Table 4-12: Tally of Independent Categorical Variables

Origin	Count	Percent	Gender	Count	Percent	Age	Count	Percent
Asia	255	48.39	Female	261	49.53	14-17	133	25.24
USA	272	51.61	Male	266	50.47	18-22	150	28.46
						23-29	121	22.96
						30-39	123	23.34

The survey had a total of N = 527 respondents. Generally, each variable shows a roughly even distribution of respondents within each category. In Table 4-12, the percentage of US respondents was slightly over 51% and non-US Asians was nearly 49%. The percentage of males and females was also almost the same (about 50%). The percentage of respondents in each age group ranged from approximately 23%-28% of the total respondents, with most respondents in the 18-22 age group. To further break down the survey respondents by subgroup, I can display a frequency table and clustered bar charts.

Rows: Origin / Gender Columns: Age

Origin	Gender	14-17	18-22	23-29	30-39	All
Asia						
	F	37	32	31	30	130
		28.46	24.62	23.85	23.08	100.00
		27.82	21.33	25.62	24.39	24.67
		7.021	6.072	5.882	5.693	24.668
	M	32	32	31	30	125
		25.60	25.60	24.80	24.00	100.00
		24.06	21.33	25.62	24.39	23.72
		6.072	6.072	5.882	5.693	23.719
USA						
	F	32	40	29	30	131
		24.43	30.53	22.14	22.90	100.00
		24.06	26.67	23.97	24.39	24.86
		6.072	7.590	5.503	5.693	24.858
	M	32	46	30	33	141
		22.70	32.62	21.28	23.40	100.00
		24.06	30.67	24.79	26.83	26.76
		6.072	8.729	5.693	6.262	26.755
All						
	All	133	150	121	123	527
		25.24	28.46	22.96	23.34	100.00
		100.00	100.00	100.00	100.00	100.00
		25.237	28.463	22.960	23.340	100.000

Cell Contents: Count
% of Row
% of Column
% of Total

Figure 4-23: Frequency of Independent Categorical Variables

Figure 4-23 shows the frequency statistics for each of the 16 demographic subgroups in the data. For example, the four values in the upper left of the table show the following: 37 respondents (the count value) were non-US Asian females in the 14-17 age group; approximately 28.46% of all Asian females in the survey were in the 14-17 age group (the % of Row value); approximately 27.82% of all the 14-17 age group were non-US Asian females (the % of Column value); non-US Asian females accounted for about 7.02% of the total respondents (% of Total value). Figure 4-23 also shows that the group with the highest frequency was US males between the ages of 18-22, who comprised about approximately 8.7% of the total respondents; the group with the lowest frequency was US females between the ages of 23-29, who comprised about 5.5% of the total respondents.

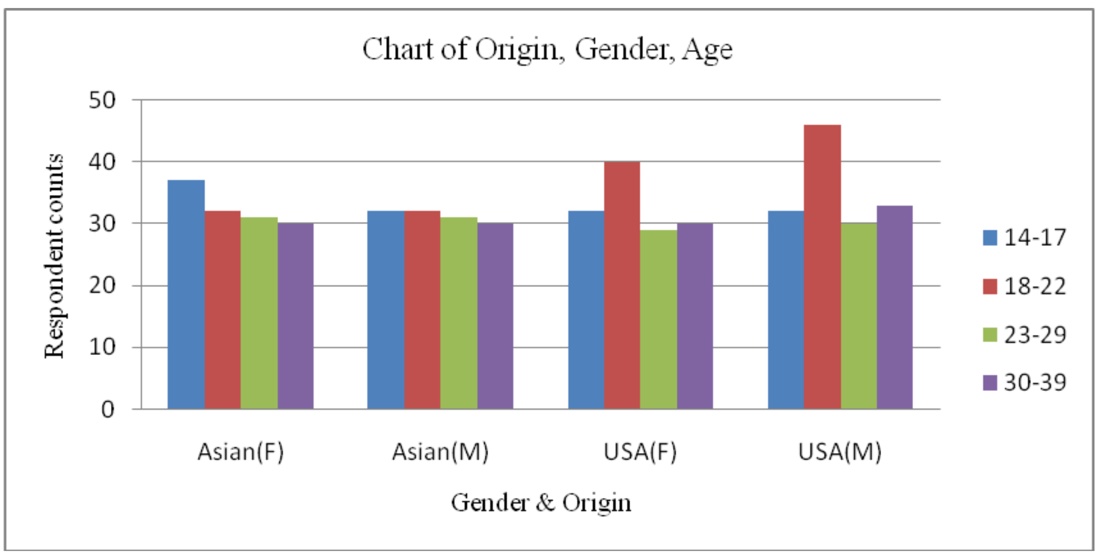

Figure 4-24: Bar Chart of Survey Respondents by Origin, Gender, and Age

Figure 4-24 indicates that each group has roughly the same amount of individuals and no group has fewer than 5 individuals. This will be important for satisfying assumptions when I evaluate the response variables using logistic regression. The bar chart does show that US citizens in the 18-22-year-old age group account for the largest share of the survey respondents. We should therefore be careful when interpreting response results and remember that this group is "over-represented" in the survey samples, and thus has the strongest influence on the overall results.

Although the bar chart results do not suggest that any of the independent variables (Origin, Gender, Age) are associated with each other (i.e., they are not dependent), I can perform a chi-square test of association to statistically determine this. For example, because so many US respondents are in the 18-22 age groups, I might wonder whether there is an association between Origin and Age in the survey data. If there is, these two variables may not be appropriate to use together as independent variables in a logistic regression model, as multicollinearity might result.

```
Rows: Origin   Columns: Age

            14-17    18-22    23-29    30-39      All

 Asia          69       64       62       60      255
            64.35    72.58    58.55    59.52   255.00

 USA           64       86       59       63      272
            68.65    77.42    62.45    63.48   272.00

 All          133      150      121      123      527
           133.00   150.00   121.00   123.00   527.00

 Cell Contents:       Count
                      Expected count

 Pearson Chi-Square = 3.017, DF = 3, P-Value = 0.389
 Likelihood Ratio Chi-Square = 3.025, DF = 3, P-Value = 0.388
```

Figure 4-25: Chi-Square Test for Origin and Age

In Figure 4-25, this result shows the actual count and expected count for each variable value. For example, the actual count of 18-22 year olds of Asian origin was 64 and the expected count was about 72. This means the number of Asians in the survey appears to be less than expected if Age and Origin are independent. Conversely, the actual count of 18-22 year olds of

US origin as 86, which is more than the expected count (for independent variables) of about 78. For the other groups, however, the expected and actual counts are fairly close.

To determine the overall chi-square test results look at the p-values. If the p-values are greater than 0.05, I conclude that the variables are independent. Because $p = 0.389$, I do not have evidence that Origin and Age are associated with each other. Thus, the variables are independent and can be used as independent predictors in the logistic regression model. In a similar manner, I can also perform the chi-square test on Origin vs. Gender, and Gender vs. Age to evaluate the independence of these predictors.

4.3.2 Dependent Variables

After I have examined the distributions and relationships between the independent variables, I can use exploratory analysis to evaluate the survey responses for Factor, Design Issues, and Form. First, I focus the analysis on the Factor variable, in which respondents chose the factor that they considered most important when purchasing a mobile phone: 1) Design 2) Price 3) Functions 4) Brand or 5) Easy User Interface.

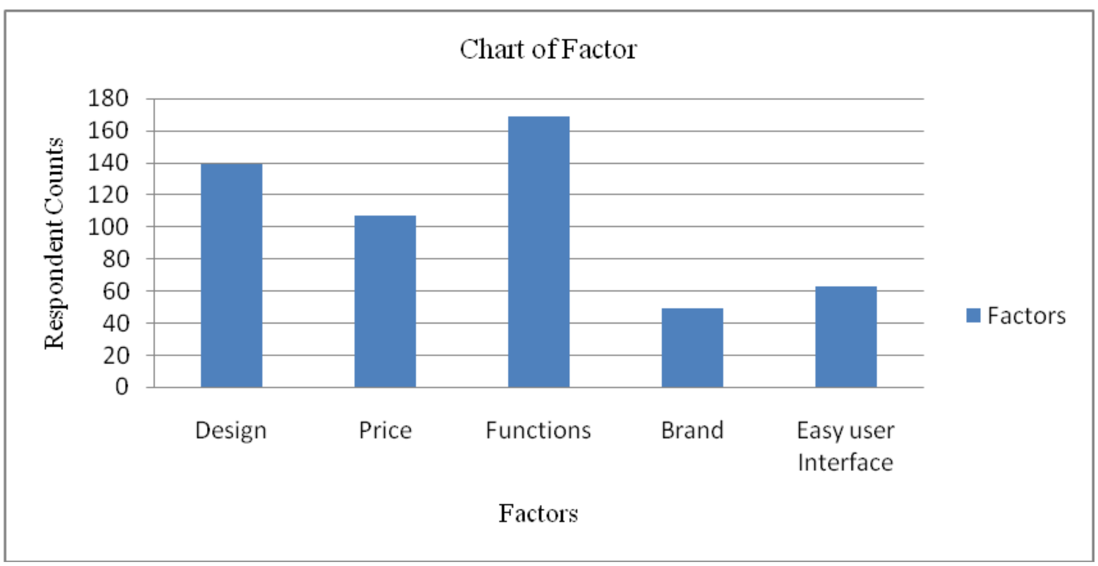

Figure 4-26: Bar Chart of Factor Response

The Figure 4-26 shows that most respondents cited Functions (3) as the most important factor when purchasing a mobile phone. Design (1) and Price were the second and third most frequent responses, respectively. Brand (4) and Easy User Interface (5) were the least cited factors.

By displaying a clustered bar chart that subdivides the results using one or more independent variables, I can learn more about how the Factor response varied among the individual groups.

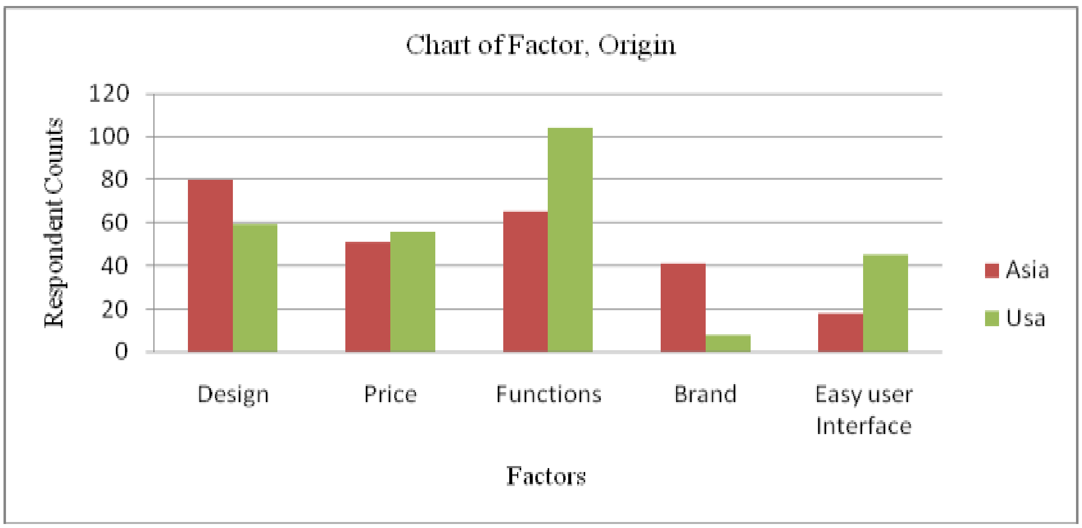

Figure 4-27: Bar Chart of Factor by Origin

The Figure 4-27 suggests that more US respondents than Asian respondents rated Functions and Easy User Interface as the most important factor for a mobile phone; more Asians than US respondents rated Design and Brand as the most important factor. Price was rated the most important factor by a nearly equal number of Asian and US respondents.

The bar chart in Figure 4-28 shows the Factor response by Age group, and is created in the similar manner as the one in Figure 4-27.

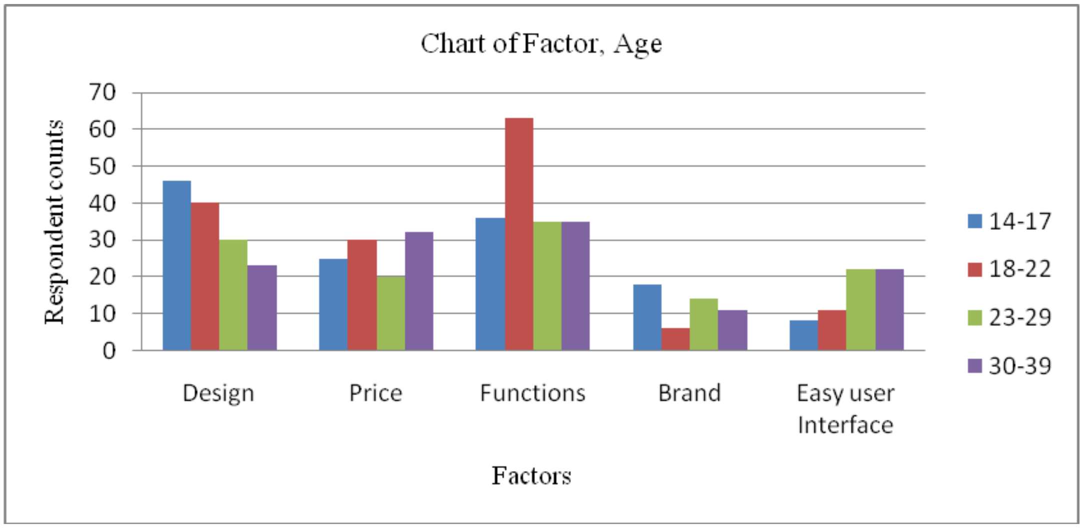

Figure 4-28: Bar Chart of Factor by Age

Clusters of bars that are either ascending or descending suggest a possible trend in the data. For example, notice that as the age group increases, the number of respondents who rate Design (1) as the most important factor generally decreases, while the number of respondents who rate Easy User Interface (5) as the most important factor generally increases. Of the respondents who rated Functions (3) as the most important factor, the greatest number of respondents was in the 18-22 age groups. Recall, however, that the number of survey respondents in that age group was greater than in the others; therefore, any trends that show greater responses for that group (or, conversely, fewer responses among groups with relatively lower frequency representation in the survey) must be interpreted with caution. I can explore other possible trends and relationships by displaying bar charts based on other combinations of other independent variables and response variables, such as Gender and Design, Gender and Form, etc.

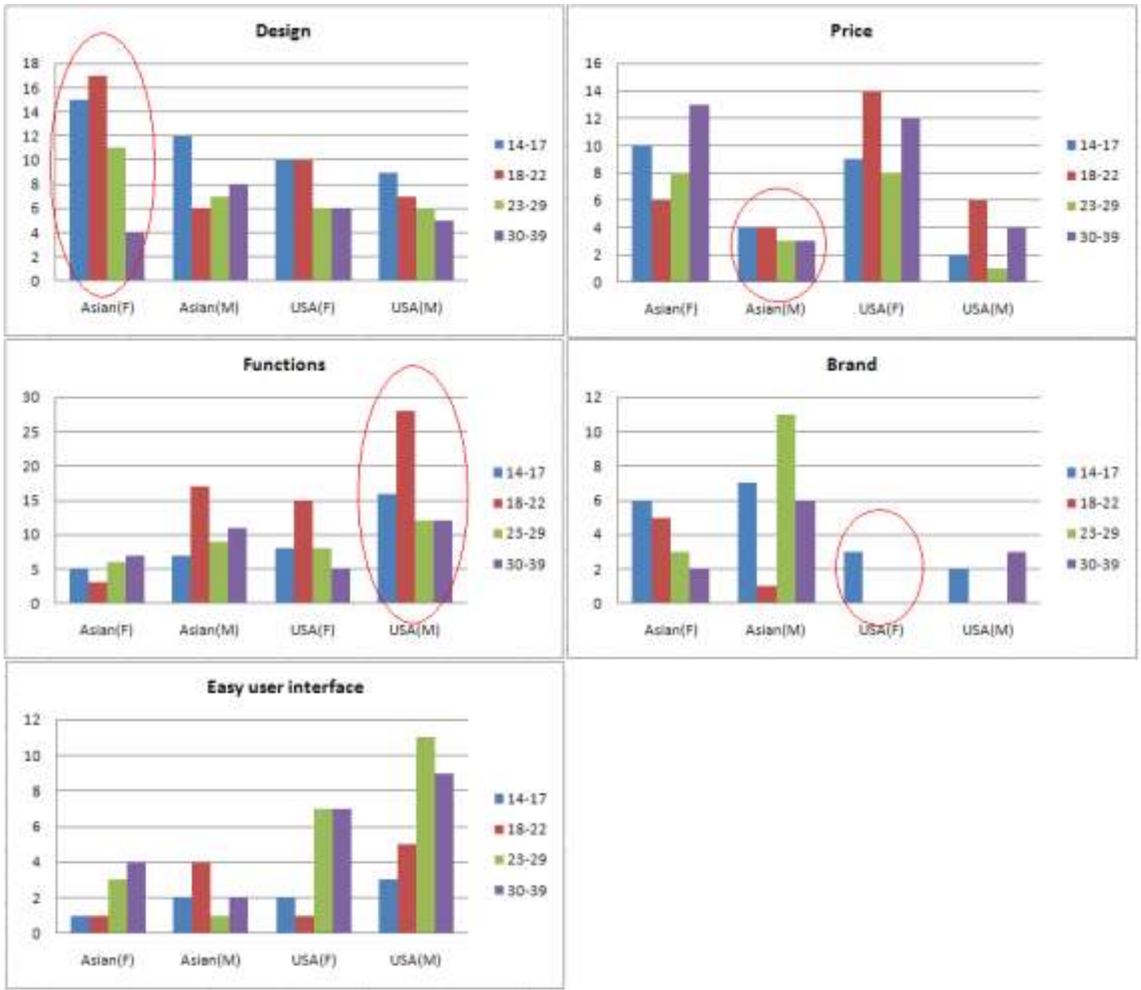

Figure 4-29: Bar Chart of Factor by Origin, Gender, and Age

Figure 4-29 shows that no female US respondents in age groups 18-22, 23-29, and 30-39 rated Brand (4) as the most important factor. Few Asian males cited Price (2) as the most important factor, and this was fairly consistent across all four age groups. Within each gender and origin, age group seems to produce the biggest disparity in response among Asian females who cited Design (1) as the most important factor, and among US males who cited Functions (3) as the most important factor. This bar chart also emphasizes that the first three factors, Design (1), Price (2), and Functions (3) are the most important.

4.3.3 Chi-Square Test of Association

Chi-Square test is to investigate the possible relationships between independent variables (Origin, Gender, and Age) and response variables (Factor, Design, and Form). I use a chi-square test of association to determine whether each pair of independent and/or dependent variables is statistically associated, or whether they are independent (Greenwood and Nikulin, 1996). The chi-square test helps set up the logistic regression model.

Figure 4-30, 4-31, and 4-32 indicate that chi-square test determine whether the independent variable Age, Origin, or Gender is associated with the response variable Factor. If the two variables are associated, this suggests that Age, Origin, or Gender will be a useful predictor in the logistic regression model with Factor as a response. At the 0.05 level of significance, the estimated p-value (0.000) suggests that predictors (Age, Origin, or Gender) and response (Factor) variables are dependent; thus, Age, Origin, or Gender is likely to be statistically significant predictors for Factor in a logistic regression model. Figure 4-30, 4-31, and 4-32 perform chi-Square method on other pairs of predictors and responses to make preliminary evaluations of what terms may likely to produce significant results in a logistic regression model for each response. There is a statistically significant association between predictors and responses. All p values are less than 0.05.

Tabulated statistics: Age, Factor

Rows: Age   Columns: Factor

	1	2	3	4	5	All
14-17	46	25	36	18	8	133
	35.08	27.00	42.65	12.37	15.90	133.00
18-22	40	30	63	6	11	150
	39.56	30.46	48.10	13.95	17.93	150.00
23-29	30	20	35	14	22	121
	31.91	24.57	38.80	11.25	14.46	121.00
30-39	23	32	35	11	22	123
	32.44	24.97	39.44	11.44	14.70	123.00
All	139	107	169	49	63	527
	139.00	107.00	169.00	49.00	63.00	527.00

Pearson Chi-Square = 37.706, DF = 12, P-Value = 0.000
Likelihood Ratio Chi-Square = 38.456, DF = 12, P-Value = 0.000

Tabulated statistics: Gender, Factor

Rows: Gender   Columns: Factor

	1	2	3	4	5	All
F	79	80	57	19	26	261
	68.84	52.99	83.70	24.27	31.20	261.00
M	60	27	112	30	37	266
	70.16	54.01	85.30	24.73	31.80	266.00
All	139	107	169	49	63	527
	139.00	107.00	169.00	49.00	63.00	527.00

Pearson Chi-Square = 51.096, DF = 4, P-Value = 0.000
Likelihood Ratio Chi-Square = 52.656, DF = 4, P-Value = 0.000

Tabulated statistics: Origin, Factor

Rows: Origin   Columns: Factor

	1	2	3	4	5	All
Asia	80	51	65	41	18	255
	67.26	51.77	81.77	23.71	30.48	255.00
USA	59	56	104	8	45	272
	71.74	55.23	87.23	25.29	32.52	272.00
All	139	107	169	49	63	527
	139.00	107.00	169.00	49.00	63.00	527.00

Pearson Chi-Square = 45.701, DF = 4, P-Value = 0.000
Likelihood Ratio Chi-Square = 48.220, DF = 4, P-Value = 0.000

Figure 4-30: Chi-square Test for Predictors (Age, Origin, Gender) vs. Factor

Tabulated statistics: Origin, Design

Rows: Origin   Columns: Design

	1	2	3	4	5	All
Asia	38	42	61	28	86	255
	40.16	25.16	84.19	35.32	70.16	255.00
USA	45	10	113	45	59	272
	42.84	26.84	89.81	37.68	74.84	272.00
All	83	52	174	73	145	527
	83.00	52.00	174.00	73.00	145.00	527.00

Pearson Chi-Square = 44.307, DF = 4, P-Value = 0.000
Likelihood Ratio Chi-Square = 46.049, DF = 4, P-Value = 0.000

Tabulated statistics: Gender, Design

Rows: Gender   Columns: Design

	1	2	3	4	5	All
F	53	39	85	24	60	261
	41.11	25.75	86.17	36.15	71.81	261.00
M	30	13	89	49	85	266
	41.89	26.25	87.83	36.85	73.19	266.00
All	83	52	174	73	145	527
	83.00	52.00	174.00	73.00	145.00	527.00

Pearson Chi-Square = 32.293, DF = 4, P-Value = 0.000
Likelihood Ratio Chi-Square = 33.176, DF = 4, P-Value = 0.000

Tabulated statistics: Age, Design

Rows: Age   Columns: Design

	1	2	3	4	5	All
14-17	28	14	46	7	38	133
	20.95	13.12	43.91	18.42	36.59	133.00
18-22	29	16	49	14	42	150
	23.62	14.80	49.53	20.78	41.27	150.00
23-29	16	13	37	19	36	121
	19.06	11.94	39.95	16.76	33.29	121.00
30-39	10	9	42	33	29	123
	19.37	12.14	40.61	17.04	33.84	123.00
All	83	52	174	73	145	527
	83.00	52.00	174.00	73.00	145.00	527.00

Pearson Chi-Square = 35.581, DF = 12, P-Value = 0.000
Likelihood Ratio Chi-Square = 35.633, DF = 12, P-Value = 0.000

Figure 4-31: Chi-square Test for Predictors (Age, Origin, Gender) vs. Design Issues

Tabulated statistics: Origin, Form

Rows: Origin   Columns: Form

	1	2	3	4	All
Asia	51	49	136	19	255
	39.68	65.32	99.19	50.81	255.00
USA	31	86	69	86	272
	42.32	69.68	105.81	54.19	272.00
All	82	135	205	105	527
	82.00	135.00	205.00	105.00	527.00

Pearson Chi-Square = 79.203, DF = 3, P-Value = 0.000
Likelihood Ratio Chi-Square = 83.222, DF = 3, P-Value = 0.000

Tabulated statistics: Gender, Form

Rows: Gender   Columns: Form

	1	2	3	4	All
F	38	80	104	39	261
	40.6	66.9	101.5	52.0	261.0
M	44	55	101	66	266
	41.4	68.1	103.5	53.0	266.0
All	82	135	205	105	527
	82.0	135.0	205.0	105.0	527.0

Pearson Chi-Square = 12.009, DF = 3, P-Value = 0.007
Likelihood Ratio Chi-Square = 12.114, DF = 3, P-Value = 0.007

Tabulated statistics: Age, Form

Rows: Age   Columns: Form

	1	2	3	4	All
14-17	36	37	43	17	133
	20.69	34.07	51.74	26.50	133.00
18-22	17	39	54	40	150
	23.34	38.43	58.35	29.89	150.00
23-29	14	26	59	22	121
	18.83	31.00	47.07	24.11	121.00
30-39	15	33	49	26	123
	19.14	31.51	47.85	24.51	123.00
All	82	135	205	105	527
	82.00	135.00	205.00	105.00	527.00

Pearson Chi-Square = 28.266, DF = 9, P-Value = 0.001
Likelihood Ratio Chi-Square = 26.657, DF = 9, P-Value = 0.002

Figure 4-32: Chi-square Test for Predictors (Age, Origin, Gender) vs. Form

4.4 Consumers' Preference Design Features by Using Nominal Logistic Regression

The chi-square analyses I performed in the previous section helped to determine which dependent and independent variables had statistically significant associations. To provide more details about these associations, a nominal logistic regression analysis is performed.

In this study, a logistic regression model for Factor will relate the predictors, Origin, Gender, and Age to the likelihood that a survey respondent selects one factor over another as the most important mobile phone feature. For the reference event, I choose any response I want to compare to all the other responses to. For example, suppose I select Price as the reference event. Each logit in the regression will then give us the odds that a survey respondent of a given origin, gender and age will choose each given response as opposed to Price. That is, logit 1 shows the odds that each category of respondent chooses Design (1) as opposed to Price (2), and is thus also represented as logit (1/2); logit 2 shows the odds that each category of respondent chooses Functions (3) as opposed to Price (2), and is thus also represented as logit (3/2); logit 3 shows the odds that each category of respondent chooses Brand (4) as opposed to Price (2), and is thus also represented as logit (4/2), and logit 4 shows the odds that each category of respondent chooses Easy User Interface (5) as opposed to Price (2), and is thus also represented as logit (5/2). Thus I identify these odds ratios, or likelihoods, as a way of predicting a certain response to the Factor survey question based on the respondent's origin, gender, and age. The 95% confidence interval for each odds ratio indicates how precise the estimate for each odds ratio is based on your data; the narrower the confidence interval, the more precise the estimate. If other parameters remain constant, I generally increase the precision of the odds ratio estimates (and thus narrow the confidence intervals) by increasing the sample size. If the value of "1" (which represents "even odds") is not covered in the range of the confidence interval, the results are statistically significant.

4.4.1 Entering and Selecting Terms for the Model

This study is to predict the mobile phone preferences of an individual based on their age, gender, and origin. Therefore, I set up and run each logistic regression model using Factor, Design or Form as the response variable. That means I simultaneously evaluate three regression models, one for each response. Because I am testing multiple models, I need to adjust the level of significance (alpha) to ensure that multiple testing does not result in finding statistically significant results purely by chance. To do this, I make the Bonferroni correction, and divide the alpha level (0.05) by the number of logistic regression models I will evaluate with data (three—one for each response variable). This gives an adjusted alpha value of 0.05/3 = .017; if the overall model is associated with a value less than 0.017, I can assume the results are statistically significant.

When I run a logistic regression, generally I start by including all the terms and interactions in the model, then eliminating each statistically insignificant term one at a time, starting with the higher-order terms first. When I decide to a reference response variable, the least frequent response variable is determined. Figure 4-33, 4-34, and 4-35 show how to set up and evaluate the logistic regression using Factor as the response, and Origin, Gender, Age as the predictors. (Recall that in section 4.3 I found that Origin, Gender, Age were likely to be significant independent predictors for factor).

Nominal Logistic Regression: Factor versus Origin, Gender, Age

```
* WARNING * Algorithm has not converged after 20 iterations.
* WARNING * Convergence has not been reached for the parameter estimates
            criterion.
* WARNING * The results may not be reliable.
* WARNING * Try increasing the maximum number of iterations.
```

```
Response Information

Variable  Value                     Count
Factor    Brand                       49  (Reference Event)
          Easy user Interface         63
          Functions                  169
          Price                      107
          Design                     139
          Total                      527

Logistic Regression Table
                                                                        95%
                                                                         CI
Predictor                  Coef   SE Coef      Z      P   Odds Ratio  Lower
Logit 1: (Easy user Interface/Brand)
Constant               -1.79176   1.08012  -1.66  0.097
Origin
 USA                    1.38629   1.41421   0.98  0.327         4.00   0.25
Gender
 M                     0.538997   1.34519   0.40  0.689         1.71   0.12
Age
 18-22                 0.182322   1.53840   0.12  0.906         1.20   0.06
 23-29                  1.79176   1.35401   1.32  0.186         6.00   0.42
 30-39                  2.48491   1.38444   1.79  0.073        12.00   0.80
Origin*Gender
 USA*M                 0.271934   1.86445   0.15  0.884         1.31   0.03
Gender*Age
 M*18-22                2.45674   2.06386   1.19  0.234        11.67   0.20
 M*23-29               -2.93689   1.88868  -1.55  0.120         0.05   0.00
 M*30-39               -2.33076   1.79616  -1.30  0.194         0.10   0.00
Origin*Age
 USA*18-22              19.0657   7489.65   0.00  0.998  1.90602E+08   0.00
 USA*23-29              19.7308   8827.04   0.00  0.998  3.70649E+08   0.00
 USA*30-39              19.1128   9165.31   0.00  0.998  1.99801E+08   0.00
Origin*Gender*Age
 USA*M*18-22           -1.71364   10448.3  -0.00  1.000         0.18   0.00
 USA*M*23-29            3.03214   14164.3   0.00  1.000        20.74   0.00
 USA*M*30-39           -18.5738   9165.31  -0.00  0.998         0.00   0.00
Logit 2: (Functions/Brand)
Constant              -0.182322  0.605530  -0.30  0.763
Origin
 USA                    1.16315  0.908295   1.28  0.200         3.20   0.54
Gender
 M                     0.182322  0.807701   0.23  0.821         1.20   0.25
Age
 18-22                -0.328504  0.948683  -0.35  0.729         0.72   0.11
 23-29                 0.875469  0.930949   0.94  0.347         2.40   0.39
 30-39                  1.43508   1.00475   1.43  0.153         4.20   0.59
Origin*Gender
 USA*M                 0.916291   1.29353   0.71  0.479         2.50   0.20
Gender*Age
 M*18-22                3.16172   1.49818   2.11  0.035        23.61   1.25
 M*23-29               -1.07614   1.16379  -0.92  0.355         0.34   0.03
 M*30-39              -0.828949   1.24612  -0.67  0.506         0.44   0.04
Origin*Age
 USA*18-22              20.8983   7489.65   0.00  0.998  1.19126E+09   0.00
 USA*23-29              19.3943   8827.04   0.00  0.998  2.64749E+08   0.00
 USA*30-39              18.4399   9165.31   0.00  0.998  1.01939E+08   0.00
Origin*Gender*Age
 USA*M*18-22           -3.69158   10448.3  -0.00  1.000         0.02   0.00
```

USA*M*23-29	0.837186	14164.3	0.00	1.000	2.31	0.00
USA*M*30-39	-19.7392	9165.31	-0.00	0.998	0.00	0.00
Logit 3: (Price/Brand)						
Constant	0.510826	0.516398	0.99	0.323		
Origin						
USA	0.587787	0.843274	0.70	0.486	1.80	0.34
Gender						
M	-1.07044	0.812111	-1.32	0.187	0.34	0.07
Age						
18-22	-0.328504	0.795822	-0.41	0.680	0.72	0.15
23-29	0.470004	0.851469	0.55	0.581	1.60	0.30
30-39	1.36098	0.918471	1.48	0.138	3.90	0.64
Origin*Gender						
USA*M	-0.0281709	1.45051	-0.02	0.985	0.97	0.06
Gender*Age						
M*18-22	2.27441	1.50870	1.51	0.132	9.72	0.51
M*23-29	-1.20967	1.24181	-0.97	0.330	0.30	0.03
M*30-39	-1.49451	1.31774	-1.13	0.257	0.22	0.02
Origin*Age						
USA*18-22	20.7115	7489.65	0.00	0.998	9.88308E+08	0.00
USA*23-29	19.6820	8827.04	0.00	0.998	3.52999E+08	0.00
USA*30-39	19.2717	9165.31	0.00	0.998	2.34199E+08	0.00
Origin*Gender*Age						
USA*M*18-22	-2.07851	10448.3	-0.00	1.000	0.13	0.00
USA*M*23-29	0.683036	14164.3	0.00	1.000	1.98	0.00
USA*M*30-39	-18.8505	9165.31	-0.00	0.998	0.00	0.00
Logit 4: (Design/Brand)						
Constant	0.916291	0.483046	1.90	0.058		
Origin						
USA	0.287682	0.816497	0.35	0.725	1.33	0.27
Gender						
M	-0.377294	0.677882	-0.56	0.578	0.69	0.18
Age						
18-22	0.307485	0.701539	0.44	0.661	1.36	0.34
23-29	0.382992	0.810910	0.47	0.637	1.47	0.30
30-39	-0.223144	0.991632	-0.23	0.822	0.80	0.11
Origin*Gender						
USA*M	0.677399	1.22636	0.55	0.581	1.97	0.18
Gender*Age						
M*18-22	0.945278	1.37296	0.69	0.491	2.57	0.17
M*23-29	-1.37397	1.05713	-1.30	0.194	0.25	0.03
M*30-39	-0.0281709	1.22523	-0.02	0.982	0.97	0.09
Origin*Age						
USA*18-22	19.6337	7489.65	0.00	0.998	3.36357E+08	0.00
USA*23-29	19.3759	8827.04	0.00	0.998	2.59936E+08	0.00
USA*30-39	20.0573	9165.31	0.00	0.998	5.13773E+08	0.00
Origin*Gender*Age						
USA*M*18-22	-1.65747	10448.3	-0.00	1.000	0.19	0.00
USA*M*23-29	1.52806	14164.3	0.00	1.000	4.61	0.00
USA*M*30-39	-20.7992	9165.31	-0.00	0.998	0.00	0.00

Log-Likelihood = -703.922
Test that all slopes are zero: G = 188.677, DF = 60, P-Value = 0.000

* NOTE * No goodness of fit test performed.
* NOTE * The model uses all degrees of freedom.

Figure 4-33: Nominal Logistic Regression Results (Full Model)

The condensed main results are shown in Figure 4-33. The warnings at the top of the output occur because with the full model that includes all the terms and interactions, some categories have very few responses, which make the odd ratios difficult to estimate in relation to all the interactions. As I eliminate terms from the model based on their statistical significance, I ignore this warning because I am not yet interpreting the odds ratios. Later, this warning will not appear in the final reduced model. Brand (4) is listed as the reference event. The overall p-value (0.000) for the model is less than alpha (0.017), suggesting that at least one of the terms in the model is statistically significant. To simplify the model, I start by looking at the highest-order term in the model—for this model, this is the 3-way interaction Origin*Gender*Age. I evaluate the p-value for this term in all of the logits; if it is less than 0.2 in one or more of the logits, I keep the term in the model.

Sometimes a value other than 0.2 is used as a cut-off value to determine which terms to keep in the model. The cut-off value 0.20 is the standard. The goal of keeping these terms in the model is to explain as much variability in the data with the model as possible (i.e., improve the model fit). However, a term that is included in the final model may or may not end up being statistically significant (in this case, when $p < 0.017$). In all the logits, none of the associated p-values for the 3-way interaction Origin*Gender*Age are less than 0.2; therefore, I will drop the interaction from the model and re-run the analysis.

Nominal Logistic Regression: Factor versus Origin, Gender, Age

```
* WARNING * Algorithm has not converged after 20 iterations.
* WARNING * Convergence has not been reached for the parameter estimates
            criterion.
* WARNING * The results may not be reliable.
* WARNING * Try increasing the maximum number of iterations.

Response Information

Variable  Value                      Count
Factor    Brand                         49  (Reference Event)
          Easy user interface           63
          Functions                    169
          Price                        107
          Design                       139
```

Total 527

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Logit 1: (Easy user interface / Brand)							
Constant	-1.52570	0.828202	-1.84	0.065			
Origin							
USA	1.06184	1.09166	0.97	0.331	2.89	0.34	24.57
Gender							
M	0.114821	1.00288	0.11	0.909	1.12	0.16	8.01
Age							
18-22	0.0105167	1.20275	0.01	0.993	1.01	0.10	10.68
23-29	1.34728	1.13001	1.19	0.233	3.85	0.42	35.24
30-39	2.47992	1.15816	2.14	0.032	11.94	1.23	115.57
Origin*Gender							
USA*M	0.672322	1.13657	0.59	0.554	1.96	0.21	18.17
Gender*Age							
M*18-22	2.76164	1.65266	1.67	0.095	15.83	0.62	403.78
M*23-29	-1.93535	1.27612	-1.52	0.129	0.14	0.01	1.76
M*30-39	-2.23506	1.27814	-1.75	0.080	0.11	0.01	1.31
Origin*Age							
USA*18-22	18.4940	4899.92	0.00	0.997	1.07607E+08	0.00	*
USA*23-29	21.3760	6486.76	0.00	0.997	1.92070E+09	0.00	*
USA*30-39	0.697337	1.28537	0.54	0.587	2.01	0.16	24.94
Logit 2: (Functions / Brand)							
Constant	-0.493928	0.578596	-0.85	0.393			
Origin							
USA	1.76709	0.832904	2.12	0.034	5.85	1.14	29.95
Gender							
M	0.703047	0.714133	0.98	0.325	2.02	0.50	8.19
Age							
18-22	0.430168	0.810863	0.53	0.596	1.54	0.31	7.53
23-29	1.19171	0.884609	1.35	0.178	3.29	0.58	18.64
30-39	1.91170	0.961228	1.99	0.047	6.76	1.03	44.51
Origin*Gender							
USA*M	-0.174071	1.01172	-0.17	0.863	0.84	0.12	6.10
Gender*Age							
M*18-22	2.08914	1.35081	1.55	0.122	8.08	0.57	114.06
M*23-29	-1.60461	1.03700	-1.55	0.122	0.20	0.03	1.53
M*30-39	-1.53072	1.06086	-1.44	0.149	0.22	0.03	1.73
Origin*Age							
USA*18-22	19.0298	4899.92	0.00	0.997	1.83883E+08	0.00	*
USA*23-29	19.8568	6486.76	0.00	0.998	4.20450E+08	0.00	*
USA*30-39	-0.685293	1.03691	-0.66	0.509	0.50	0.07	3.85
Logit 3: (Price / Brand)							
Constant	0.455230	0.489903	0.93	0.353			
Origin							
USA	0.790233	0.811202	0.97	0.330	2.20	0.45	10.81
Gender							
M	-0.995419	0.742246	-1.34	0.180	0.37	0.09	1.58
Age							
18-22	-0.307044	0.761134	-0.40	0.687	0.74	0.17	3.27
23-29	0.604724	0.823441	0.73	0.463	1.83	0.36	9.20
30-39	1.63000	0.906608	1.80	0.072	5.10	0.86	30.17
Origin*Gender							
USA*M	-0.367235	1.05278	-0.35	0.727	0.69	0.09	5.45
Gender*Age							
M*18-22	2.28264	1.39449	1.64	0.102	9.80	0.64	150.78
M*23-29	-1.61173	1.13841	-1.42	0.157	0.20	0.02	1.86
M*30-39	-1.66154	1.11928	-1.48	0.138	0.19	0.02	1.70
Origin*Age							
USA*18-22	19.8596	4899.92	0.00	0.997	4.21627E+08	0.00	*
USA*23-29	20.3891	6486.76	0.00	0.997	7.15928E+08	0.00	*
USA*30-39	0.427583	1.10079	0.39	0.698	1.53	0.18	13.26

```
Logit 4: (Design/ Brand)
Constant          0.841034  0.458045   1.84  0.066
Origin
 USA              0.532301  0.779381   0.68  0.495         1.70   0.37    7.85
Gender
 M               -0.232539  0.636872  -0.37  0.715         0.79   0.23    2.76
Age
 18-22            0.284634  0.682912   0.42  0.677         1.33   0.35    5.07
 23-29            0.440236  0.789492   0.56  0.577         1.55   0.33    7.30
 30-39            0.487331  0.920267   0.53  0.596         1.63   0.27    9.89
Origin*Gender
 USA*M            0.153193   1.00212   0.15  0.879         1.17   0.16    8.31
Gender*Age
 M*18-22           1.13352   1.31732   0.86  0.390         3.11   0.23   41.08
 M*23-29          -1.47305  0.993414  -1.48  0.138         0.23   0.03    1.61
 M*30-39          -1.13671   1.05580  -1.08  0.282         0.32   0.04    2.54
Origin*Age
 USA*18-22         18.9368   4899.92   0.00  0.997  1.67546E+08   0.00       *
 USA*23-29         20.2562   6486.76   0.00  0.998  6.26812E+08   0.00       *
 USA*30-39        0.341277   1.05677   0.32  0.747         1.41   0.18   11.16

Log-Likelihood = -708.222
Test that all slopes are zero: G = 180.078, DF = 48, P-Value = 0.000

Goodness-of-Fit Tests
Method    Chi-Square  DF      P
Pearson      8.06465  12  0.780
Deviance     8.59855  12  0.737
```

Figure 4-34: Nominal Logistic Regression Results (Reduced Model)

Figure 4-34, the two-way interactions are examined to the highest order terms in the reduced model. Origin*Age generally has the highest p-values and its p-values are greater than 0.2 in all of the logits, so remove this interaction from the model and re-run the analysis. Keep repeating this procedure until the model only includes terms with p < 0.2 or terms that are associated with an interaction whose p-value is less than 0.2. (If an interaction has a p-value < 0.2, do not remove the individual terms that make up the interaction, even if their associated p-values are greater than 0.2). Repeating this procedure, I obtain the final reduced model in Figure 4-35:

Nominal Logistic Regression: Factor versus Origin, Gender, Age

```
Response Information

Variable  Value                    Count
Factor    Brand                       49  (Reference Event)
          Easy user interface         63
          Functions                  169
          Price                      107
```

Design	139
Total	527

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Logit 1: (Easy user interface / Brand)							
Constant	-2.21633	0.717231	-3.09	0.002			
Origin							
USA	2.62528	0.483843	5.43	0.000	13.81	5.35	35.65
Gender							
M	0.472476	0.894100	0.53	0.597	1.60	0.28	9.25
Age							
18-22	0.0139372	1.08865	0.01	0.990	1.01	0.12	8.57
23-29	2.41524	0.959809	2.52	0.012	11.19	1.71	73.44
30-39	2.92137	1.03887	2.81	0.005	18.57	2.42	142.25
Gender*Age							
M*18-22	2.77403	1.62775	1.70	0.088	16.02	0.66	389.33
M*23-29	-1.66650	1.20506	-1.38	0.167	0.19	0.02	2.00
M*30-39	-2.11175	1.27947	-1.65	0.099	0.12	0.01	1.49
Logit 2: (Functions / Brand)							
Constant	-0.400605	0.464774	-0.86	0.389			
Origin							
USA	2.10366	0.424506	4.96	0.000	8.20	3.57	18.83
Gender							
M	0.539870	0.608811	0.89	0.375	1.72	0.52	5.66
Age							
18-22	0.776824	0.685707	1.13	0.257	2.17	0.57	8.34
23-29	1.26143	0.787123	1.60	0.109	3.53	0.75	16.51
30-39	1.51691	0.893470	1.70	0.090	4.56	0.79	26.26
Gender*Age							
M*18-22	2.09394	1.29405	1.62	0.106	8.12	0.64	102.54
M*23-29	-1.49414	0.970478	-1.54	0.124	0.22	0.03	1.50
M*30-39	-1.49995	1.06600	-1.41	0.159	0.22	0.03	1.80
Logit 3: (Price / Brand)							
Constant	0.172701	0.431174	0.40	0.689			
Origin							
USA	1.76329	0.443664	3.97	0.000	5.83	2.44	13.91
Gender							
M	-1.17775	0.678539	-1.74	0.083	0.31	0.08	1.16
Age							
18-22	0.526562	0.657354	0.80	0.423	1.69	0.47	6.14
23-29	0.998460	0.759675	1.31	0.189	2.71	0.61	12.03
30-39	1.85355	0.849488	2.18	0.029	6.38	1.21	33.74
Gender*Age							
M*18-22	2.18343	1.35215	1.61	0.106	8.88	0.63	125.67
M*23-29	-1.55655	1.10233	-1.41	0.158	0.21	0.02	1.83
M*30-39	-1.68552	1.12811	-1.49	0.135	0.19	0.02	1.69
Logit 4: (Design / Brand)							
Constant	0.657840	0.404407	1.63	0.104			
Origin							
USA	1.31673	0.427025	3.08	0.002	3.73	1.62	8.62
Gender							
M	-0.191915	0.564877	-0.34	0.734	0.83	0.27	2.50
Age							
18-22	0.584912	0.630497	0.93	0.354	1.79	0.52	6.18
23-29	0.762438	0.742734	1.03	0.305	2.14	0.50	9.19
30-39	0.639549	0.871478	0.73	0.463	1.90	0.34	10.46
Gender*Age							
M*18-22	1.13411	1.28145	0.89	0.376	3.11	0.25	38.31
M*23-29	-1.40915	0.942141	-1.50	0.135	0.24	0.04	1.55

```
 M*30-39         -1.10939   1.05634  -1.05  0.294   0.33   0.04    2.61

Log-Likelihood = -719.015
Test that all slopes are zero: G = 158.492, DF = 32, P-Value = 0.000

Goodness-of-Fit Tests

Method    Chi-Square  DF      P
Pearson      26.9082  28  0.523
Deviance     30.1850  28  0.354
```

Figure 4-35: Nominal Logistic Regression Results (Final Model)

In the final model of Figure 4-35, each term has an associated p-value < 0.2 in at least one of the logits. Origin and the interaction Gender * Age have at least one p-value < 0.2 in all the logits; Age has at least one p-value < 0.2 in logits 1, 2, and 3; and Gender has at least one p-value < 0.2 in logit 3. The p-values for the overall goodness of fit tests are 0.523 and 0.354. Because these values are much greater than 0.05, there is no evidence that the model does not adequately fit the data. In other words, I can assume that my final reduced model fits the data adequately.

4.4.2 Nominal Logistic Regression – Factor vs. Origin, Gender, Age

To determine which terms in the final model are statistically significant in Figure 4-36, I compare their p-values with the adjusted alpha level of 0.017 in each logit. In Logit 1, which evaluates the response of Easy User Interface compared to Brand (5/4), the Origin p-value is 0.000, which is < 0.017, so the term is statistically significant. The associated odds ratio is 13.81, which means that US survey respondents are nearly 14 times as likely as Asians survey respondents to cite Easy User Interface rather than Brand as the most important factor when purchasing a mobile phone.

For the Age factor in Logit 1, I see that the results for age groups 23-29 and 30-39 are also statistically significant (p-values = 0.012 and 0.005, respectively). Compared to those in the youngest (14-17) age group, respondents in the 23-29 year-old age group are 11.19 times as likely to cite Easy User Interface over Brand as the most important factor; respondents in the 30-39 year-old age group are 18.57 times as likely to cite Easy User Interface over Brand as the most important factor.

```
Response Variable Value                       Count
Factor          Brand                          49  (Reference Event)
                Easy user interface            63
                Functions                     169
                Price                         107
                Design                        139
                Total                         527
Logistic Regression Table
                                                         Odds       98.3% CI
Predictor              Coef    SE Coef        Z       P  Ratio  Lower    Upper
Logit 1: (Easy user interface/Brand)
Constant          -2.21633  0.717231  -3.09  0.002        0.01963    0.605
Origin
USA                2.62528  0.483843   5.43  0.000  13.81 4.34445   43.889
Gender
M                 0.472476  0.894100   0.53  0.597   1.60 0.18930   13.591
Age
18-22            0.0139372   1.08865   0.01  0.990   1.01 0.07518   13.678
23-29              2.41524  0.959809   2.52  0.012  11.19 1.12896  110.962
30-39              2.92137   1.03887   2.81  0.005  18.57 1.55032  222.357
Gender*Age
M*18-22            2.77403   1.62775   1.70  0.088  16.02 0.32749  783.962
M*23-29           -1.66650   1.20506  -1.38  0.167   0.19 0.01060    3.366
M*30-39           -2.11175   1.27947  -1.65  0.099   0.12 0.00569    2.576
Logit 2: (Functions/Brands)
Constant         -0.400605  0.464774 -0.86  0.389         0.22060    2.034
Origin
USA                2.10366  0.424506   4.96  0.000   8.20  2.97158   22.606
Gender
M                 0.539870  0.608811   0.89  0.375   1.72  0.40044    7.352
Age
18-22             0.776824  0.685707   1.13  0.257   2.17 0.42231   11.197
23-29              1.26143  0.787123   1.60  0.109   3.53 0.53806   23.165
30-39              1.51691  0.893470   1.70  0.090   4.56 0.53876   38.563
Gender*Age
M*18-22            2.09394   1.29405   1.62  0.106   8.12 0.36831  178.879
M*23-29           -1.49414  0.970478  -1.54  0.124   0.22 0.02207    2.283
M*30-39           -1.49995   1.06600  -1.41  0.159   0.22 0.01746    2.851
Logit 3: (Prices/Brands)
Constant          0.172701  0.431174  0.40  0.689         0.42409    3.331
Origin
USA                1.76329  0.443664   3.97  0.000   5.83 2.01968   16.838
Gender
M                 -1.17775  0.678539  -1.74  0.083   0.31 0.06084    1.559
Age
18-22             0.526562  0.657354   0.80  0.423   1.69 0.35186    8.147
23-29             0.998460  0.759675   1.31  0.189   2.71 0.44168   16.678
30-39              1.85355  0.849488   2.18  0.029   6.38 0.83801   48.610
Gender*Age
M*18-22            2.18343   1.35215   1.61  0.106   8.88 0.35057  224.768
M*23-29           -1.55655   1.10233  -1.41  0.158   0.21 0.01513    2.939
M*30-39           -1.68552   1.12811  -1.49  0.135   0.19 0.01250    2.747
Logit 4: (Design/Brands)
Constant          0.657840  0.404407  1.63  0.104         0.73441    5.075
Origin
USA                1.31673  0.427025   3.08  0.002   3.73 1.34465   10.353
Gender
M                -0.191915  0.564877  -0.34  0.734   0.83 0.21396    3.184
Age
18-22             0.584912  0.630497   0.93  0.354   1.79 0.39773    8.099
23-29             0.762438  0.742734   1.03  0.305   2.14 0.36324   12.649
30-39             0.639549  0.871478   0.73  0.463   1.90 0.23615   15.217
Gender*Age
M*18-22            1.13411   1.28145   0.89  0.376   3.11 0.14536   66.471
M*23-29           -1.40915  0.942141 -1.50  0.135   0.24 0.02571    2.322
M*30-39           -1.10939   1.05634  -1.05  0.294   0.33 0.02641    4.118

Log-Likelihood = -719.015
Test that all slopes are zero: G = 158.492, DF = 32, P-Value = 0.000

Goodness-of-Fit Tests
Method    Chi-Square  DF      P
Pearson      26.9082  28  0.523
Deviance     30.1850  28  0.354
```

Figure 4-36: Nominal Logistic Regression-Factors vs. Origin, Gender, Age

Logit 2 is the parameter estimates of the change in logits of functions relative to the reference event, brand. The p- value 0.000 for origin, which is < 0.017, respectively, indicate that there is sufficient evidence, if the p-values are less than your acceptable α -level. The associated odds ratio is 8.2, which means that US survey respondents are 8 times higher Asian survey respondents to cite functions rather than brand as the most important factor. Logit 3, which evaluates the response of price compared to brand, origin p-value is 0.000, which is < 0.017, so the term is statistically significant. The associated odds ratio is 5.83, which means that US survey respondents are nearly 6 times as likely as Asian survey respondents to cite price rather than brand as the most important factor when purchasing a mobile phone. In logit 4, the associated odds ratio is 3.73 and p-value of origin is 0.002. Thus, US survey respondents are 3.7 times higher Asian survey respondents to cite design rather than brand as the most important factor.

Recall that I am using an adjusted alpha of 0.05/3 = 0.017, rather than 0.05, to account for performing the logistic regression analysis three times, once for each of the 3 response variables. Therefore, the 95% confidence intervals for the odds ratios shown in the output, which are based on an alpha value of 0.05, are not appropriate to use. With an alpha of 0.017, you would need to use 100-0.017 % = 98.3% confidence intervals to accurately evaluate the precision of the odds ratio estimates. The 98.3% confidence intervals will be wider than the 95% confidence intervals because the alpha level is more conservative. For odds ratios associated with p-values less than 0.017, the lower bound of the 98.3% confidence intervals would be greater than 1; for odds ratios associated with p-values greater than 0.017, the lower bound of the 98.3% confidence interval would be less than 1. Appendix B indicates how to calculate these intervals from the output.

4.4.3 Nominal Logistic Regression – Design Issue vs. Origin, Gender, Age

```
Response Variable Value       Count
Design    Body color           52 (Reference Event)
          Form                145
          Screen Size          73
          Keypad              174
          Weight               83
          Total               527
Logistic Regression Table
                                                        Odds       98.3% CI
Predictor            Coef    SE Coef       Z       P   Ratio  Lower     Upper
Logit 1: (Form/Body Color)
Constant          1.62772   0.481841    3.38   0.001         1.60981       16.108
Origin
 USA              2.19940    1.06447    2.07   0.039    9.02 0.70845      114.833
Gender
 F               -1.20730   0.405732   -2.98   0.003    0.30 0.11338        0.789
Age
 14-17          -0.215610   0.504829   -0.43   0.669    0.81 0.24119        2.694
 18-22          -0.384917   0.495017   -0.78   0.437    0.68 0.20846        2.221
 23-29          -0.195987   0.511946   -0.38   0.702    0.82 0.24183        2.794
Origin*Gender
 USA*F           -1.45174    1.16064   -1.25   0.211    0.23 0.01462        3.752
Logit 2: (Screen size/body color)
Constant          1.61953   0.507732    3.19   0.001        1.50088       16.997
Origin
 USA              2.85677    1.08557    2.63   0.008   17.41 1.29984      233.058
Gender
 F               -2.06545   0.564538   -3.66   0.000    0.13 0.03289        0.489
Age
 14-17           -2.05434   0.619426   -3.32   0.001    0.13 0.02917        0.563
 18-22           -1.72946   0.552802   -3.13   0.002    0.18 0.04733        0.665
 23-29          -0.984152   0.547052   -1.80   0.072    0.37 0.10110        1.382
Origin*Gender
 USA*F          -0.703088    1.23801   -0.57   0.570    0.50 0.02568        9.543
Logit 3: (keypad/body color)
Constant          1.20417   0.493298    2.44   0.015        1.02550       10.839
Origin
 USA              3.40386    1.06772    3.19   0.001   30.08 2.34437      385.952
Gender
 F              -0.638301   0.430057   -1.48   0.138    0.53 0.18898        1.476
Age
 14-17          -0.392908   0.497355   -0.79   0.430    0.68 0.20565        2.216
 18-22          -0.687379   0.488999   -1.41   0.160    0.50 0.15628        1.618
 23-29          -0.539246   0.508360   -1.06   0.289    0.58 0.17304        1.965
Origin*Gender
 USA*F           -1.81158    1.15438   -1.57   0.117    0.16 0.01035        2.579
Logit 4: (Weight/Body Color)
Constant        0.0528556   0.575574    0.09   0.927        0.26640        4.172
Origin
 USA              2.20966    1.10510    2.00   0.046    9.11 0.64952      127.849
Gender
 F              -0.701528   0.473389   -1.48   0.138    0.50 0.15994        1.537
Age
 14-17           0.571399   0.574410    0.99   0.320    1.77 0.44868        6.988
 18-22           0.316229   0.566719    0.56   0.577    1.37 0.35408        5.316
 23-29          0.0704537   0.602299    0.12   0.907    1.07 0.25435        4.527
Origin*Gender
 USA*F          -0.583444    1.20166   -0.49   0.627    0.56  0.03157       9.860
Log-Likelihood = -735.968
Test that all slopes are zero: G = 124.226, DF = 24, P-Value = 0.000
Goodness-of-Fit Tests
Method    Chi-Square  DF      P
Pearson      46.4362  36  0.114
Deviance     50.6631  36  0.053
```

Figure 4-37: Nominal Logistic Regression -Design Issues vs. Origin, Gender, Age

In Figure 4-37, logit 2, which evaluates the response of screen size compared to body color, origin p-value is 0.008, which is < 0.017, so the term is statistically significant. The associated odds ratio is 17.41, which means that US survey respondents are nearly 17 times as likely as Asian survey respondents to cite screen size rather than body color as the most important design issue when purchasing a mobile phone.

Logit 3 is the parameter estimates of the change in logits of keypad design relative to the reference event, body color. The p- value 0.001 for origin, which is < 0.017, respectively, indicate that there is sufficient evidence, if the p-values are less than your acceptable α -level. The associated odds ratio is 30.08, which means that US survey respondents are 30 times higher Asian survey respondents to cite keypad rather than body color as the most important design issue.

In first set of estimated logits, labeled Logit 1 (Form/ Body Color) are the parameter estimates of the change in logits of form relative to the reference event, body color. The p-vale 0.003 for gender, which is < 0.017, points out that the term is statistically important. The associated odds ratio is 0.80, which means that females are 0.8 times higher for males to cite forms rather than body color. There is sufficient evidence to conclude that females prefer forms to body color. However, without gender, consumers' preference about form and body color is almost the similarity by identifying odds ratio value, 0.80.

4.4.4 Nominal Logistic Regression – Form vs. Origin, Gender, Age

```
Response Variable Value               Count
Form      Clamshell folder             82 (Reference Event)
          Flip                        105
          Slider                      205
          Block                        135
          Total                        527
Logistic Regression Table
                                                   Odds       98.3% CI
Predictor              Coef   SE Coef       Z       P  Ratio     Lower     Upper
Logit 1: (Flip/clamshell folder)
Constant          -2.72976  0.556309  -4.91  0.000           0.08062     1.118
Origin
 USA               3.21680  0.628661   5.12  0.000  24.95   5.55251   112.095
Gender
 M                 1.36614  0.599191   2.28  0.023   3.92   0.93615    16.416
Age
 18-22             1.52653  0.429591   3.55  0.000   4.60   0.07783     0.607
 23-29             1.17680  0.468579   2.51  0.012   3.24   0.23241     2.138
 30-39             1.29508  0.455828   2.84  0.004   3.65   0.26982     2.333
Origin*Gender
 USA*M            -2.03133  0.763777  -2.66  0.008   0.13   0.02114     0.814
Logit 2: (slider/clamshell folder)
Constant          0.188412  0.283786   0.66  0.507            1.47480    7.239
Origin
 USA              0.355317  0.460069   0.77  0.440   1.43   0.47509     4.284
Gender
 M                0.127242  0.339819   0.37  0.708   1.14   0.50413     2.558
Age
 18-22            0.995598  0.359726   2.77  0.006   2.71   0.15640     0.873
 23-29             1.26069  0.374473   3.37  0.001   3.53   0.49143     3.458
 30-39             1.01327  0.372716   2.72  0.007   2.75   0.38522     2.689
Origin*Gender
 USA*M           -0.879439  0.590425  -1.49  0.136   0.42   0.10121     1.702
Logit 3: (Block/clamshell folder)
Constant         -0.861733  0.351197  -2.45  0.014            0.36231    2.377
Origin
 USA               2.41061  0.477189   5.05  0.000  11.14   3.56132    34.851
Gender
 M                0.715965  0.409710   1.75  0.081   2.05   0.76856     5.448
Age
 18-22            0.787061  0.389150   2.02  0.043   2.20   0.17958     1.154
 23-29            0.575458  0.422184   1.36  0.173   1.78   0.27295     2.400
 30-39            0.783527  0.405841   1.93  0.054   2.19   0.34851     2.849
Origin*Gender
 USA*M            -2.62926  0.627878  -4.19  0.000   0.07   0.01608     0.323

Log-Likelihood = -627.824
Test that all slopes are zero: G = 143.090, DF = 18, P-Value = 0.000

Goodness-of-Fit Tests
Method    Chi-Square  DF      P
Pearson      31.7682  27  0.241
Deviance     33.3903  27  0.184
```

Figure 4-38: Nominal Logistic Regression -Form vs. Origin, Gender, Age

In Figure 4-38, logit 1, which evaluates the response of flip compared to clamshell folder, origin p-value is 0.000, which is < 0.017, so the term is statistically significant. The associated odds ratio is 24.95, which means that US survey respondents are nearly 25 times as likely as Asian survey respondents to prefer flip form rather than clamshell folder as the most important form style when purchasing a mobile phone.

In logit 1, the results for age groups 18-22 (p-value=0.000), 23-29 (p-value=0.0129), and 30-39 (p-value=0.004) are also statistically significant. Compared to those in the youngest (14-17) age group, respondents in the 18-22 years age group are 4.6 times as likely to prefer flip form over clamshell folder as the most important design issue; respondents in the 23-29 years age groups are 3.24 times higher than those; respondents in the 30-39 years age groups are 3.65 times as likely to prefer flip form over clamshell folder as the most important form style.

In logit 2, the results for age groups 18-22 (p-value=0.000), 23-29 (p-value=0.001), and 30-39 (p-value=0.007) are also statistically significant. Compared to those in the youngest (14-17) age group, respondents in the 18-22 years age group are 2.7 times as likely to prefer slider form over clamshell folder as the most important design issue; respondents in the 23-29 years age groups are 3.5 times higher than those; respondents in the 30-39 years age groups are 2.75 times as likely to prefer slider form over clamshell folder as the most important form style.

Logit 3 is the parameter estimates of the change in logits of block form relative to the reference event, clamshell folder. The p- value 0.000 for origin, which is < 0.017, respectively, indicate that there is sufficient evidence, if the p-values are less than your acceptable α -level. The associated odds ratio is 11.14, which means that US survey respondents are 11 times higher Asian survey respondents to prefer block style rather than clamshell folder as the most important form style.

In summary, the logistic regression model provides specific odds rations that confirm the general trends through exploratory survey data research. By looking at the logits in the model, I compare the probabilities in relation to each response for each factor, design issue, form style, and evaluate whether the results are statistically significant. All of the three logistic regression models have 0.000 of p-value, and I can identify goodness of fit test. The p-value for the Person test is 0.523, and the p-value for the deviance test is 0.354 in logistic regression model for factor. The p-value for the Person test is 0.114, and the p-value for the deviance test is 0.053 in logistic regression model for design issues. The p-value for the Person test is 0.241, and the p-value for the deviance test is 0.184 in logistic regression model for form style. If the p-value for person test is less than p-value 0.017, the test would indicate that the model does not fit the data. Thus, there are sufficient evidences that three models fit the data adequately.

4.5 Comparison Between the Historical and Future Oriented Data Mining

By comparing the past and current data, design forms and functions are growing in importance extending in the global market. Companies are continuously increasing product productions demanding through proper choice of design factors. Moreover, companies developing creative and innovative products are more concerned about consumers' emotional responses than high quality and technology. After all, suitable design criteria are related to increases in manufacturers' production and brand quality.

In terms of the historical data analysis, I found that consumers were more likely to be influenced by basic functions and selected form related characteristics (e.g., weight). This result might be interpreted as consumers did not prefer a product that includes many functions because additional functionality might complicate the user interface and pressure the power supply. Only basic functions were enough to make a product successful, and consumers preferred uncomplicated forms like the "block". The significant characteristics, identified using the regression analysis, are provided in Figure 4.39. However, one drawback with regards to the data pertaining to the historical data mining should be mentioned. Because it is very hard to reach data that is categorized based on consumer demographics, the results only reflect trends in general, not provide differences across groups with origin, age, and gender differences. This fact indeed might limit customization ability of the company based on historical data.

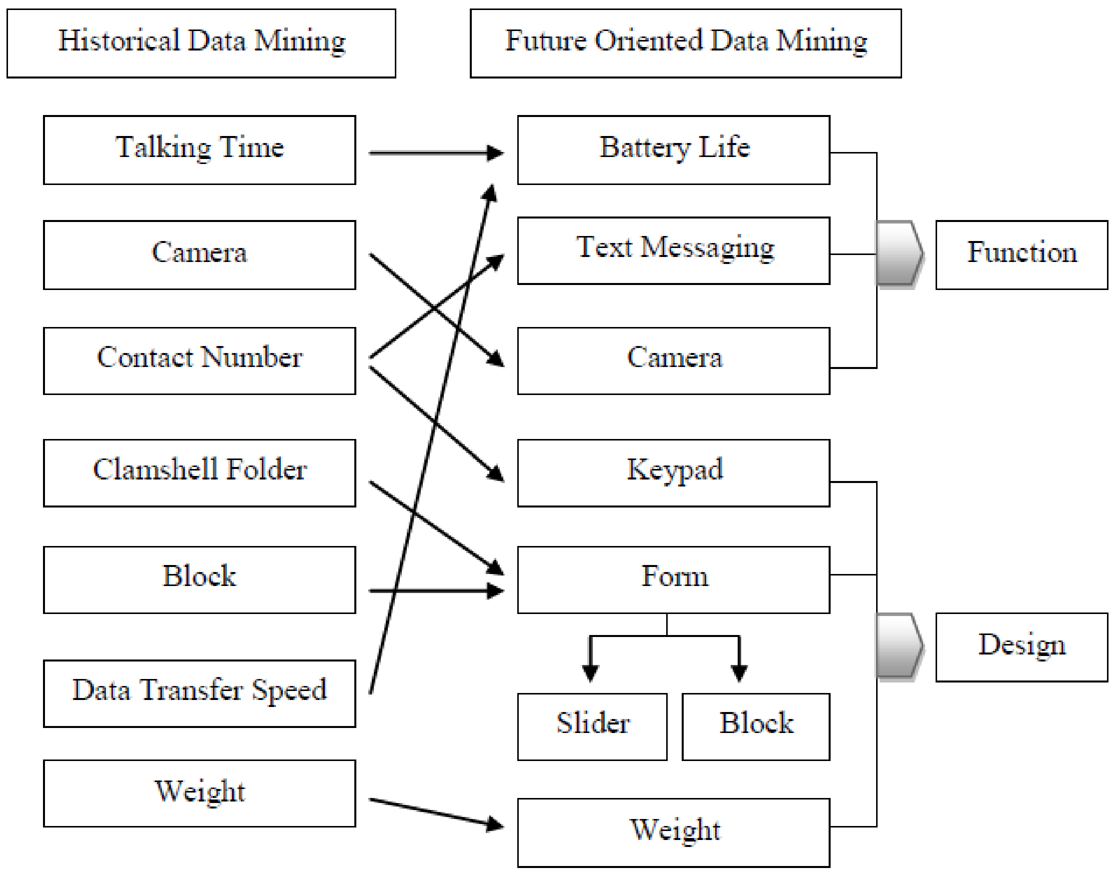

Figure 4-39: A Mapping of the Significant Mobile Phone Characteristics Identified Using Historical Data Mining & Future Oriented Data Mining.

In terms of current survey analysis, I found that consumers considered similar characteristics as the whole group. Indeed, Figure 4-39 shows a one-to one mapping of the characteristics between historical data mining and future oriented data mining results. However, because I have the data for the consumer demographics as collected as a part of the survey questions, I further investigated the differences across origin, age, and gender groups. According to origin, age, and gender, consumers' preferences criteria were different. They considered text messaging and call features as the most necessary function of mobile phone. Moreover, consumers preferred QWERTY keypad to be included, which could be used to send easily text messages or E-mail. Long battery life was considered the most essential function for phone call users. Accordingly, I found that consumers wanted improved essential functions such as text messaging and battery life more than new functions for mobile phones.

The results of the logistic regression are presented in Table 4-13. In the table, I have placed the significantly different consumer preferences under each grouping variable (origin, age, gender). As one can see from the table, female consumers pay more attention to form and screen size, and older age groups have a preference for flip and slider phones. However, looking at the results presented in Figure 4-39, these inferences cannot be made, and thus, it is recommended that consumer clusters are studies for their potentially different preferences when data is available. However, when completing future oriented data mining, it should also be kept in mind that the results only indicate consumers' opinion without actual sales, and hence, not as actual information.

Table 4-13: Summary Table for Significant Preferences of Consumer Groups

Segment / Preferences		Origin		Age				Gender	
		USA	Asia	14-17	18-22	23-29	30-39	Male	Female
Purchasing Factor (Brand)		Easy User Interface				Easy User Interface	Easy User Interface		
		Function							
		Price							
		Design							
Design Issues (Color)									Form
		Screen Size		Screen Size	Screen Size				Screen Size
		Keypad							
Form (Clamshell)		Flip			Flip	Flip	Flip		
					Slider	Slider	Slider		
		Block							

Looking ahead, consumers might be more likely to be influenced by a sensory experience with their mobile phone. In the past, product design focused on form and color as visual factors while recent product designs have been influenced by other senses. Sensors of hearing and touch have been recently regarded as important factors to improved physical interaction between user and product. Thus, manufacturer and consumers both consider design styles and forms as important factors in purchasing decision.

In the near future, product design will combine high technology, innovative form, and sensors to motivate consumers' psychological emotion. Even though products with high technology motivate consumers' curiosity, product design should be focused on humans' convenience and happiness. In the competitive market, evolution of technology design will create innovative user interface. As becoming more difficulty to distinguished forms and functions in consumers' choice, emotional user interfaces will occupy on the important role.

Chapter 5

Conclusion and Future Research

In this study, the relative importance of nine design factors was determined by the partial regression coefficients method. Based on the ranking of design factors, customers prefer the block and clamshell type, and consider a digital camera to be an important factor. Also, the data suggests that market share of mobile phone companies can improve as these manufacturers increase the average number of contacts, the average data transfer speed, and the average talk time of their line of phones. Based on these experimental results, battery life and available contact storage space are first order functions which must satisfy consumers' requirements. Because consumers are tired of new and increasingly complex functions, improving the quality of primary functions may lead to positive market share changes for mobile phone companies. In addition, the survey employed here made it possible to better understand current consumers' preferences and suggested a guide for what the future of mobile phone design might look like. Accordingly, future product development ought to concentrate on creating highly efficient primary characteristic functions, and manufacturers should create more designs based on particular desires of various demographics based on age, gender, and origin. Reflecting the preference factors of different groups enables manufacturers to produce new design models which appeal more specifically to these constituents. In the future research, sales numbers and prices of released products will be investigated as factors that affect market share, and further questionnaire data will be collected to determine consumers' preferred features.

Manufacturers of mobile phones are pursuing a functional design which can be used easily by consumers. Even though manufactures of mobile phones obtain highly advanced technology, they do not always meet consumers' satisfaction. Doing so more effectively could increase market share. Because consumers are not likely to use a number of new functions, easy-to-handle functions are essential. As a function of relating to senses, a phone's User Interface (UI) becomes an increasingly important factor that influences competitiveness in the marketplace. Because these full touch phones do not have significant differences with respect to design form, comfortable and intuitive user interfaces can lead to improved overall design and consumer satisfaction. Recently, the popularity of sensitive touch screen phones has led to the development of 3D User Interface. This interface shows how to use mobile phones automatically through more intuitive three dimensional screens than previous two dimensional plane screens. In terms of industry research, the development of sensor technology such as sound, light, temperature, and pressure will lead to even more innovative products in the near future.

In the product design process, designers, engineers, marketers and psychologists can maintain a balance between design and usability by communicating both internally and with customers, and also by determining important design factors of a product that distinguishes it from others. The top design decision maker can overcome several difficult limitations and develop innovative and creative products by eliminating unimportant factors. The methods of this study will also help save manufacturing time in the research and development phase and increase consumers' satisfaction and the perceived value they find in a product. This study demonstrates how investments should be directed in the next product design process.

Bibliography

Agarwal, M., Cagan, J., and Constantine, K.G., 1999, Influencing generative design through continuous evaluation: Associating costs with the coffeemaker shape grammar, *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM*, Vol. 13, Issue 4, p 253-275

Alcantara, E., Artacho, M.A, Gonzalez, J.C., and Garcia, A.C., 2005, Application of product semantics to footwear design. Part I - Identification of footwear semantic space applying differential semantics, *International Journal of Industrial Ergonomics*, Vol. 35, Issue 8, p 713-725

Anderberg, M.R., 1973. Cluster Analysis for Applications: Probability and Mathematical Statistics, Academic Press, New York.

Andree, P., Daniel, B., Gerard, M., and Louise, C., 1990, Pink or blue Environmental gender stereotypes in the first two years of life, *Sex Roles*, Vol. 22, Issue 5-6, p 359-367

Bangle, C., 2001, The Ultimate Creativity Machine: How BMW Turns Art into Profit, *Harvard Business Review*, Vol. 79, Issue 1, p 4-11

Barone, S., Lombardo, A., and Tarantino, P., 2007, A weighted logistic regression for conjoint analysis and kansei engineering, *Quality and Reliability Engineering International*, Vol. 23, Issue 6, p 689-706

Bloch, P.H., 1995, Seeking the Ideal Form: Product Design and Consumer Response, *The Journal of Marketing*, Vol. 59, Issue 3, p 16-29

Brown, T., 2008, Design Thinking, *Harvard Business Review*, Boston: June, Vol. 86, Issue 6, p 84-92

Chang, C.C., 2008, Factors influencing visual comfort appreciation of the product form of digital cameras, *International Journal of Industrial Ergonomics,* Vol. 38, Issue 11-12, p 1007-1016

Chang, H.C., Lai, H.J, and Chang, Y.M., 2006, Expression modes used by consumers in conveying desire for product form: A case study of a car, *International Journal of Industrial Ergonomics*, Vol. 36, Issue 1, p 3-10

Chuang, M.C., Chang, C.C., and Hsu, S.H., 2001, Perceptual factors underlying user preferences toward product form of mobile phones, *International Journal of Industrial Ergonomics*, 27, 247–258

Clark, K., and Smith, R., 2008, Unleashing the Power of Design Thinking, *Design Management Review*, Boston: Summer, Vol. 19, Issue 3, p 8-17

Demirbilek, O. And Sener, B., 2003, Product design, semantics and emotional response, *Ergonomics*, Vol. 46, Issue 13-14, p 1346-1360

Desbarats, G., 2005, A Holistic Approach to Delivering Ergonomic Innovation, *Design Management Review*, Boston: Fall, Vol. 17, Issue 4, p 39-48

Dittmar, M., 2001, Changing Color Preference with aging: A comparative study on Younger and Older Native Germans Aged 19-90 Years, *Gerontology*, Vol.47, Issue 4, p 219-226

Du, X., Jiao, J., and Tseng, M.M., 2006, Understanding customer satisfaction in product customization, *International Journal of Advanced Manufacturing Technology*, Vol. 31, Issue 3-4, p 396-406

Dubberly, H., 2008, Learning Curves for Design, *Interactions,* New York: Jun/August. Vol. 15, Issue 4, p 13-17

Emory, W. C. and Cooper, R. D., 1991, Business Research Methods 4th Edition, ISBN-0-256-09265-6

Francis, E.H. and Gu, J. 2003, A methodology for evolutionary product design, *Engineering with computer*, Vol. 19, p 160-173

Fu, K. and Geng, D., 2007, using the data mining approach to determine the product preferences of target customers, *Proceedings - ICSSSM'07: 2007 International Conference on Service Systems and Service Management*

Geraldi, P., Kowalski, B.R., 1986, Partial least-squares regression: a tutorial. Analytica Chimica Acta, 185, 1–17

Greenwood, P.E. and Nikulin, M.S., 1996, A guide to chi-squared testing. Wiley, New York, ISBN 047155779X

Han, S.H. and Yang, H.C., 2004, screening important design variables for building a usability model: genetic algorithm-based partial least-squares approach. *International Journal of Industrial Ergonomics*, 33, 159–171

Han, S.H., Yun, M.H., Kim, K., and Kwahk, J., 2000, Evaluation of product usability: development and validation of usability dimensions and design elements based on empirical models, *International Journal of Industrial Ergonomics*, 26, 477–488

Horn, Diana and Salvendy, Gavriel., 2006, Consumer-based assessment of product creativity: A review and reappraisal, *Human Factors and Ergonomics In Manufacturing*, Vol. 16, Issue 2, p 155-175

Hosmer, D.W., 2000, applied logistic regression, 2th Edition, Wiley, New York

Hsiao, S.W. and Tsai, H.C., 2005, Applying a hybrid approach based on fuzzy neural network and genetic algorithm to product form design, *International Journal of Industrial Ergonomics*, Vol. 35, Issue 5, p 411-428

Hsiao, S.W. and Chou, J. R., 2004, A creativity-based design process for innovative product design, *International Journal of Industrial Ergonomics*, Vol. 34, Issue 5, p 421-443

Hsu,S.H., and Chunang, M.C., 2000, A semantic differential study of designers' and users' product form perception. *International Journal of Industrial Ergonomics,* 25, 375–391

Isiklar, G. and Buyukozkan, G., 2007, Using a multi-criteria decision making approach to evaluate mobile phone alternatives, *Computer Standards and Interfaces*, Vol. 29, Issue 2, p 265-274

Jiao, J. and Tseng, M.M., 2004, Customizability analysis in design for mass customization, *CAD Computer Aided Design*, Vol. 36, Issue 8, p 745-757

Jolliffe, I.T., 1986, Principal Component Analysis. *Springer*, New York.

Kelly, J. and Papalambros, P.Y., 2007, Use of Shape Preference Information in Product Design, *In the International Conference on Engineering Design*, Paris, France.

Kelly, J., Wakefield, G.H., and Papalambros, P.Y., 2006, The Development of a Tool for the Preference Assessment of the Visual Aesthetics of an Object Using Interactive Genetic Algorithms. *In the Generative Art Conference*, Milan, Italy, December 13-15

Kim, P. and Ding, Y., 2004, Optimal engineering system design guided by data-mining methods, 2004, *IIE Annual Conference and Exhibition*, p1285

Lai, H.H., Lin, Y.C., and Yeh, C.H., 2005, Form design of product image using grey relational analysis and neural network models, *Computers and Operations Research*, Vol. 32, Issue 10, p 2689-2711

Lee, H.C. and Tang, M. X., 2006, Generating stylistically consistent product form designs using interactive evolutionary parametric shape grammars, *2006 7th International Conference on Computer-Aided Industrial Design and Conceptual Design, CAIDC*, p 4127160

Li, H., Guo, H., Guo, H., and Meng, Z., 2008, Data mining techniques for complex formation evaluation in petroleum exploration and production: A comparison of feature selection and classification methods, *2008 Pacific-Asia Workshop on Computational Intelligence and Industrial Application*, Vol. 1, p 37-43

Lin, R., Lin, C.Y., and Wong, J., 1996, Application of multidimensional scaling in product semantics, *International Journal of Industrial Ergonomics*, Vol. 18, Issue 2-3, p 193-204

Lin, Y.C., Yeh, C.H., and Lai, H.H., 2007, Consumer-oriented product form design based on fuzzy logic: A case study of mobile phones, *International Journal of Industrial Ergonomics*, Vol. 37, Issue 6, p 531-543

Lojacono, G. and Zaccai, G., 2004,The Evolution of the Design-Inspired Enterprise, *MIT Sloan Management Review.* Cambridge: Spring, Vol. 45, Issue 3, p 75-79

Mallows, C.L. 1973, Some Comments on Cp, *Technometrics*, 15, 661-675

Mariëlle, E. H. Creusen and Jan P. L. Schoormans, 2005**,** The Different Roles of Product Appearance in Consumer Choice, *Journal of Product Innovation Management*, 22, 63-81

Matsubara, Y. and Nagamachi, M., 1997, Hybrid Kansei Engineering System and design support, *International Journal of Industrial Ergonomics,* Vol. 19, Issue 2, p 81-92

McLachlan, G. 2004, Discriminant Analysis and Statistical Pattern Recognition, Wiley Interscience, ISBN 0471691151

McCullagh, K., 2006, Strategy for the Real World, *Design Management Review,* Boston: Fall. Vol. 17, Issue 4, p 48-56

Montgomery, D.C., 1991, Design and Analysis of Experiments, Third Edition, John Wiley & Son

Montgomery D.C. and Peck E.A., 1982, Introduction to Linear Regression Analysis, John Wiley & Sons

Nancy, E., Bridget F., Georgia, F., and Susanna, H., 2006, Business 2.0 & Frog Design Present the 2nd Annual Bottom Line Design Awards, *Business 2.0.* San Francisco: Apr, Vol. 7, Issue 3, p113-117

Nagamachi, M., 1995, Kansei engineering: a new ergonomic consumer-oriented technology for product development, *Journal of Industrial Ergonomics,* Vol. 15, Issue 1, p 3-11

Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W., 1996, Applied Linear Statistical Model, 4th Edition. R.D. Irwin, Boston.

Orsborn, S., Cagan, J., and Boatwright, P., 2008, A methodology for creating a statistically derived shape grammar composed of non-obvious shape chunks, *Research in Engineering Design*, Vol. 18, Issue 4, p 181-196

Pedhazur, E.J. 1997, Multiple Regressions in behavioral research, 3th Edition, New York: Wadsworth

Peters, G., 2007, Aesthetic primitives of images for visualization, *11th International Conference Information Visualization*, p 316-325

Petiot, J., and Grognet, S., 2002, A multidimensional scaling approach for product design and preference modeling, *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Vol. 4, p 54-59

Pugh, S., 1991, Total design: Integrated methods for successful product engineering, 2nd ed. *Addison- Wesley*, New York

Pugliese, M. J. and Cagan. J., 2002, Capturing a rebel: Modeling the Harley-Davidson brand through a motorcycle shape grammar, *Research in Engineering Design*, Vol. 13, Issue 4, 139-156

Pullman, M.E., Moore, W.L., and Wardell, D.G., 2002, A comparison of quality of functional deployment and conjoint analysis in the new product design, *Journal of Product Innovation Management*, Vol. 19, Issue 5, p 354-364

Ravasi, D. and Lojacono, G., 2005, Managing design and designers for strategic renewal, *Long Range Planning*, London: Feb. Vol. 38, Issue 1, p 51-77

Rawsthorn, A., 2009, Good or Bad? The verdict(s), *New York Daily News*, New York: Jan 30, 2009. p.3

Saaty, T.L., 1980, The Analytic Hierarchy Process, *McGraw-Hill*, New York.

Shieh, M.D. and Yang, C.C., 2008, Classification model for product form design using fuzzy support vector machines, *Computers and Industrial Engineering*, Vol. 55, Issue 1, p 150-164

Thomke, S. and Feinberg, B., 2009, Design Thinking and Innovation at Apple, *Harvard Business Review*, Prod. #: 609066-PDF-ENG, p.12

Veryzer, R.W., 1993, Aesthetic Response and the influence of Design Principles on Product Preferences, *in Advances in Consumer Research*, Vol. 20, p 224-228

Veryzer, R.W., 1995, The Place of Product Design And Aesthetics In Consumer Research, *in Advances in Consumer Research,* Vol. 22, p 641-645

Wang, J., Zhang, J., and Wei, X, 2006, Evolutionary Muti-Objective Optimization Algorithm with Preference for Mechanical Design, *Advances in Machine Learning and Cybernetics - 4th International Conference*, Vol. 3930, p 497-506

Wang, K.C. and Chen, S.M., 2007, Product form design using ANFIS-Kansei engineering model, *2007 IEEE International Conference on Systems, Man, and Cybernetics, SMC* , 2034-2043

Wang, J., 2002, Improved engineering design concept selection using fuzzy sets, *International Journal of Computer Integrated Manufacturing*, Vol. 15, Issue 1, p 18-27

Waters, R., 2007, Design with Heart, *Design Management Review*. Boston, Vol. 18, Issue 1, p 16-23

Wilks, D.S., 1995, Statistical methods in the atmospheric sciences: Academic Press, p.467

Yamamoto, M. and Lambert, D.R., 1994, The Impact of Product Aesthetics on the Evaluation of Industrial Products, *Journal of Product Innovation Management*, Vol. 11, Issue 4, p 309-324

Yan, J. and Li, W., 2008, Group technology based feature extraction methodology for data mining, *5th International Conference on Fuzzy Systems and Knowledge Discovery*, Vol.2, p 235-239

Yoffie, D, B. and Slind, M., 2008, Apple Inc, *Harvard Business Review*, Prod. #: 708480-PDF-ENG, p.32

Yun, M. H., Han, S. H., Hong, S. W. Kim. J.S., 2003, incorporating user satisfaction into the look-and-feel of mobile phone design, *Ergonomics*, Vol. 46, Issue 13, p 1423 – 1440

Zhang, Q., Lu, C., Wu, T., and Yu, S., 2006, A product-form aesthetic evaluation system, *2006 7th International Conference on Computer-Aided Industrial Design and Conceptual Design*, *CAIDC*, p 4127099

Zhu, Y., Luo, S.J., and Xu, X.H. 2006, Research on designers' product color perception of mobile phone.*7th International Conference on Computer-Aided Industrial Design and Conceptual Design*, November 17-19, 1-5

Appendix A: Survey Form

0. What is your origin?
a. American b. Asian

1. What is your age?
a. 14-17 b. 18-22 c. 23~29 d. 30-39

2. What is your gender?
a. Male b. Female

3. What is your education level?
a. High school student
b. Only high school diploma
c. Undergraduate student
d. Graduate student with Bachelor degree or Master degree
e. Bachelor Degree
f. Masters Degree
g. Ph.D. Degree

4. What factor do you consider **as most** important when purchasing a mobile phone?
a. Design b. Price c. Functions d. Brand e. Easy user interface f. other __

5. What is **the most significant design** issue in mobile phones in your opinion?
a. Weight b. Body color c. Keypad design d. Screen size e. Form f. Texture

6. What mobile phone form do you prefer?


a. Clamshell folder


b. block


c. Slider


d. Folder (Flip)

7. What are **the most essential** 5 functions of mobile phones, in your opinion?
(Please indicate the ranking for first 5)

Function	**Rank 1~5** **1** most important
a. 3G Network	
b. Battery Life Time	
c. Bluetooth	
d. Game	
e. Memory capacity	
f. MP3 Player	
g.GPS	
h. Multimedia	
i. Camera	
j. Instant messaging	
k. Sensitive sensor	
l. Call log Memory	
m. E-mail	
n. Text messaging	
o. Other _________	

Function	Descriptions
3G Networks	Represent an international standard for wide-area cellular networks that are replacing 2G networks.
Instant Messaging	The function that sends real time messages to another mobile phone user.
Bluetooth	A wireless hand free kit that uses low power radio communications over short distances from fixed and mobile
Sensitive sensor	Sensor function that has Touch sensitive scroll wheel, Touch screen, or QWERTY Keypad
GPS	Global Positioning System (GPS) is a satellite-based navigation system to help accurately determine their locations
Multimedia	it can mean various mobile internet services such as shopping, game, and video with your phone browser.

8. Recently, manufacturers are releasing the mobile phones that emphasize specific function based on consumers' requirements. Please choose mobile phone that you would like to purchase **the most.**

1. *mobile phone that has good MP3 player function*
2. *mobile phone that has large memory capacity*
3. *mobile phone that has decent instant messenger, cam messenger, or text message*
4. *mobile phone that has sensitive function like Touch screen or QWERTY key pad*
5. *mobile phone that has exciting game feature*
6. *mobile phone that has good camera quality like similar to digital camera*
7. *mobile phone that has easy user interface menu organization*

9. Please pick only one built in function that you personally use **the most**
a. Text messaging   b. phone call   c. MP3 Player   d. game   e. camera   f. E-mail

10. On average, how often are you using the functions listed below?

① *Text messaging*	*a. 0-5*	*b. 6--10*	*c. 11-15*	*d. 16-20*	*e. over 20*
② *Phone call*	*a. 0-5*	*b. 6--10*	*c. 11-15*	*d. 16-20*	*e. over 20*
③ *Instant Messaging*	*a. 0-2*	*b. 3-5*	*c. 6-9*	*d. 10-13*	*e. over 13*
④ *Camera*	*a. 0-2*	*b. 3-5*	*c. 6-9*	*d. 10-13*	*e. over 13*
⑤ *Mp3 player*	*a. 0-2*	*b. 3-5*	*c. 6-9*	*d. 10-13*	*e. over 13*
⑥ *Game*	*a. 0-2*	*b. 3-5*	*c. 6-9*	*d. 10-13*	*e. over 13*
⑦ *E-Mail*	*a. 0-2*	*b. 3-5*	*c. 6-9*	*d. 10-13*	*e. over 13*
⑧ *Multimedia*	*a. 0-2*	*b. 3-5*	*c. 6-9*	*d. 10-13*	*e. over 13*

11. Please write down your mobile phone model and wireless company
(For example- LG chocolate, Sprint) ________________________

Appendix B: Calculating Confidence Intervals for Nominal Logistic Regression

The formula that Minitab uses to calculate confidence intervals in nominal logistic regression is given in the Methods and Formulas section of Help.

The confidence interval for each coefficient ß is

$$ß_{kj} \pm Z_{\alpha/2} * \text{(standard error)} \quad (1)$$

where z measures the units of standard deviation from the mean under the normal curve
(a table z-values is included in most statistical textbooks).

To obtain the confidence interval for the associated odds ratio, you take the exponential function of the above formula:

$$e^{ß_{kj} \pm Z_{\alpha/2} * \text{(standard error)}} \quad (2)$$

For example, consider the partial output from the final model for Logit 1 shown below
(See Figure 3-13 for the full output results).

Logistic Regression Table

Predictor	*Coef*	*SE Coef*	*Z*	*P*	*Odds Ratio*	*95% CI Lower*	*95% CI Upper*
Logit 1: (easy user interface/Brand)							
Constant	*-2.21633*	*0.717231*	*-3.09*	*0.002*			
Origin							
USA	*2.62528*	*0.483843*	*5.43*	*0.000*	*13.81*	*5.35*	*35.65*

Using formula (2) above, with $\alpha = 0.05$, the 95% confidence interval for
the odds ratio (13.81) associated with Origin-USA variable in Logit 1 is calculated as follows:

$$e^{2.62528 \pm z_{0.025} * 0.483843} = e^{2.62528 \pm 1.96 * 0.483843} = e^{2.62528 \pm .94833228}$$

$$= (e^{1.67694772}, e^{3.57361228}) = (5.35, 35.65)$$

Notice that this matches the 95% CI lower and upper bounds shown in the Minitab output above.

To calculate the 98.3% CI for each odds ratio, simply substitute a value of $\alpha = 0.017$
in formula (2). Thus, for the Origin-USA variable in Logit 1, the 98.3% confidence interval is:

$$e^{2.62528 \pm z_{0.0085} * 0.483843} = e^{2.62528 \pm 2.39 * 0.483843}$$
$$= (e^{1.46889523}, e^{3.78166477}) = (4.34, 43.89)$$

As expected, the 98.3% confidence interval is wider than the 95% confidence interval
because the alpha value is lower and thus the confidence interval is more conservative.