AN ABSTRACT OF THE THESIS OF

Shuo Naini (Shuo Wang) for the degree of Master of Science in Food Science and Technology presented on May 9, 2003.

Title: A Simulation Tool for Evaluating Sensory Data Analysis Methods.

Abstract approved:

Mina R. McDaniel

In cross-cultural studies, respondents from specific cultures may have different product preferences and scale usage. Combining data from different cultures will result in departures from the basic assumptions of analysis of variance (ANOVA) and loss of power in testing capability of finding product and culture differences. However, the result of violations on power of ANOVA is unknown by sensory researchers. The objectives of this research were by simulating consumer product evaluation data, to evaluate the robustness and testing power of ANOVA under different cross-cultural situations.

The study was conducted in two parts. First, an Empirical Logit simulation model was employed for generating sensory data. This model included respondent, product, consumer segment and product by segment interaction effects. Four underlying distributions: Binomial, Beta-Binomial, Hypergeometric, and Beta-

Hypergeometric were used to increase or decrease the dispersion of the responses. Alternatively, instead of using these four distributions, the same applications were achieved by a binning step. The entire simulation procedure including the Empirical Logit model and the binning step was called Discrete Empirical Logit model. In the second part of the study, the Discrete Empirical Logit model was chosen to generate specified data sets under six different cross-cultural cases. After analyzing these data sets by ANOVA reduced and full models, the empirical power of ANOVA under different cases was calculated and compared.

The results showed that both Beta-Hypergeometric and Discrete Empirical Logit were flexible on simulating sensory responses, but the Discrete Empirical Logit was relatively simple to use. Comparing with the ANOVA reduced model, the full model gave better information on evaluating the case that segments differ in product preferences. This suggested segmentation was very important in cross-cultural data analysis. Under the situations that sample sizes were equal and respondents performed consistently within segment (MSE $\approx$ 1), ANOVA was very robust to different scale usage, losing at worst 18% in power.

From the scope of this study, we recommend using the ANOVA full model in the cross-cultural research. Results from different cultures could be combined when consistency within segments was high.

A SIMULATION TOOL FOR EVALUATING SENSORY DATA ANALYSIS METHODS

by

Shuo Naini

A THESIS

submitted to

Oregon State University

in partial fulfillment of

the requirement for the

degree of

Master of Science

Presented May 9, 2003

Commencement June, 2003

Master of Science thesis of Shuo Naini presented on May 9, 2003.

APPROVED:

___

Major Professor, representing Food Science and Technology

___

Chair of Department of Food Science and Technology

___

Dean of Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

___

Shuo Naini, Author

ACKNOWLEDGEMENTS

The first acknowledgement must be given to my professor, Dr. David Lundahl, who first introduced me to Sensory Evaluation and Statistics. When I first came to Corvallis, he and his family welcomed me with a happy cozy home. It is him who gave me guidance and constant support, who understood me and challenged me. Also it is him who encouraged me to go for my second degree – Statistics.

I also wish to express my sincere gratitude to Dr. Dave Plaehn, for sharing the great work he had done. His contributions to this study are immeasurable and his friendship is and will always remain in my heart.

I would like to thank Dr. Mina McDaniel for dedicating her time to help me editing the thesis. Her encouragement and kindness has given me a tremendous positive influence. Also, I am grateful for all the supports and advise I received over the last few years from the sensory group.

I would like to thank Dr. Dave Butler for his statistical and programming guidance. Thanks also goes to Dr. Annie Qu, Dr. Cliff Pereira and Dr. Park from the Statistics Department, my committee members, and all those in Food Science and Technology Department, Statistics Department and graduate school who have shared their expertise, time and patience to help me complete this program.

Words will never be strong enough to express my gratitude to my parents, my sister and my husband Kaveh, for their great support and unconditional belief in me, and their sacrifices for my graduate studies.

CONTRIBUTION OF AUTHORS

Dr. David Lundahl was guiding the whole study and involved in the design of sensory experiment and writing of the manuscript. Dr. Dave Plaehn was involved in the development of models, programs and the writing of the manuscript. Dr. Mina McDaniel was involved in the writing of the manuscript.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

LIST OF FIGURES

LIST OF FIGURES (CONTINUED)

LIST OF FIGURES (CONTINUED)

LIST OF FIGURES (CONTINUED)

LIST OF TABLES

LIST OF TABLES (CONTINUED)

A SIMULATION TOOL FOR EVALUATING SENSORY DATA ANALYSIS METHODS

CHAPTER 1

THESIS INTRODUCTION

As a well-known traditional statistical method, analysis of variance or ANOVA has been broadly applied in the sensory field, generally to study the differences between products (Gacula, 1993). It has come to be associated with research designs, mathematical models, and analytical tools (Edwards, 1993). The validity of estimates and tests of hypotheses for analyses derived from the linear model rests on the merits of several key assumptions. The random experimental errors are assumed to be independent, normally distributed with a mean of zero, and have a common variance for all treatment groups (Kuehl, 2000). Applied to sensory analyses on scaled data, the assumptions are implicit that the scores are independent, have equal random variance, are normally distributed, and on the same scale of measurement (Gay and Mead, 1992). Any disagreement between the data and one or more of these assumptions affects the estimates of the treatment means and tests of significance from the analysis of variance.

It should be stressed that in cross-cultural studies, ideal conditions are seldom realized. Respondents from specific cultures may have different product preference and scale usage (Bertino and Chan, 1986; Prescott et al., 1992, 1997, 1998; Wilkinson and Yuksel, 1997; Ayabe-Kanamura et al., 1998; Yeh and

Lundahl, 1999; Cox et al., 2001). For example, one or more consumer segments (or groups represent different cultures), while sharing a common perception of the differences between products, consistently use some portion of the scale instead of the whole. This can cause the distribution of the data to be centered ("central tendency") (Amerine et al., 1965) or skewed ("dislike avoidance"). On the other hand, segments may share the same scale but disagree in the product preference ("crossover effect"). Furthermore, the disagreement among segments results in heterogeneous segment location variation; within a segment, the acceptance rating differences among respondents causes heterogeneous acceptance variation. All these cross-cultural cases may lead to non-normal distributed consumer responses with unequal variances and therefore violate ANOVA assumptions.

In cross-cultural research, there may be a need to combine data from the respective different responding cultures. But if above cases exist, the power of ANOVA to detect product differences may be decreased. As far as we know, Monte Carlo studies (statistical simulation studies that utilizes sequences of random numbers to perform the simulation to show how *p*-values for a statistical test or estimation method functions) suggest the ANOVA F tests are relatively robust as far as breaking these assumptions is concerned. However, the degree of robustness is unclear in sensory research (O'Mahony, 1986), especially when combining data from different cultures.

Due to the importance of ANOVA in analyzing cross-cultural data, a research project was initiated to evaluate the robustness of this analysis and testing power to various extraneous sources of cross-cultural related errors in measurement. The study was conducted in two parts. The first part involved the development of a generalized simulation tool for categorical sensory response. The second part applied this simulation tool to generate many data sets such that empirical power calculations could be conducted under various cases where cross-cultural differences exist.

The generalized simulation tool is based upon Thurstonian principles of psychophysics developed by Leon Louis Thurstone. As Moskowitz mentioned (1983): "Thurstonian methods utilize these principles and represent a psychological approach to sensory analysis and hedonics. They joint together psychological measurement principles with real-world stimuli." Thurstone hypothesized that when people evaluate the acceptability of stimuli, they do so based upon an underlying or internal psychological scale of hedonics or liking. For example, direct preferences by paired comparison test where the individual chooses one stimulus over another, represents the outcome of the use of this underlying psychological scale. The closer the preference rating lies to 50/50 (no preference) the more likely that the stimuli lie near each other on the underlying psychological scale of liking (Moskowitz, 1983).

Let's imagine a panelist tasting a product, by Thurstonian principles this panelist's internal response is based on a psychological scale that is evoked by sensory stimuli. This internal scale is translated to an external response on a categorical scale. In this study, in order to model this process, a random parameter, *p*, which locates on a psychological standardized scale from 0 to 1, was generated using an empirical logit simulator. This random variate was then transformed to discrete random data on a sensory 9-point hedonic scale. Four underlying statistical distributions – binomial, beta-binomial, hypergeometric, beta-hypergeometric were applied to increase or decrease the dispersion of category interval responses by controlling the input parameters (*p*, number of scale categories and truncation parameters). Alternatively, data were also simulated using the Discrete Empirical Logit model, simply a transformation of the parameter *p*. Comparing with the other four models, our study indicated that the Discrete Empirical Logit model was the most flexible, easily capable of producing the designed cases for the next step of the research.

In the second part of the study, the Discrete Empirical Logit model was chosen to generate the product evaluation data under a simplified situation – 2 segments with 100 respondents in each, evaluating 2 products. Six common cases for cross-cultural differences were studied, such as on scale usage ("central tendency" and "dislike avoidance" cases), on product preference ("crossover" and "segment effect"), heterogeneity within and among segments. A truncation procedure was employed to mimic central tendency and dislike avoidance cases by modifying the range of the scale being used. In each case, 10000 data sets combining the responses from the two segments were analyzed by ANOVA using a reduced and full model. The empirical power of ANOVA under 10000 simulations was calculated as the percentage of times when significant product differences were detected.

This study applied a brand new way to address difficult issues in sensory data analysis. A generalized simulation tool was proposed and then applied to a critical problem in the field of sensory evaluation that could not be solved in any other way. As a result, this study demonstrated the power of simulation which can be applied to a wide range of sensory evaluation problems where there are questions as to what is the most appropriate analysis.

CHAPTER 2

LITERATURE REVIEW

Cross-Cultural Sensory Studies

Cross-cultural product development is becoming increasingly important as companies strive to compete in an international marketplace. The importance of cultural patterns in determining food preferences and tastes must be stressed (Amerine et al., 1965). Cultural influence may play a substantial role on consumer decision-making. In recent years, more and more attention has been given to cross-cultural studies on consumer scale usage. Some psychologists suggested that culture rules reasoning styles. For instance, East Asians take a "holistic" approach to reasoning tasks. They make little use of categories and formal logic, and instead focus on relations among objects and the context in which they interact. They direct their attention into a complex, conflict-strewn environment. While people in the United States adopt an "analytic" perspective. They look for the traits of objects while largely ignoring their context, categorize items by applying formal logic and explicit rules, and try to resolve any contradictions that turn up (Bower, 1999).

For both strategic and applied research, knowledge of scale use and scale differences would be extremely valuable in understanding cross-cultural consumers' response behavior. In the study of Wilkinson and Yuksel (1997), they introduced that respondents may differ in their use of the measurement scale, both in location (where on the scale the scores tend to be located) and in dispersion (the range of the scale used). Thus, some respondents will use the whole available range for scoring while others will concentrate on one part of the scale (differences in dispersion). Moreover, some respondents may primarily score in the lower part of the scale while others may score largely in the middle or upper region (differences in location) (Næs, 1990).

Bertino and Chan (1986) investigated relationships between taste perception and diet in individuals with Chinese and European ethnic backgrounds. The Chinese subjects assigned higher pleasantness ratings to higher concentrations of sucrose in water and showed a tendency to rate sucrose in cookies as tasting more pleasant. They also assigned higher pleasantness ratings to and preferred higher concentrations of salt in crackers. However, they suggested that, it was possible that the Chinese, in an attempt to be polite, assigned higher pleasantness ratings to all taste stimuli.

Prescott et al. conducted a series of studies between Japanese and Australians, e.g. comparing the hedonic responses to taste solutions (1992), and the responses to manipulations of sweetness in foods (1997). In 1992, they found that one or both groups were not using the scale end points. In a lately study, Prescott (1998) found there was little evidence for cross-cultural influences on respondents' assessment behavior, such as scale usage.

In the study of odor perception between Japanese and German subjects by Ayabe-Kanamura et al. (1998), differences in pleasantness ratings between the two populations were explained partly due to differences in the use of the rating scale, with the Germans tending to give more extreme ratings.

The psychological error of "central tendency" (Amerine et al., 1965) is frequently observed in scoring when the extreme values of a scoring scale are seldom used. Yeh and Lundahl (1999) compared 9-point hedonic scale usage between consumers from the US and Pacific Rim cultures (Chinese, Koreans, and Thais). The 9-point hedonic scale was translated directly from English to their respective languages. The results indicated that Pacific Rim respondents used the 9-point hedonic scale differently from American respondents; they tend to agree with each other ("central tendency") and do not use extremely low scores ("dislike avoidance") such as Thais. The study posed that different scale usage may result in less statistical power to detect differences among products. They also found convincing evidence of adaptive behavior among the Pacific Rim respondents to respond more like Americans with longer time in residence in the US.

In 2001, Cox, Clark and Mialon conducted a study to test whether there were cultural, scale and gender interactions between European-origin Australian (n = 61) and Malaysian (n = 54) consumers' hedonic responses to food and drink stimuli. One group was using a labeled 9-point hedonic category scale and the other used an unstructured-anchored line scale, both using computerized responses. The result found no systematic cultural bias except the gender bias. They concluded that the result was consistent with the concerns expressed by Yeh and Lundahl (1999) over Asian avoidance of category scale extremes, therefore a preference for line scales was cautiously suggested.

Power of Analysis of Variance (ANOVA)

Analysis of Variance (ANOVA) is a common data analysis method used in sensory studies. Power is the probability of rejecting the null hypothesis when it is false. It gives a method of discriminating between competing tests of the same hypothesis, the test with the higher power being preferred (Everitt, 1998). Therefore, the power of ANOVA represents the chance of rejecting a null hypothesis when a difference in the population actually exists. Empirically, the power of ANOVA can be estimated as the proportion of times when a significant product difference is detected.

Sources of Variation of ANOVA and Experimental Design

ANOVA has been the statistical tool most often employed to separate the total variation of sensory data into sources that affect sensory responses. These sources of variation usually include among respondent, among treatment and respondent by treatment interaction effects. More complicated experimental designs may include other effects or factorial treatment sets as possible sources of variation (Lundahl and McDaniel, 1988).

The major sources of variation in sensory data are often due to the respondents and treatments. The respondent source of variation is from respondents' use of different parts of the rating scale. Differences among respondents in response patterns to treatments contribute to the respondent by treatment interaction source of variation. Experimental designs including this source of variation test the inconsistency of the panel in evaluating samples (Lundahl and McDaniel, 1988).

In the experimental design, the question of whether we should treat the respondent effect as random or fixed has received some attention in sensory science literature (O'Mahony 1986; Lundahl and McDaniel 1988; Lawless 1997; Næs and Langsrud 1998). Naïve (inexperienced) consumer respondents, for example, are always random because they must, by definition, relate to the population of consumers (Lundahl and McDaniel, 1988). In our study, the respondent effect was treated as random since the consumer respondents were simulated to represent randomly selected individuals from different cultures.

ANOVA Assumptions

It is worth considering how the violation of ANOVA assumptions will affect its power. O'Mahony (1986) summarized the four fundamental assumptions that relate to the two-way ANOVA design, specifically in sensory scaling data analysis. They are:

1. Samples must be randomly picked from their respective populations.
2. The scores within a treatment must be independent.
3. Samples of scores under each treatment must come from normally distributed populations of scores and this implies that data must be on an interval or ratio scale.
4. Samples of scores under each treatment must come from populations with the same variance (homoscedasticity).

Wilcox (1993) has emphasized that violating the usual equal variance assumption as well as the usual assumption of normality can have serious consequences in terms of both Type I errors and power. On the other hand, based on the practical Central Limit Theorem, ANOVA is robust against small departures from a particular assumption. Therefore, inferences are valid even when some assumptions are not met. Meanwhile, Monte Carlo simulation study suggests that ANOVA is generally robust to moderate violations of the assumption of normality, particularly when sample sizes are equal (Glass, Peckham and Sanders, 1972).

O'Mahony (1986) discussed that using unequal-interval scales results in the categorical scaled data breaking the assumption of normality. Homoscedasticity can also be a problem. For example, the spread or variance of the sample will not be as great at the end of the scale as in the center. Thus, to compare the mean values from the center and from the end of the scale will be to compare means of samples drawn from populations with different variances. Furthermore, the skewing of the distribution for a mean at the end of the scale means that it is no longer normal.

In cross-cultural studies, if we want to combine data from different responding cultures, it is not known how much power of ANOVA to detect product differences will be eliminated. Further, we currently do not know the degree of violation of these ANOVA assumptions that can be tolerated. Those are the questions we wish to answer by this study.

Simulation Studies

To evaluate the robustness and power of ANOVA under different situations, different data sets need to be collected as inputs. The problem arises when there is just little or no data available – the real consumer data required may not exist, or it may be difficult or costly to obtain from different cultures within a limited time period. If we could effectively employ an input data processor to simulate data as from real-world cases, the effect of changing certain input parameters and extraneous errors in measurement on the power of ANOVA will be easily observed.

Simulation studies are frequently used when investigating expected measurement outcomes and examining expected statistical outcomes (Bang et al., 1998). Simulation makes relative ease with which samples can often be generated from a probability distribution, even when the density function cannot be explicitly integrated. In performing simulations, it is helpful to consider the duality between a probability density function and a histogram of a set of random draws from the distribution. Given a large enough sample, a histogram can provide practically complete information about the density and in particular, various sample moments, percentiles, and other summary statistics provide estimates of any aspect of the distribution, to a level of precision that can be estimated (Gelman et al., 1997). Usually Monte Carlo simulation is an approach to examine expected statistical outcomes by using random-number generators. For example, one can use Monte Carlo to show how p-values vary for a statistical test or estimation method function. It can also be applied to make inferences about the population from which a sample has been drawn when no convenient real data exist (Kelly, 1999). However, only a few simulation studies have been applied in the sensory field.

Lundahl (1992) used a simulation strategy to determine the influence of special consumer groups as a small subset (10 to 30%) of respondents on the outcomes of consumer acceptance tests. He concluded that a segment by treatment interaction with crossover patterns or minority acceptance can contribute to errors of type II or I, and large crossovers from 10% special consumers or small crossovers from 20 to 30% differently responding ("special") consumers can significantly increase type II error.

Næs and Langsrud (1998) advocated the use of the mixed ANOVA models to reduce the interaction effect by removing the scaling effect mathematically. They applied Monte Carlo simulation and power function to compute Type I errors and power of the F-tests for a number of cases with different use of the scale. Their conclusion is that with moderate differences among the assessors and with a realistic size of sensory panel (>10) the mixed model F-test for products can be used as usual without losing much information.

In the study of using a beta-binomial model to assess the results of difference testing methods, Ennis and Bi (1998) conducted Monte Carlo experiments to investigate the behavior of difference or preference tests with over-dispersed binomial data based on the binomial model and the beta-binomial (BB) model. They found when inter-trial variation exists due to the noise among judgments or/and samples, a compound distribution of the beta and binomial distributions will be a better fit for the over-dispersed binomial data.

Thurstonian Method

The generalized simulation tool developed in this thesis is based upon Thurstonian principles of psychophysics. Thurstonian methods, developed by Leon Louis Thurstone, represent a psychological approach to sensory analysis and hedonics. They joint together psychological measurement principles with real-world stimuli. Thurstone hypothesized that when people evaluate the acceptability of stimuli, they do so based upon an underlying or internal psychological scale of hedonics or liking (Moskowitz, 1983). As Thurstone explained (1927):

> "Suppose we are confronted with a series of stimuli or specimens such as a series of gray values, cylindrical weighs, handwriting specimens, or any other series of stimuli that are subject to comparison. The first requirement is a specification as to what it is that we are to judge or compare. It may be gray values, or weights, or excellence, or any other quantitative or qualitative attribute about which we can think "more" or "less" for each

> specimen. This attribute which may be assigned, as it were, in differing amounts to each specimen defines what we shall call the psychological continuum or scale for that particular project in measurement."

He defined or constructed, the psychological scale is the frequencies of the respective discriminal processes for any given stimulus form a normal distribution on the psychological scale (Thurstone, 1927). However, this involves no assumption of a normal distribution or of anything else. The psychological scale is at best an artificial construct. It is so spaced off that the frequencies of the discriminal processes for any given stimulus form a normal distribution on the scale (Thurstone). He also defined the separation on the scale between the discriminal process for a given stimulus on any particular occasion and the modal discriminal process for that stimulus is called the discriminal deviation on that occasion (Thurstone).

Observable behavior, such as direct preferences by paired comparison (where the individual chooses one stimulus over another), represents the outcome of the use of this underlying psychological scale. The closer the preference rating lies to 50/50 (no preference) the more likely that the stimuli lie near each other on the underlying psychological scale of liking (Moskowitz, 1983). In the sensory evaluation field, Thurstonian models assume the underlying responses to each sensory stimuli are each normally distributed, independent and with constant variance. Also, they assume that panelists use specific rules to come up with a response to the sensory difference test (Lundahl, 1997).

Empirical Logit Model

General Theoretical Background – Johnson System

To generate sensory consumer data with certain distributions, a guessing model is needed. Traditionally, the normal distribution had played a dominant role in both theoretical and applied statistics. However, it is apparent that the normal curve cannot provide an adequate representation of many of the distributions encountered in statistical practice. The constructed systems of frequency curves should be capable of representing a wider variety of distributions than those for which a normal curve would suffice (Johnson, 1949).

When faced with problem of summarizing many data sets by means of a mathematical function, a common practice is to use a flexible family of distributions, so called empirical distributions (Slifker and Shapiro, 1980). To accomplish this, often a family of distribution proposed by Johnson (1949), the so-called Johnson System, can be created.

The system contains three families of distributions, denoted the S_U, S_B, and S_L (log-normal) distribution respectively, which are generated by transformations of the form

$$z = \gamma + \eta k_i(x; \lambda, \varepsilon) \tag{2.1}$$

where i = 1, 2, 3 for S_U, S_B, and S_L respectively; z is a standard normal variable and $k_i(x; \lambda, \varepsilon)$ is chosen to cover a wide range of possible shapes. For S_B distribution,

$$k_2(x;\lambda,\varepsilon) = ln(\frac{x-\varepsilon}{\lambda+\varepsilon-x}) \tag{2.2}$$

The S_B distribution by nature bounded on $(\varepsilon, \varepsilon+\lambda)$, appears to be suited to represent certain classes of variables which have physical or natural constraints on their range (Mage, 1980). The Empirical Logit model introduced in this study is similar to the S_B.

Logit Transformation

The term logit or logistic is named for transformations (Christensen, 1990). With a two-category response variable, we will examine models for $log(p_1/p_2)$, where p_1 is the chance for one outcome and $p_2 = 1 - p_1$ for another. When these models are ANOVA-type models, they are often referred to as logit models. When these models are regression-type models, they are called logistic regression models (Christensen, 1990).

A logistic regression model describes how a binary (0 or 1) response variable is associated with a set of explanatory variables. It is advocated when the dependent variable is expressed as a proportion (p). So the logistic regression model specifies that a probability is related to a regression-like function of explanatory variables (Ramsey and Schafer, 1997). This model is used when relationships between the non-response factors (explanatory variables) are not of interest. It is implicit in the definition of a logit model that no structure between the explanatory variables is taken into account (Christensen, 1990).

If p is used for the population mean of binary response (0 or 1), the probabilities for the responses are modeled as a regression-like function of explanatory variables:

$$logit(p) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p \tag{2.3}$$

and $X_1, \cdots, X_p$ are the explanatory variables. The inverse of the logit function is called the logistic function. If $logit(p) = \eta$, then

$$p = \frac{exp(\eta)}{1 + exp(\eta)} \tag{2.4}$$

The logit transformation replaces the value p of the dependent variable by their transformed value *log(odds)* or $log(\frac{p}{1-p})$. This has the effect of allowing the dependent variable to take any value in the range $(-\infty, \infty)$ rather than be constrained to the range (0, 1). Thus the regression model may be better justified with a normally distributed random component (Cooper and Weekes, 1983).

Adding Dispersion to the Empirical Logit Model

The sensory literature discusses dispersion in terms of adding variation to a response with a binomial distribution. Below is a review of the literature as may apply to the problem of adding dispersion to a response based upon the Empirical Logit model.

Binomial Distribution

A binomial experiment has two parameters, n and p where a response Y consists of n identical, independent trials, and each trial results in one of two outcomes. The probability of success on a single trial is equal to p and remains the same from trial to trial (so, the probability of a failure is equal to $q = (1 - p)$. A random variable Y is said to have a binomial distribution based on n trials with success probability p if and only if

$$P(Y = y) = \binom{n}{y} p^y q^{n-y}, y = 0, 1, 2, \ldots, n \quad 0 \le p \le 1 \tag{2.5}$$

(Wackerly et al., 1996).

The mean and variance associated with a binomial random variable are

$$E(Y) = np,\ V(Y) = npq \tag{2.6}$$

Beta Distribution

A random variable Y is said to have a beta probability distribution with parameters α and β if and only if the density function of Y is

$$f(y) = \begin{cases} \dfrac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha, \beta)}, & \alpha, \beta > 0; 0 \le y \le 1 \\ 0, & \text{elsewhere} \end{cases} \tag{2.7}$$

where

$$B(\alpha, \beta) = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1}dy = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{2.8}$$

(Wackerly et al., 1996).

The mean and variance of Y from this Beta distribution are

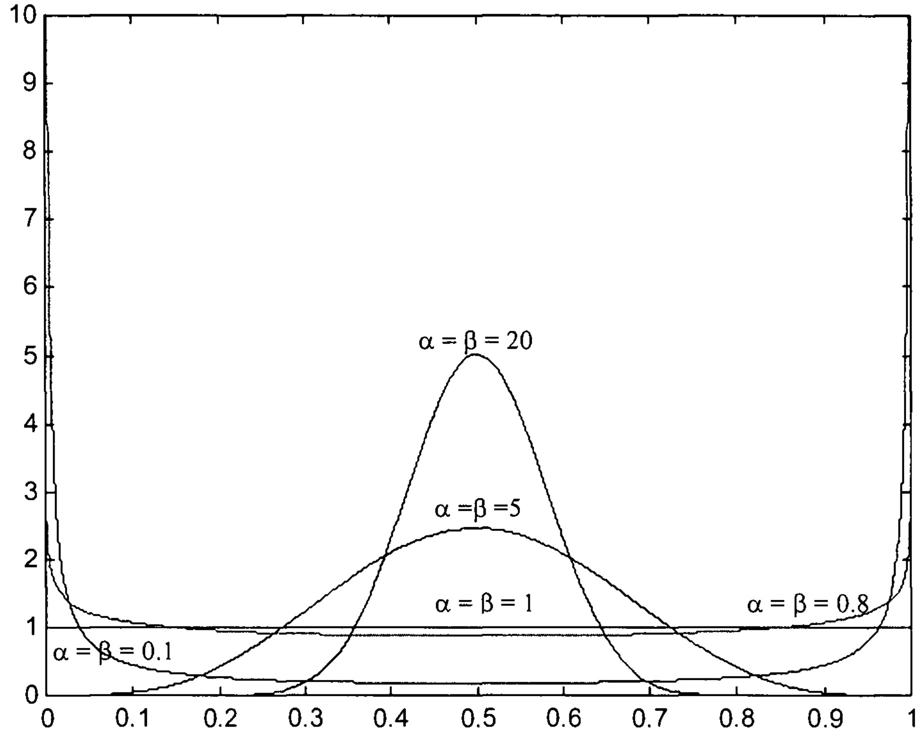
$$E(Y) = \frac{\alpha}{\alpha+\beta}, \; V(Y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{2.9}$$

As the parameters α and β vary, the beta distribution takes on many shapes. The probability density function (*pdf*) can be strictly increasing ($\alpha > 1$, $\beta = 1$), strictly decreasing ($\alpha = 1$, $\beta > 1$), U-shaped ($\alpha < 1$, $\beta < 1$) or unimodel ($\alpha > 1$, $\beta > 1$). The case $\alpha = \beta$ yields a *pdf* symmetric about $\frac{1}{2}$ with mean $\frac{1}{2}$ (necessarily) and variance $\frac{1}{4(2\alpha+1)}$. The *pdf* becomes more concentrated as α increases, but stays symmetric (Figure 2.1). Finally, if $\alpha = \beta = 1$, the beta distribution reduces to the uniform (0, 1) (Casella and Berger, 1990).

Reparameterizing by letting $p = \frac{\alpha}{\alpha+\beta}$, $q = 1 - p = \frac{\beta}{\alpha+\beta}$, and $\gamma = \frac{1}{\alpha+\beta+1}$, then the mean and variance are

$$E(Y) = p, \; V(Y) = pq\gamma \tag{2.10}$$

Figure 2.1 Symmetric beta densities.



Note: The case $\alpha = \beta$ yields a *pdf* symmetric about $\frac{1}{2}$ with mean $\frac{1}{2}$ (necessarily) and variance $\frac{1}{4(2\alpha+1)}$. The *pdf* becomes more concentrated as α increases, but stays symmetric. If $\alpha = \beta = 1$, the beta distribution reduces to the uniform (0, 1).

Beta-Binomial Distribution (Binomial-Beta Distribution)

The study of the beta-binomial (BB) distribution has received much attention in the past decade. It was first used by in the study of chromosomes (Skellam, 1948) and has been widely applied to behavior science, e.g. TV show loyalty (Sabavala and Morrison, 1977), TV schedules (Rust and Klompmaker, 1981), and marketing (Chatfield and Goodhardt, 1970, Morrison, 1979). In addition, the BB distribution has been used in disease incidence (Griffiths, 1973, Madden and Hughes, 1994), teratology (Williams, 1975, Yamamoto and Yanagimoto, 1994), and mutagenesis (Otake and Prentice, 1984) and has been studied by many biostatisticians.

Ennis and Bi (1998) applied the BB distribution to sensory data in replicated difference and preference tests. Commonly, sensory difference and preference tests are analyzed by the binomial test under the assumptions that choices made by the respondents are independent and choice probabilities do not vary from trial to trial. However, when inter-trial variation (variations due to judgments and samples) exists, the assumption of a binomial model is violated. An alternative is to model the variation in inter-trial choice probabilities with a beta distribution to fit the over-dispersed binomial data.

The BB distribution is a compound distribution of the beta and the binomial distributions. It is a natural extension of the binomial model. Recall that if a random variable Y follows the binomial distribution with probability p, then

$$P(Y=y)=\binom{n}{y}p^{y}(1-p)^{n-y}, y=0, 1, \ldots, n \quad (2.11)$$

In order to make the resulting distribution tractable and flexible, assume p is not a constant but a random variable which follows the beta distribution with parameters α and β. The beta distribution exhibits a fairly wide variety of shapes on the unit interval and may prove useful for describing the true variation in p's.

The probability function of the BB distribution is:

$$P(Y=y)=\binom{n}{y}\frac{\Gamma(\alpha+y)\Gamma(\beta+n-y)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+n)\Gamma(\alpha)\Gamma(\beta)} \quad (2.12)$$

where $y = 0, 1, 2, \ldots, n$, α, $\beta > 0$ (Kleinman, 1973). The quantity $\Gamma(n)$ is known as the gamma function and $\Gamma(n) = (n-1)!$, provided that n is an integer, y is the number of choices of a particular type out of n.

The mean and variance of the BB distribution can be obtained as:

$$E(Y)=\frac{n\alpha}{\alpha+\beta} \quad (2.13)$$

$$V(Y)=\frac{n\alpha\beta(n+\alpha+\beta)}{(\alpha+\beta)^2(1+\alpha+\beta)} \quad (2.14)$$

It is convenient and more meaningful to reparameterize as

$$E(Y) = np \quad (2.15)$$

$$V(Y) = np(1-p)[1+(n-1)\gamma] \quad (2.16)$$

where

$$p=\frac{\alpha}{\alpha+\beta}, \gamma=\frac{1}{\alpha+\beta+1} \quad (p>0, \gamma>0) \quad (2.17)$$

The scale parameter γ, which varies between 0 and 1, measures the variation of the random parameter p. The inflation factor $[1 + (n - 1)\gamma]$ is always greater than or equal to 1, and models the over-dispersion due to the variation of p. Therefore the variance of Y for the same p is always larger under the BB distribution than binomial.

Note that $0 < \gamma < 1$. As $\gamma \to 0$, $V(Y) \to np(1 - p)$ which is equal to the variance of binomial. As $\gamma \to 1$, $V(Y) \to n^2p(1 - p)$, this is the maximum variance of the BB distribution.

Hypergeometric Distribution

Unlike the binomial distribution where the trials are independent and sampling is with replacement, in the hypergeometric distribution (HG), trials are not independent and sampling is without replacement.

Suppose there are total M populations, K males and $(M - K)$ females. Let us perform n trials of an experiment in which a person is chosen at random, its gender is observed, and then the person is not replaced after being chosen. In such a case, if Y is the random variable for the number of males chosen (successes) in n trials ($M \geq n$), then the random variable Y is said to have a HG probability distribution. The probability of a success can be modeled as

$$P(Y=y) = \frac{\binom{K}{y}\binom{M-K}{n-y}}{\binom{M}{n}} \tag{2.18}$$

where y is an integer 0, 1, 2, ..., n, subject to the restriction $y \leq K$ and $n - y \leq M - K$ (Spiegel, 1975).

The mean and variance of Y are

$$E(Y) = \frac{nK}{M}\ ,\ V(Y) = n\left(\frac{K}{M}\right)\left(\frac{M-K}{M}\right)\left(\frac{M-n}{M-1}\right) \tag{2.19}$$

Since we have M total number of blue and red marbles, the proportions of blue and red marbles are $p = \dfrac{K}{M}$ and $q = 1 - p = \dfrac{M-K}{M}$, respectively. We can also re-express the Equation 2.18 and 2.19 as

$$P(Y=y) = \frac{\binom{pM}{y}\binom{(1-p)M}{n-y}}{\binom{M}{n}} \tag{2.20}$$

$$E(Y) = np,\ V(Y) = npq\left(\frac{M-n}{M-1}\right) \tag{2.21}$$

Therefore we can view the term

$\dfrac{M-n}{M-1} \in (0,\ 1)$ as a variance reduction factor in $V(Y)$.

Note for fixed n, as $M \to \infty$ (or M is large compared with n),

$$\frac{M-n}{M-1} \to 1 \text{ and } \mu = E(Y) \to np,\ V(Y) \to npq \tag{2.22}$$

and the variance of HG reduces to the variance of binomial.

To our knowledge, the HG distribution has not been reported to be applied to sensory data in the literature. In terms of sensory responses on a hedonic scale, n is associated with the number of scale categories, and p represents an internal response on a psychological standard scale ranging from 0 to 1. HG model translates this internal response p to external response Y. The parameter M can be varied to reduce the variance of Y from a maximum of npq when M is very large, to a small quantity when M is approximately equal to n ($M \rightarrow n$). Therefore, this model should be able to describe categorical data where the underlying variance is less than that of the binomial distribution.

CHAPTER 3

SIMULATING SENSORY DATA USING DIFFERENT MODELS

Shuo Naini, Dave Plaehn, David Lundahl and Mina McDaniel

To be submitted to the Journal of Food Quality and Preference

May 9, 2003

Abstract

Five different generalized simulation approaches were applied to generate consumer product evaluation data (9-point hedonic data). These methods utilized a scale location parameter (*p*), which represents a respondent's scale location preference on a psychological standard scale (from 0 to 1). The value of *p* was randomly generated using an Empirical Logit model. This model included respondent (random), product, consumer segment, product by segment interaction effects and a random error effect. These five methods differed in how the scale location parameter was applied to generate discrete random responses on a 9-point hedonic scale. The parameter *p* was applied to one of the four underlying statistical distributions (binomial, beta-binomial, hypergeometric and beta-hypergeometric), or alternatively simply transformed directly to a scaled response by a binning step. Briefly, these five models were named as Logit Binomial (LB), Logit Beta-Binomial (LBB), Logit Hypergeometric (LHG), Logit Beta-Hypergeometric (LBHG) and Discrete Empirical Logit (DL) model respectively. These five simulation models were evaluated for their ability to generate simulated data with distributions characteristic of expected real situations.

Results showed that in the situation when the mean of segment responses located near the center of the scale, the LB and LBB model were not flexible enough to simulate affected sensory data with small variation within segment. The LHG model was not flexible for data with large variance. The LBHG model was

flexible to simulate data with small or large variances. Compared with the other four models, the DL model was relative simple and capable of simulating a wide range of consumer data analysis situations.

Introduction

In order to evaluate the robustness of ANOVA and calculate its empirical power under different situations, different data sets need to be collected as inputs. The problem arises when there is little or no data available – the real consumer data required may not exist, or it may be difficult or costly to obtain from different cultures in a limited time period. Simulation studies have been frequently used in this kind of situation to investigate expected measurement outcomes and to examine expected statistical outcomes (Bang et al., 1998). Therefore, the objective of this study is to employ a generalized simulation tool for generating sensory data such that a broad range of sensory situations can be included and used to test new data analysis technology.

Characters of Hedonic Data and Cross-cultural Consumer Data

Both in the past and currently, the use of the labeled categorical scale has been popular (Lawless & Heymann, 1998). The labeled-category-scale is considered advantageous in pairing the (semantic) label with a number allowing for ease of use by (untrained) consumers (Lawless and Malone, 1986), ease of interpretation and, furthermore greater and relatively simple analysis (data can be treated as both continuous and categorical data) (Cox et al., 2001). However, the labeled categorical scale has its disadvantages. It is well documented by Cardello (1996) that there are problems inherent in category scales, particularly with regard to the fact that the categorical labels do not constitute equal intervals; the neutral category reduces its efficiency and the avoidance of end-categories.

Consider the case where the mean of a normal distribution is located in the center of the scale. As the location moves toward the end of the scale, there is not enough room for the entire normal distribution to fit. The spread of the distribution will not be so great when the mean is at an end of the scale. Therefore, when comparing the mean values of responses generated from distributions at the center and the ends of the scale, the respective means will have different variance. In addition, the skewness of the distribution for a mean at the end of the scale means the distribution may no longer be normal. Furthermore, because of different scale usage, the intervals (the category sizes) in the category scale may not be equal (O'Mahony, 1982).

Many studies have suggested that consumers from different cultures may have different scale usage as well as different taste preferences (Bertino and Chan, 1986; Prescott et al., 1992; Ayabe-Kanamura et al., 1998; Yeh and Lundahl, 1999). As an example, in a study comparing 9-point hedonic scale usage between consumers from the US and Chinese, Korean and Thai Consumers, Pacific Rim respondents tend to agree with each other and are reluctant to use the ends of the scale (central tendency). They usually give positive responses and avoid using the lower part of the scale (dislike avoidance) (Yeh and Lundahl, 1999).

Cross-cultural data include information on flavor preference and cultural combinations. Thus, in order to be successful in studying the effects of cross-cultural scaling, a new simulation model must be generated which is not subject to the problems of end of scale distribution compression. It should be able to provide the control necessary to address cross-cultural issues in central tendency and dislike avoidance. In addition, it should have a psychophysical basis.

Thurstonian Psychophysical Basis

In this study, a simulation tool was established on Thurstonian principles of psychophysics. Thurstonian methods represent a psychological approach to sensory analysis and hedonics. They join together psychological measurement principles with real-world stimuli. The hypothesis is that when people rate stimuli, they do so based upon an underlying or internal unbounded psychological scale of hedonics $(-\infty,\infty)$ or liking (Moskowitz, 1983). Therefore, we assume that when a panelist evaluates a product, his or her internal response is based on a psychological scale that is evoked by sensory stimuli. When a subject/observer answers the questionnaire, this internal response is translated automatically to an external response on a categorical scale, such as 9-point hedonic scale. In translating the internal to an external response, these can be a source of error, e.g., from replication to replication of the same stimuli, then a different external response is given.

General Theoretical Background of the Simulation Model

Simulation forms a central part in statistics because of the relative ease with which samples can often be generated from a probability distribution or a certain model. In performing simulations, it is helpful to consider the duality between a probability density function and a histogram of a set of random draws from the distribution. Given a large enough sample, the histogram can provide practically complete information about the distribution, to a level of precision that can be estimated (Gelman et al., 1997).

Simulation is nothing new to sensory modeling. Lundahl (1992), Næs and Langsrud (1998) used the normal distribution to simulate sensory data. However, it is apparent that the normal curve does not provide an adequate simulation basis for all types of sensory category scale data. A broader system of frequency curves is needed to represent a wider variety of situations.

In order to generate sensory consumer hedonic data under a broader range of situations, a flexible distribution family is needed. This study introduced a very flexible family of distributions formed using the Empirical Logit model in a fashion similar to the Johnson System (Jonson, 1949). Such a family offers the added benefit that it can be applied to simulated data from a Thurstonian perspective.

The Johnson System contains three alternative families of distributions, denoted S_U, S_B, and S_L respectively. These involve transformations of the form

$$z = \gamma + \eta k_i(x; \lambda, \varepsilon), \tag{3.1}$$

$i = 1, 2, 3$ for S_U, S_B, and S_L respectively

where z is a standard normal variable and the $k_i(x; \lambda, \varepsilon)$ are chosen to cover a wide range of possible distribution shapes. The S_B distribution applies to certain classes of variables which have physical or natural constraints on their range (Mage, 1980). It takes on the form $k_2(x; \lambda, \varepsilon) = ln(\frac{x - \varepsilon}{\lambda + \varepsilon - x})$ with bounds ε and $\varepsilon + \lambda$. The Empirical Logit model is similar to the S_B distribution.

Logit transformation is one of the Johnson transformations. Let $\gamma = -g(\beta), \eta = 1, \varepsilon = 0, \lambda = 1,$ Equation 3.1 becomes

$$\begin{aligned} z &= \gamma + \eta \kappa_2(x; \lambda, \varepsilon) = \gamma + \eta \ln(\frac{x - \varepsilon}{\lambda + \varepsilon - x}) \\ &= -g(\beta) + \ln(\frac{x}{1 - x}) \end{aligned} \tag{3.2}$$

A logistic regression model describes how a binary (0 or 1) response variable is associated with a set of explanatory variables. It is advocated when the dependent variable is expressed as a proportion (p). So the logistic regression model specifies that a probability is related to a regression-like function of explanatory variables (Ramsey and Schafer, 1997):

$$logit\ (p) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p \tag{3.3}$$

and X_1, $\cdots$, X_p are the explanatory variables. The inverse of the logit function is called the logistic function. If *logit (p)* = η, then

$$p = \frac{exp(\eta)}{1 + exp(\eta)} \tag{3.4}$$

The logit transformation replaces the value p of the dependent variable by their transformed value *log (odds)* or $log(\frac{p}{1-p})$. This has the effect of allowing the dependent variable to take any value in the range $(-\infty, \infty)$ rather than be constrained to the range (0, 1). Thus the regression model, with a normally distributed random component, may be better justified (Cooper and Weekes, 1983). This model is for use when relationships between the non-response factors (explanatory variables) are not of interest. It is implicit in the definition of a logit model that no structure between the explanatory variables is taken into account (Christensen, 1990).

Simulation Models

In this study, a simulation was performed in two steps. First, a scale location preference parameter (on a psychological standard scale from 0 to 1) was generated using the Empirical Logit model with three main effects – respondent, product, consumer segment and a product by segment interaction effect.

In the second stage, the scale location parameter (p) was applied to one of the four underlying statistical distributions: binomial, beta-binomial, hypergeometric, and beta-hypergeometric respectively. These distributions were used to add variation to the internal response due to translation to an external response, and to generate the category scale response (equal intervals). The beta-binomial is a compound distribution of beta and binomial. The beta-hypergeometric is a compound distribution of beta and hypergeometric distribution. The increasing or decreasing of the distribution of category interval responses was controlled by the input parameters – scale location preference parameter (p), number of scale categories (n) and truncation parameters. The four simulation models were named as Logit Binomial (LB), Logit Beta-Binomial (LBB), Logit Hypergeometric (LHG), and Logit Beta-Hypergeometric (LBHG) respectively. On the other hand, data can also be simulated simply by a binning step or a transformation of the parameter p – multiple by n and rounded to integers between 1 and 9. This model was named as Discrete Empirical Logit (DL) model.

The random discrete data sets (10000), which represent the sensory responses on a 9-point hedonic scale, were generated for each simulation model with added random noise. Also, a unique truncation method was introduced to mimic two common scale usage cases in cross-cultural research – central tendency and dislike avoidance.

The simulation algorithm was written in MATLAB™ language (Version 5.0). MATLAB™ is a high-level programming language that has a built in random (more precisely, pseudorandom) number generator – multiplicative linear congruential random-number generators (LCG's).

Methods

Simulation of an Internal Response

A simulation model was created to mimic the following situation:

- j^{th} segment
- i^{th} respondent within j^{th} segment
- i^{th} respondent measures k^{th} product
- l^{th} response for i^{th} respondent on k^{th} product.

In the present study, the situation was simplified as:

- 2 segments, so $j = 1, 2$
- 100 respondents in each segment, so $i = 1, \ldots, 100$
- each respondent measures 2 products – A and B, so $k = 1, 2$
- no replications, so $l = 1$.

A logit "response", which can take on any score between negative and positive infinity, represents the i^{th} underlying internal response (I) by a randomly selected subject, responding to a stimulus on a given replication (Figure 3.1).

The logit internal response, p, was simulated by the Empirical Logit model:

$$I = logit(p_{i(j)kl}) = log[p_{\,i(j)kl} / (1 - p_{\,i(j)kl})]$$

$$= R_{i(j)} + g(X_j, X_k \mid \beta_0, \beta_1, \beta_2, \beta_3) + (E_{i(j)kl}) \qquad (3.5)$$

and $g(X_j, X_k \mid \beta_0, \beta_1, \beta_2, \beta_3) = \beta_0 + \beta_1 X_j + \beta_2 X_k + \beta_3 X_j X_k$

$i = 1, 2, \ldots, r \;\; j = 1, 2, \ldots, a \quad k = 1, 2, \ldots, \mathrm{b} \;\; (r = 100, \;\; a = b = 2)$

where β_0 is the grand mean, β_1 is the fixed effect for factor segment, β_2 is the fixed effect for factor product, and therefore β_3 is the fixed effect for segment by product interaction. $g(X_j, X_k \mid \beta_0, \beta_1, \beta_2, \beta_3)$ or g is a function of β's. X_j and X_k are both indicator variables, with X_j taking on a value of 0 if the respondent is from the first segment, and 1 if from the second segment; X_k takes on a value 0 if the respondent evaluates product A, and 1 if evaluates product B. Both $R_{i(j)}$ and $E_{i(j)kl}$ are independent, normally distributed random noise; $R_{i(j)} \sim N(0, \sigma_R^2)$ and $E_{i(j)kl} \sim N(0, \sigma_E^2)$. $R_{i(j)}$ represents a random location effect for the i^{th} respondent within the j^{th} segment, that is, the variance associated with the mean scores location on the scale. $E_{i(j)kl}$ denotes within segment inconsistency. It reflects the within segment heterogeneity in acceptance variation. In this study, σ_R and σ_E are independent of product and segment.

From this internal response measure, the next step in the development of a simulator is to translate the internal response into an external response (Figure 3.2). To do so, five different models were generated, each adding a range of noise associated with a subject's translation of his/her respective internal response into an external response (Figure 3.3).

Simulation of an External Response

By controlling the input parameters, g, σ_R and σ_E, the logit form of parameter $p_{i(j)kl}$ was randomly generated by the Empirical Logit model for the i^{th}

Figure 3.1 An internal i^{th} response takes on any value between negative infinity and positive infinity.

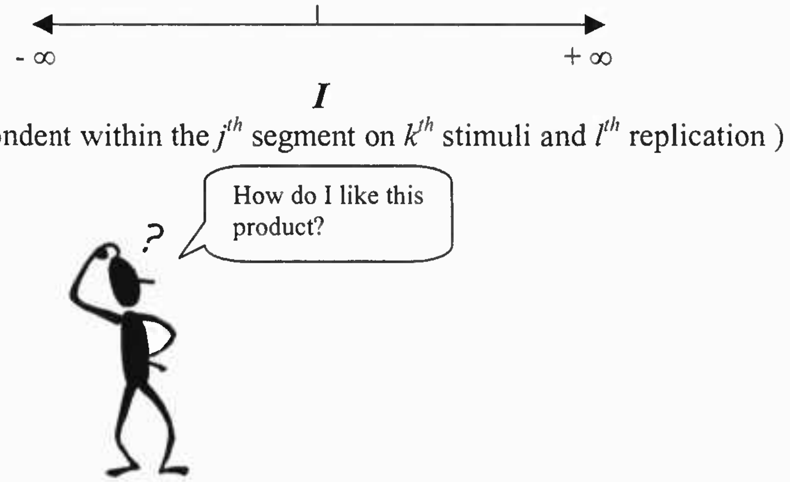


Figure 3.2 Translating an internal response into an external response Y.

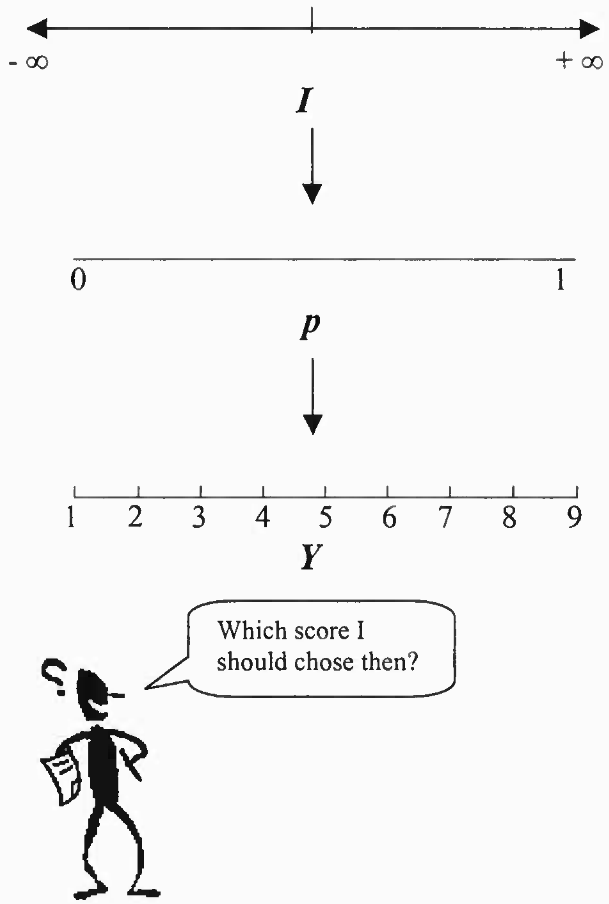
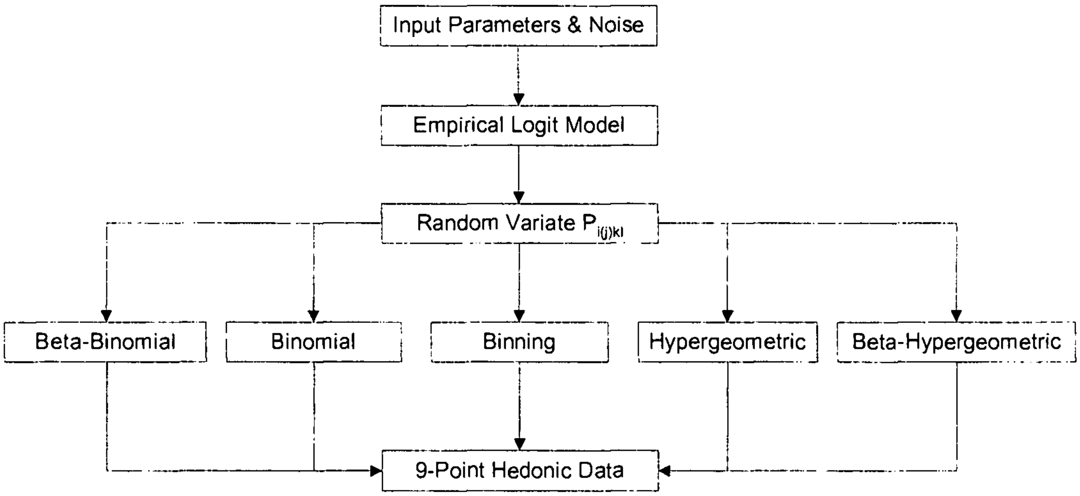


Figure 3.3 Flowchart of the simulation program.



respondent within the j^{th} segment, k^{th} product (stimulus), l^{th} replication with equation

$$p_{i(j)kl} = \frac{exp(logit(p_{i(j)kl})}{1 + exp(logit(p_{i(j)kl})} \tag{3.6}$$

$p_{i(j)kl}$ stands for the standardized score of liking on a 0 to 1 scale, where 0 represents dislike extremely, 0.5 represents neither like nor dislike and 1 represents like extremely (Figure 3.2). This standardized response also represents a case where there is no added variation in translating an internal into an external response.

For the LB, LBB, LHG, LBHG model, four underlying distributions – binomial, beta-binomial, hypergeometric and beta-hypergeometric were applied respectively to simulate a random response based upon the random value of $p_{i(j)kl}$ (Figure 3.3). They added an additional flexibility into the response.

In the case of Discrete Empirical Logit model (DL), the resulting random variate $p_{i(j)kl}$ was directly transformed into a scaled response which broke up the (0, 1) scale into nine equal categories with no added variation.

Summary of the Five Models

Model 1 – Logit Binomial Model (LB)

The basic Logit Binomial model (LB) follows a combined distribution of the Empirical Logit and binomial. At the first step, the scale location preference parameter $p_{i(j)kl}$ was generated from the Empirical Logit model without the extra random error ($E_{i(j)kl}$)

$$logit(p_{i(j)kl}) = \beta_0 + R_{i(j)} + \beta_1 X_j + \beta_2 X_k + \beta_3 X_j X_k \tag{3.7}$$

This variation can be considered as simulating a "baseline" randomness to within subject response error (variation the most at $p = 0.5$) as an alternative to using $E_{i(j)kl}$ in the Empirical Logit model. At the second step, this model assumes that the individual responses $Y_{i(j)kl}$ for the i^{th} respondent within j^{th} segment, k^{th} product and l^{th} replication follows a binomial distribution with the probability mass function (*pmf*)

$$P(Y_{i(j)kl} = y_{i(j)kl}) = \binom{n}{y} (p_{i(j)kl})^{y} (1 - p_{i(j)kl})^{n-y} \tag{3.8}$$

where $y_{i(j)kl} = 0, 1, 2, \ldots, n$ and $n = 8$ (number of scale categories minus one), $0 \leq p_{i(j)kl} \leq 1$. Thus, the random responses $Y_{i(j)kl} \in [1, 9]$ were generated by the binomial distribution

$$Y_{i(j)kl} \sim \text{binomial}(n, p_{i(j)kl}) + 1 \tag{3.9}$$

with two sources of variation, one from $R_{i(j)}$ and one from the binomial distribution. The binomial variation

$$V(Y_{i(j)kl}) = n(p_{i(j)kl})(1 - np_{i(j)kl}) \tag{3.10}$$

The truncation and double-truncation methods were used for the scaling effects of central tendency and dislike avoidance. Two truncation parameters k_1 and k_2 were added to the basic model and the truncation model becomes:

$$Y_{i(j)kl} \sim \text{binomial}(n - k_1, p_{i(j)kl}) + k_2 + 1 \tag{3.11}$$

where $p_{i(j)kl}$ was generated from the Equation 3.7.

- In the central tendency case, the 9-point hedonic scale was double truncated by k_1 and k_2 where k_1 is an even integer and

$$k_2 = k_1 / 2 \tag{3.12}$$

  That is, if k_1 is chosen as 2, 4, 6, then k_2 is equal to 1, 2, 3, and the scores of response $Y_{i(j)kl}$ will vary from 2 to 8, 3 to 7, 4 to 6 respectively (Table 3.1).

- In the Dislike Avoidance Case, the scale was truncated on one side where k_1 is an integer and

$$k_2 = k_1 \tag{3.13}$$

  If k_1 is chosen as 1, 2, 4, then k_2 is equal to 1, 2, 4, and the scores will be allowed to range from 2 to 9, 3 to 9, 5 to 9 respectively (Table 3.1).

Model 2 – Logit Hypergeometric Model (LHG)

Similarly as the LB model, the basic Logit Hypergeometric model (LHG) follows a compound distribution of the Empirical Logit and hypergeometric. The parameter $p_{i(j)kl}$ was generated from the Empirical Logit model without the extra random error ($E_{i(j)kl}$) (Equation 3.7). The model assumes that the individual responses $Y_{i(j)kl} \in [1, 9]$ follows the hypergeometric distribution with the *pmf* as in equation 2.18

Table 3.1 Simulated random responses by the LB truncation model for central tendency and dislike avoidance.

	Central Tendency			Dislike Avoidance		
k_1	2	4	6	1	2	4
k_2	1	2	3	1	2	4
$Y_{i(j)kl} \sim$ binomial $(n - k_1, p_{i(j)kl}) + k_2 + 1$	2 ~ 8	3 ~ 7	4 ~ 6	2 ~ 9	3 ~ 9	5 ~ 9

Note: k_1, k_2 – truncation parameters, $Y_{i(j)kl}$ – respondent, $p_{i(j)kl}$ – standardized liking score on a 0 to 1 scale, n is the number of categories which $n = 8$ and other parameters are as previously defined. For example, when $k_1 = 2$, $k_2 = 1$, the score range simulated by the LB truncation model for central tendency case is from 2 (when $p_{i(j)kl} = 0$) to 8 (when $p_{i(j)kl} = 1$).

$$P(Y_{i(j)kl} = y_{i(j)kl}) = \frac{\binom{K}{y}\binom{M-K}{n-y}}{\binom{M}{n}} \tag{3.14}$$

where $y_{i(j)kl}$ is an integer 0, 1, 2, …, n subject to the restriction $y \le K$ and $n - y \le M - K$. M, $K_{i(j)kl}$ and n are input parameters where $K_{i(j)kl} = Round(M \times p_{i(j)kl})$, n = number of scale categories minus one = 8. The random response $Y_{i(j)kl}$ was generated by

$$Y_{i(j)kl} \sim \text{Round}\ [\text{HG}(M, p_{i(j)kl}, n)] + 1 \tag{3.15}$$

There are two sources of variation involved, one from $R_{i(j)}$ and one from rounding and the hypergeometric distribution as in equation 2.21

$$V(Y_{i(j)kl}) = np_{i(j)kl}(1 - p_{i(j)kl})\left(\frac{M-n}{M-1}\right) \tag{3.16}$$

Here, M can be considered as a variance reduction factor. As M gets very large, the hypergeometric variation goes to maximum $np_{i(j)kl}(1 - p_{i(j)kl})$, allowing for simulations where there is inconsistent translation from internal to external responses. As M is approximately equal to n, the variance reduced to a small quantity to simulate a consistent translation from internal to external responses. This model was used to generate categorical data where the underlying variance is less than that of the binomial model.

The truncated hypergeometric model is

$$Y_{i(j)kl} \sim Round\ [\text{HG}(M, p_{i(j)kl}, (n - k_1))] + k_2 + 1 \tag{3.17}$$

where $p_{i(j)kl}$ was generated from the Equation 3.7 and other parameters are as previously defined (Table 3.2).

Model 3 – Logit Beta-Binomial Model (LBB)

Unlike the binomial distribution, the beta-binomial distribution has an additional source of variation from $p_{i(j)kl}$, which is assumed to be a random variable for each respondent following the beta distribution with parameters $\alpha_{i(j)kl}$ and $\gamma_{i(j)kl}$. In Logit Beta-Binomial model, we assume the responses

$$Y_{i(j)kl} \sim \text{LBB}\ (n, p_{i(j)kl}, \gamma_{i(j)kl}) + 1 \qquad (3.18)$$

where n = number of scale categories minus one = 8, $\gamma_{i(j)kl}$ is the variance inflation factor for individuals, $p_{i(j)kl}$ is as previously defined. If $\gamma_{i(j)kl} = 0$, then every respondent varies replication responses same as the binomial. Since the scale parameter $\gamma_{i(j)kl}$ varies between 0 and 1, the inflation factor $[1 + (n - 1)\gamma_{i(j)kl}]$ is always greater than or equal to 1. Apparently, the beta-binomial distribution has a larger variance than the binomial distribution and is a better fit for over-dispersed data.

The procedure including the Empirical Logit model and the beta-binomial model is called the Logit Beta-Binomial model (LBB). The theoretical mean and variance of LBB are listed in the Results.

In the truncated LBB model, there are two additional truncation parameters involved (Table 3.3),

Table 3.2 Simulated random responses by the LHG truncation model for central tendency and dislike avoidance.

	Central Tendency			Dislike Avoidance		
k_1	2	4	6	1	2	4
k_2	1	2	3	1	2	4
$Y_{i(j)kl} \sim$ Round [HG(M, $p_{i(j)kl}$, $(n - k_1)$)] $+k_2+1$	2 ~ 8	3 ~ 7	4 ~ 6	2 ~ 9	3 ~ 9	5 ~ 9

Note: k_1, k_2 – truncation parameters, $Y_{i(j)kl}$ – respondent, $p_{i(j)kl}$ – standardized liking score on a 0 to 1 scale, n is the number of categories which n = 8 and other parameters are as previously defined. For example, when k_1 = 2, k_2 = 1, the score range simulated by the LHG truncation model for central tendency case is from 2 to 8.

Table 3.3 Simulated random responses by the LBB truncation model for central tendency and dislike avoidance.

	Central Tendency			Dislike Avoidance		
k_1	2	4	6	1	2	4
k_2	1	2	3	1	2	4
$Y_{i(j)kl} \sim$ LBB$(n - k_1, p_{i(j)kl}, \gamma_{i(j)kl}) + k_2 + 1$	2 ~ 8	3 ~ 7	4 ~ 6	2 ~ 9	3 ~ 9	5 ~ 9

Note: k_1, k_2 – truncation parameters, $Y_{i(j)kl}$ – respondent, $p_{i(j)kl}$ – standardized liking score on a 0 to 1 scale, n is the number of categories which n = 8 and other parameters are as previously defined. For example, when k_1 = 2, k_2 = 1, the score range simulated by the LBB truncation model for central tendency case is from 2 to 8.

$$Y_{i(j)kl} \sim \text{LBB}\ (n - k_1,\ p_{i(j)kl},\ \gamma_{i(j)kl}) + k_2 + 1 \tag{3.19}$$

Model 4 – Logit Beta-Hypergeometric Model (LBHG)

The algorithm is similar to the one in the beta-binomial model. We assume $p_{i(j)kl}$ is not constant for each respondent but a random variable follows the beta distribution with parameter $\alpha_{i(j)kl}$ and $\gamma_{i(j)kl}$, the individual responses $Y_{i(j)kl}$ has a hypergeometric distribution with parameters M, $p_{i(j)kl}$ and n,

$$Y_{i(j)kl} \sim Round\ [\text{BHG}\ (M, p_{i(j)kl}, \gamma_{i(j)kl}, n)] + 1 \tag{3.20}$$

where M, $p_{i(j)kl}$, and n are previously defined in Model 2 – LHG.

In this model, M is the variance reduction factor, and $\gamma_{i(j)kl}$ is the variance inflation factor for individuals. These two parameters allow the beta-hypergeometric model generate data with small and large variances. As M decreases approximately to n, the variance of beta-hypergeometric reduced to simulate a consistent translation from internal to external responses. As $\gamma_{i(j)kl}$ increases, the variance increases to simulate an inconsistent translation from internal to external responses. For example, when $M \to \infty$, and $\gamma = 1$, the variance gets its maximum; when $M \to \infty$, and $\gamma = 0$, the variance is close to the variance of the LHG model; as $M \to n$, and $\gamma = 0$, the variance gets its minimum (close to zero). Therefore, the beta-hypergeometric model becomes a generalized model for simulating the variation of replicated sensory response from zero variance to large variance.

The procedure including the Empirical Logit model and the beta-hypergeometric model is called Logit Beta-Hypergeometric model (LBHG). The theoretical mean and variance of LBHG are listed in the Results. The truncated LBHG model is (Table 3.4)

$$Y_{i(j)kl} \sim Round\ [\mathrm{BHG}\ (M, p_{i(j)kl}, \gamma_{i(j)kl}, (n - k_1))] + k_2 + 1 \qquad (3.21)$$

Model 5 – Discrete Empirical Logit (DL)

Conceptually, the algorithm used to construct the response data is as follows (Figure 3.4). First, the scale location preference parameter $p_{i(j)kl}$ was randomly generated for each segment by the Empirical Logit model (Equation 3.5)

$$logit(p_{i(j)kl}) = \beta_0 + R_{i(j)} + \beta_1 X_j + \beta_2 X_k + \beta_3 X_j X_k + E_{i(j)kl} \qquad (3.22)$$

This model includes two sources of variation, from the variance of $R_{i(j)} \sim N(0,\ \sigma_R^2)$ and $E_{i(j)kl} \sim N(0,\ \sigma_E^2)$. Let

$$V = R + E \text{ and } var = \sigma_R^2 + \sigma_E^2 \qquad (3.23)$$

In the second step, the responses $Y_{i(j)kl}$ were calculated from a binning step

$$Y_{i(j)kl} = Floor\ (n \times p_{i(j)kl}) + 1 \qquad (3.24)$$

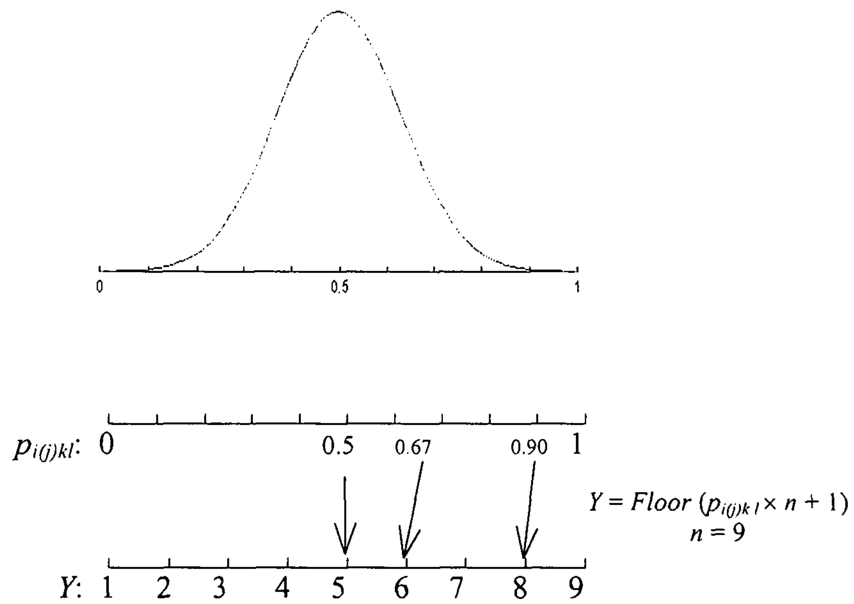
where $n = 9$ (number of response scale categories). Command "*floor*" means *round* the elements to the nearest integers towards minus infinity. This step represented the situation where the translation of an internal to an external response is deterministic (i.e., not adding variation). The procedure including the Empirical

Table 3.4 Simulated random responses by the LBHG truncation model for central tendency and dislike avoidance.

	Central Tendency			Dislike Avoidance		
k_1	2	4	6	1	2	4
k_2	1	2	3	1	2	4
$Y_{i(j)kl}$ ~ Round [BHG (M, $p_{i(j)kl}$, $\gamma_{i(j)kl}$, (n − k_1))] + k_2 +1	2 ~ 8	3 ~ 7	4 ~ 6	2 ~ 9	3 ~ 9	5 ~ 9

Note: k_1, k_2 – truncation parameters, $Y_{i(j)kl}$ – respondent, $p_{i(j)kl}$ – standardized liking score on a 0 to 1 scale, n is the number of categories which n = 8 and other parameters are as previously defined. For example, when k_1 = 2, k_2 = 1, the score range simulated by the LBHG truncation model for central tendency case is from 2 to 8.

Figure 3.4 Binning step of the DL model.



Logit model and the binning step was named as Discrete Empirical Logit model (DL).

If we simply use

$$Y_{i(j)kl} = \text{DL }(p_{i(j)kl}, \text{ and other parameters}) + 1 \qquad (3.25)$$

to represent the DL basic model with dispersion parameters ($p_{i(j)kl,}$ and other parameters distinct to each model), then the truncation DL model can be written as

$$Y_{i(j)kl} = \text{DL }(p_{i(j)kl}, \ k_1, \text{ and other parameters}) + k_2 + 1 \qquad (3.26)$$

with additional truncation parameter k_1 and k_2 (Table 3.5).

Results and Discussion

Theoretical Results

LB, LHG, LBB, and LBHG Model

Let μ_1' be the first moment of parameter $p_{i(j)kl}$ (as previously defined), and μ_2' be the second moment of $p_{i(j)kl}$,

$$\begin{aligned}\mu_1'(g,\sigma) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\frac{e^{g+x}}{1+e^{g+x}}e^{-x^2/2\sigma^2}dx \\ &= \frac{1}{\sqrt{2\pi(\sigma_R^{\ 2}+\sigma_E^{\ 2})}}\int_0^1 p\frac{e^{-(\ln\frac{p}{1-p}-g)^2/2(\sigma_R^{\ 2}+\sigma_E^{\ 2})}}{p(1-p)}dp\end{aligned} \qquad (3.27)$$

$$\mu_2'(g,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\left(\frac{e^{g+x}}{1+e^{g+x}}\right)^2 e^{-x^2/2\sigma^2}dx \qquad (3.28)$$

Table 3.5 Simulated random responses by the DL truncation model for central tendency and dislike avoidance.

	Central Tendency			Dislike Avoidance		
k_1	2	4	6	1	2	4
k_2	1	2	3	1	2	4
$Y_{i(j)kl} = \text{Floor}\,((n - k_1)\, p_{i(j)kl}) + k_2 + 1$	2 ~ 8	3 ~ 7	4 ~ 6	2 ~ 9	3 ~ 9	5 ~ 9

Note: k_1, k_2 – truncation parameters, $Y_{i(j)kl}$ – respondent, $p_{i(j)kl}$ – standardized liking score on a 0 to 1 scale, n is the number of categories which n = 9 and other parameters are as previously defined. For example, when k_1 = 2, k_2 = 1, the score range simulated by the DL truncation model for central tendency case is from 2 to 8.

then the theoretical results of the four models except the DL model are as follows:

LB: $E(Y) = n\mu_1' + 1$; $V(Y) = n\mu_1' + n(n-1)\mu_2' - n^2\mu_1'^2$ (3.29)

LHG: $E(Y) = n\mu_1' + 1$;

$$V(Y) = \frac{10(M-10n)n\mu_1' + 10nM(10n-1)\mu_2'}{100(M-1)} - n^2\mu_1'^2 \quad (3.30)$$

LBB: $E(Y) = n\mu_1' + 1$; $V(Y) = \dfrac{n(n-1)(\theta\mu_1' + \mu_2')}{1+\theta} + n\mu_1' - n^2\mu_1'^2$ (3.31)

LBHG: $E(Y) = n\mu_1' + 1$;

$$V(Y) = \frac{100\theta\, n^2\mu_1' + 100n^2\mu_2' + \dfrac{10n(M-10n)(\mu_1' - \mu_2')}{M-1}}{100(\theta + 1)} - n^2\mu_1'^2 \quad (3.32)$$

where parameters θ, $g(\beta)$, σ_R, σ_E, p, n, M, are all previously defined.

DL Model

From the Empirical Logit model (3.5)

$$log(\frac{p_{i(j)kl}}{1-p_{i(j)kl}}) = g + R_{i(j)} + E_{i(j)kl}$$

where both $R_{i(j)}$ and $E_{i(j)kl}$ are independent, normally distributed, $R_{i(j)} \sim N(0, \sigma_R^2)$ and $E_{i(j)kl} \sim N(0, \sigma_E^2)$. Therefore $R_{i(j)} + E_{i(j)kl} \sim N(0, var)$ where $var = \sigma_R^2 + \sigma_E^2$ and the variation from the model $V = R + E$ (Equation 3.24). After the transformation, the *pdf* of $p_{i(j)kl}$ was calculated as:

$$f_{p_{i(j)kl}} = [\frac{1}{\sqrt{2\pi(\sigma_R^2+\sigma_E^2)}}\exp(\frac{-[\log\frac{p}{1-p} - g]^2}{2(\sigma_R^2+\sigma_E^2)})](\frac{1}{p(1-p)}) \qquad (3.33)$$

The distribution of $p_{i(j)kl}$ is very flexible and can take on various shapes.

If $g = 0$, there are no segment, product or segment by product interaction effects. In this case, the *pdf* stays symmetric about 0.5 and varies from Bell-shape (e.g., *var* = 0.3 or 1.2) to U-shape (e.g., *var* = 5.2) as variance (σ_R^2 + σ_E^2) increases (Figure 3.5). If *var* = 2.2, the *pdf* roughly looks like a uniform.

If $g < 0$ or $g > 0$ (i.e. segment, product and/or segment by product effects are greater than zero), the *pdf* shifts to the left or right respectively and no longer stays symmetric (Figure 3.6 with *var* = 0.3, Figure 3.7 with *var* = 2.2, and Figure 3.8 with *var* = 5.2).

Figure 3.9 visualized the 3-D *pdf* distribution as variance increased at $g = 0$. Figure 3.10 and 3.11 displayed the 3-D *pdf* distributions with $g = -1.0$ and -2.0 respectively as variance increased. Figure 3.12, 3.13, 3.14 showed the *pdf*s with *var* = 9.0, 4.0, 0.25 respectively as *g* varied between −5.0 and 5.0 interval.

Figure 3.15 showed the theoretical results of the DL model on the 9-point hedonic scale with the inputs: σ_R^2 = 0.4, σ_E^2 = 0.35, or *var* = 0.75, and g = 0.00. Therefore, by varying the inputs in the simulation study, the DL model could possibly mimic difference cases of sensory cross-cultural consumer responses.

Figure 3.5 Illustration of symmetric probability density function (*pdf*) of $p_{i(j)kl}$ for the DL model as variance ($var = \sigma_R^2 + \sigma_E^2$) varies, $g = 0.0$.

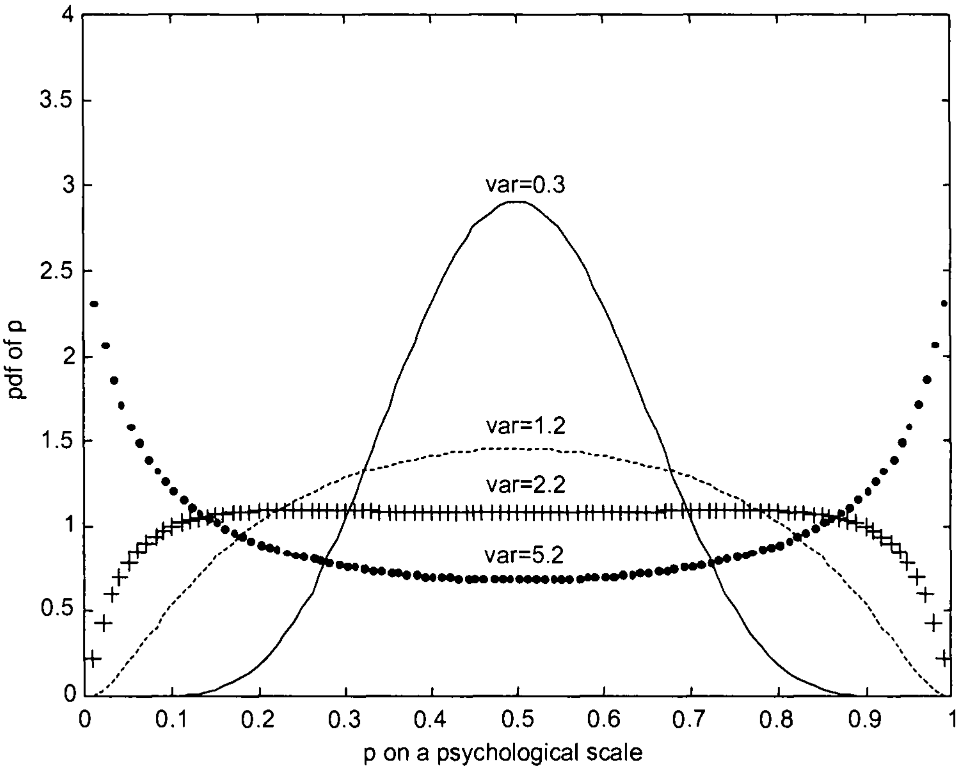


Figure 3.6 Illustration of probability density function (*pdf*) of $p_{i(j)kl}$ for the DL model as g varies, $var = \sigma_R^2 + \sigma_E^2 = 0.3$.

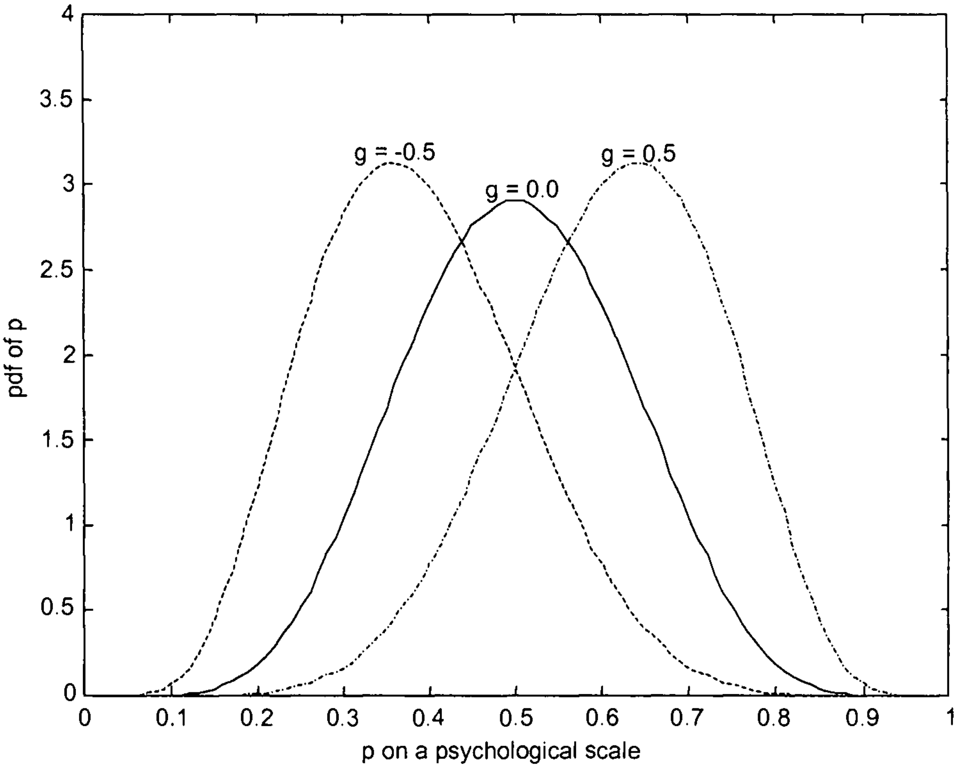


Figure 3.7 Illustration of probability density function (*pdf*) of $p_{i(j)kl}$ for the DL model as g varies, $var = \sigma_R^2 + \sigma_E^2 = 2.2$.

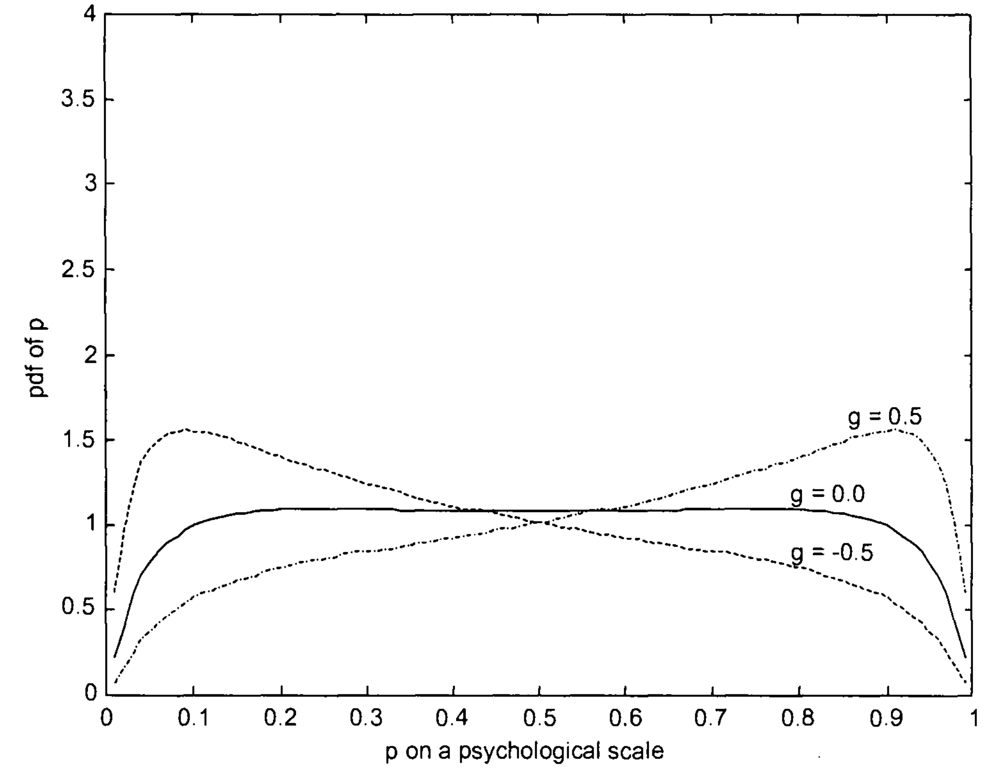


Figure 3.8 Illustration of probability density function (*pdf*) of $p_{i(j)kl}$ for the DL model as g varies, $var= \sigma_R^2 + \sigma_E^2 = 5.2$.

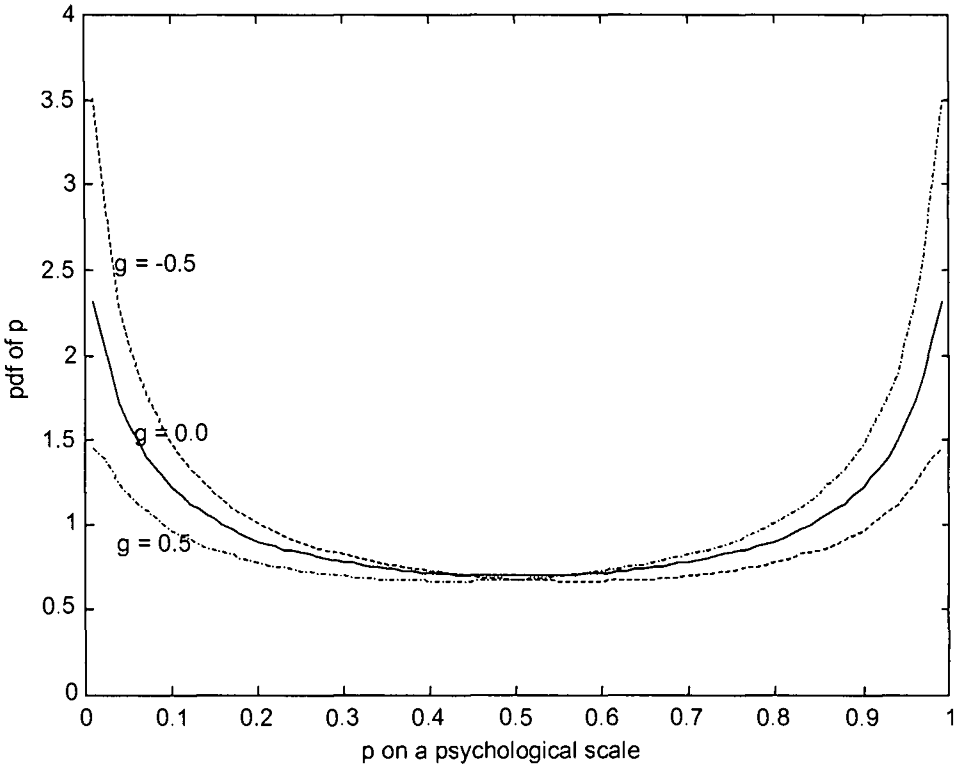


Figure 3.9 3-D graphic display of probability density function (*pdf*) of $p_{i(j)kl}$ for the DL model as variance (or standard deviation = *sqrt(var)* or $\sqrt{var}$) varies, $g = 0.0$.

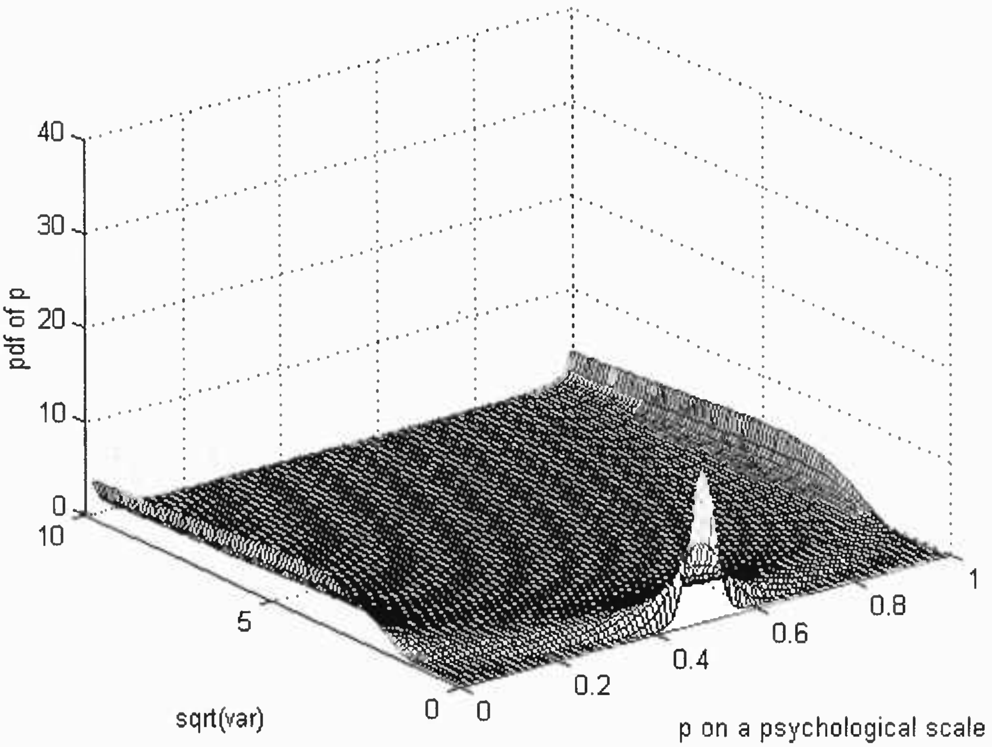


Figure 3.10 3-D graphic display of probability density function (*pdf*) of $p_{i(j)kl}$ for the DL model as variance (or standard deviation = *sqrt(var)* or $\sqrt{var}$) varies, g = -1.0.

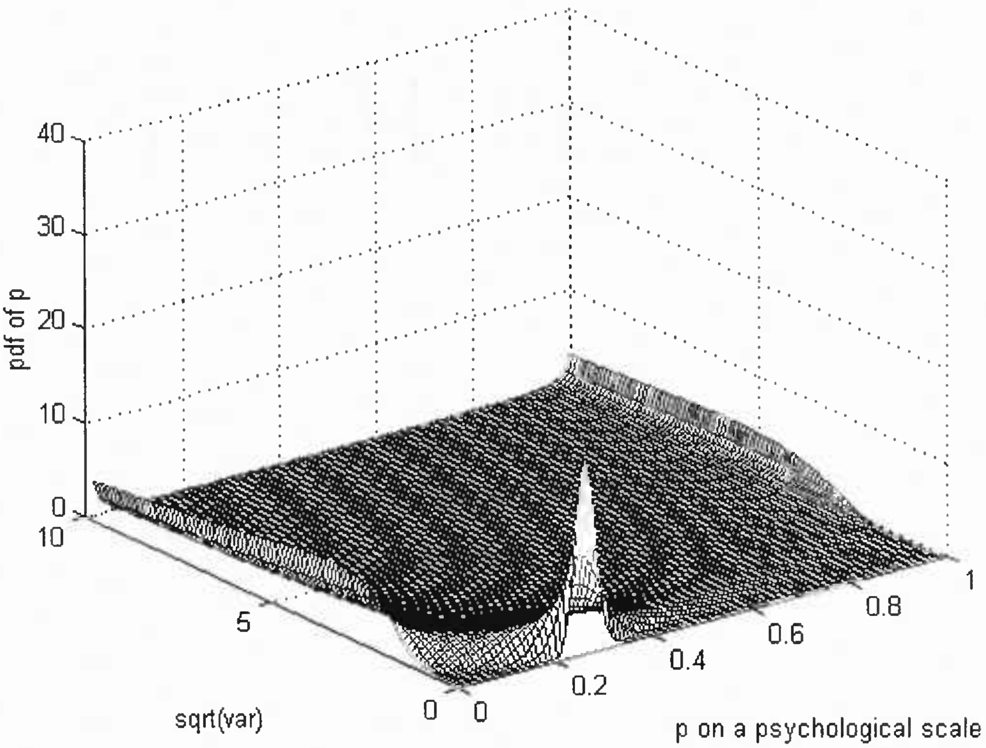


Figure 3.11 3-D graphic display of probability density function (*pdf*) of $p_{i(j)kl}$ for the DL model as variance (or standard deviation = *sqrt(var)* or $\sqrt{var}$)) varies, g = -2.0.

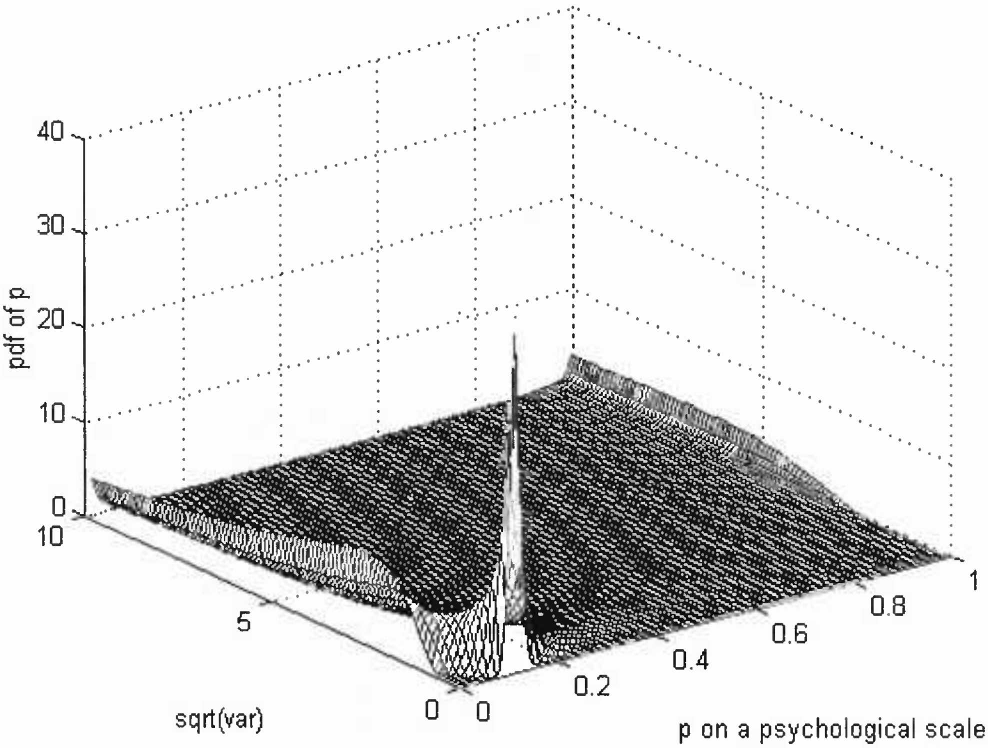


Figure 3.12(a) 3-D graphic display of probability density function (*pdf*) of $p_{i(j)kl}$ for the DL model as the location of g varies, $var = 9.0$.

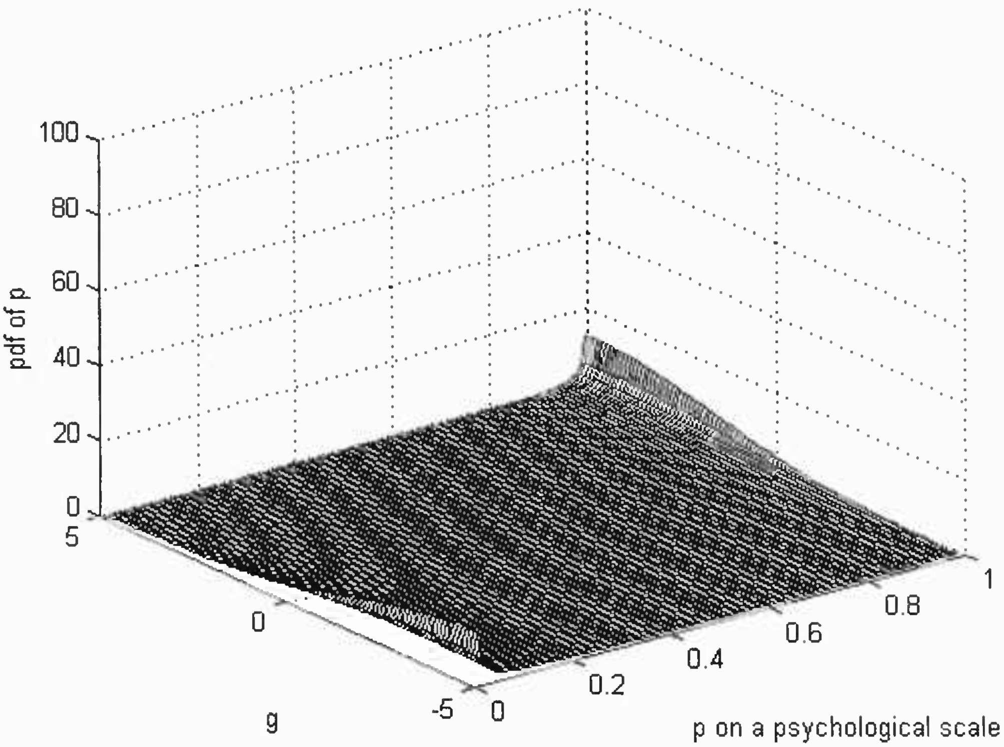


Figure 3.12(b) A close look at the 3-D graphic display of probability density function (*pdf*) of $p_{i(j)kl}$ for the DL model as the location of *g* varies, *var* = 9.0.

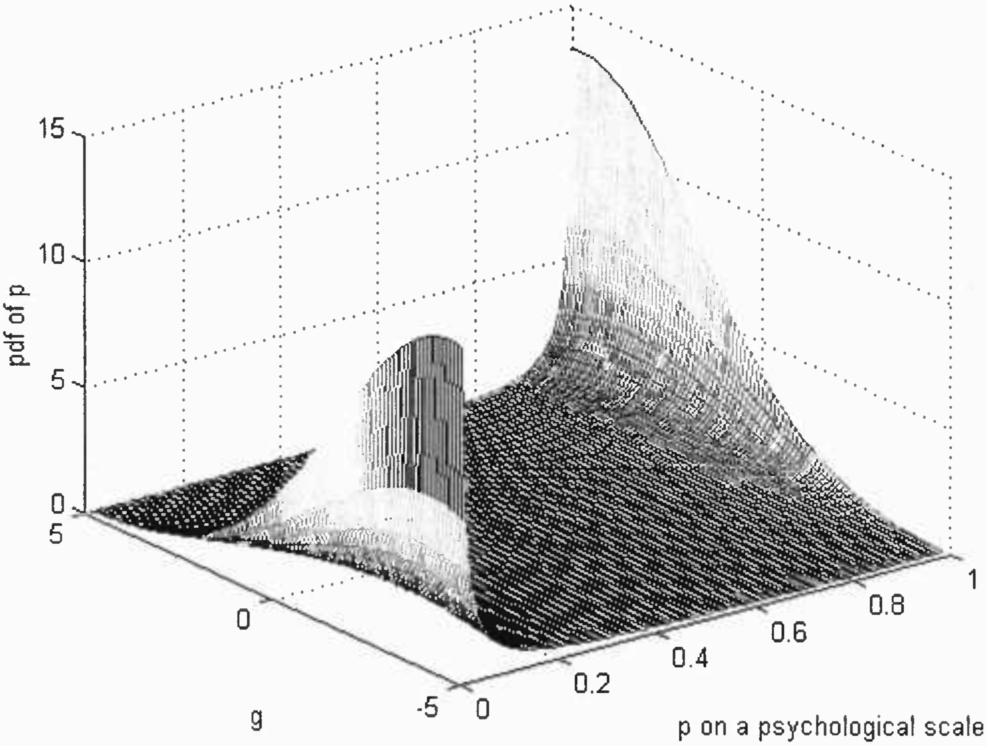


Figure 3.13(a) 3-D graphic display of probability density function (*pdf*) of $p_{i(j)kl}$ for the DL model as the location of *g* varies, *var* = 4.0.

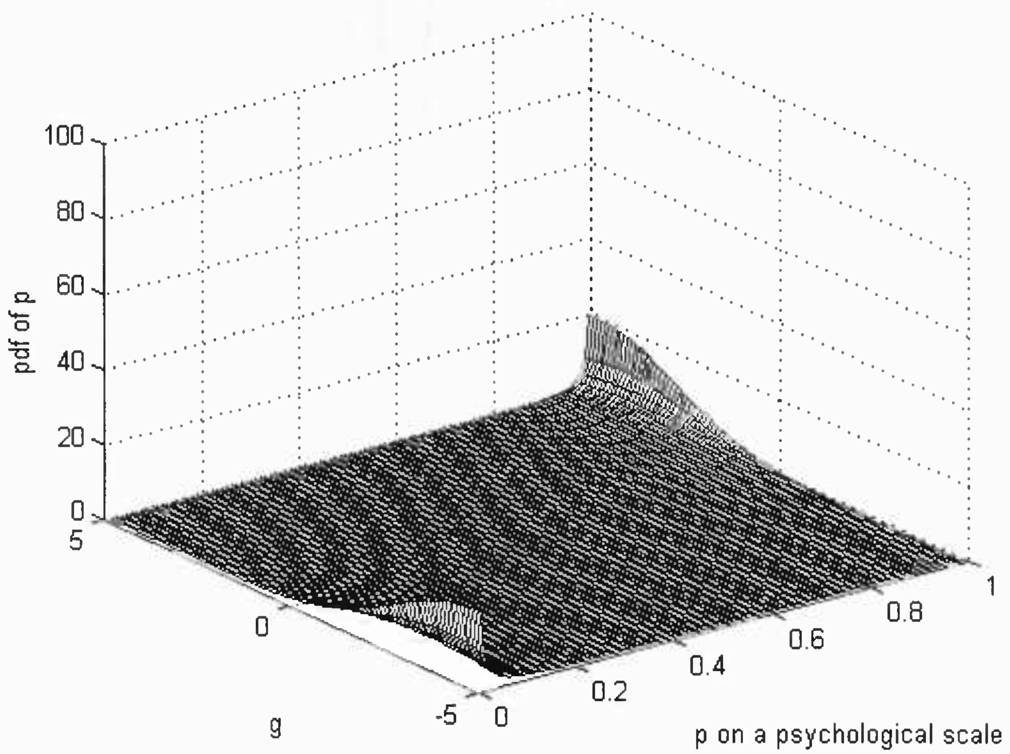


Figure 3.13(b) A close look at the 3-D graphic display of probability density function (*pdf*) of $p_{i(j)kl}$ for the DL model as the location of g varies, $var = 4.0$.

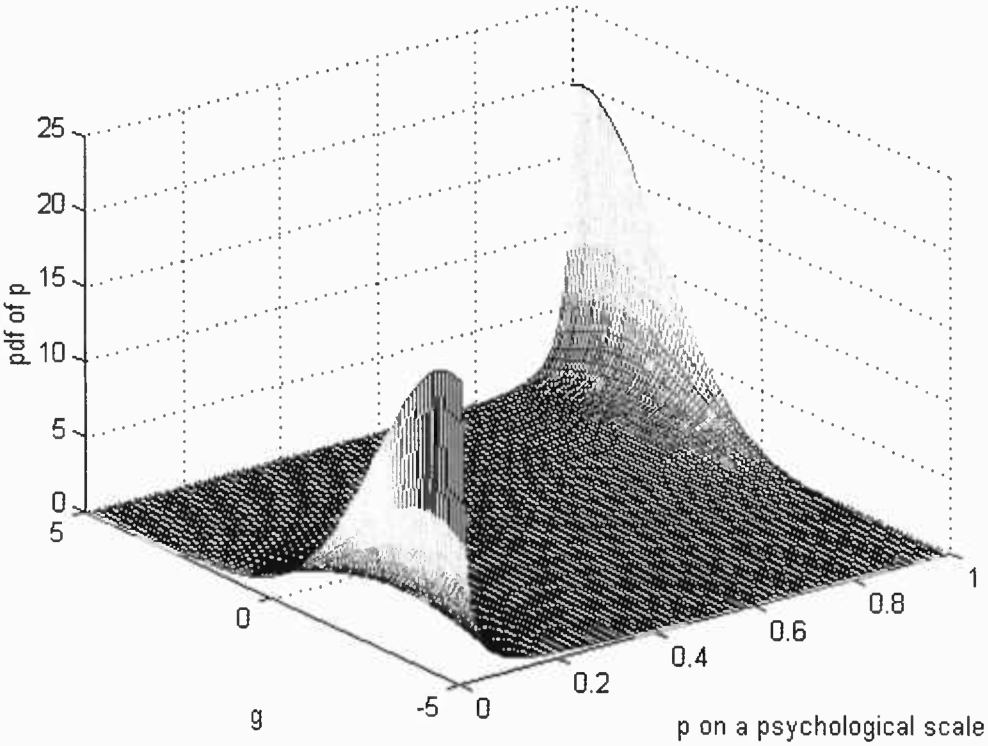


Figure 3.14 3-D graphic display of probability density function (*pdf*) of $p_{i(j)kl}$ for the DL model as the location of *g* varies, *var* = 0.25.

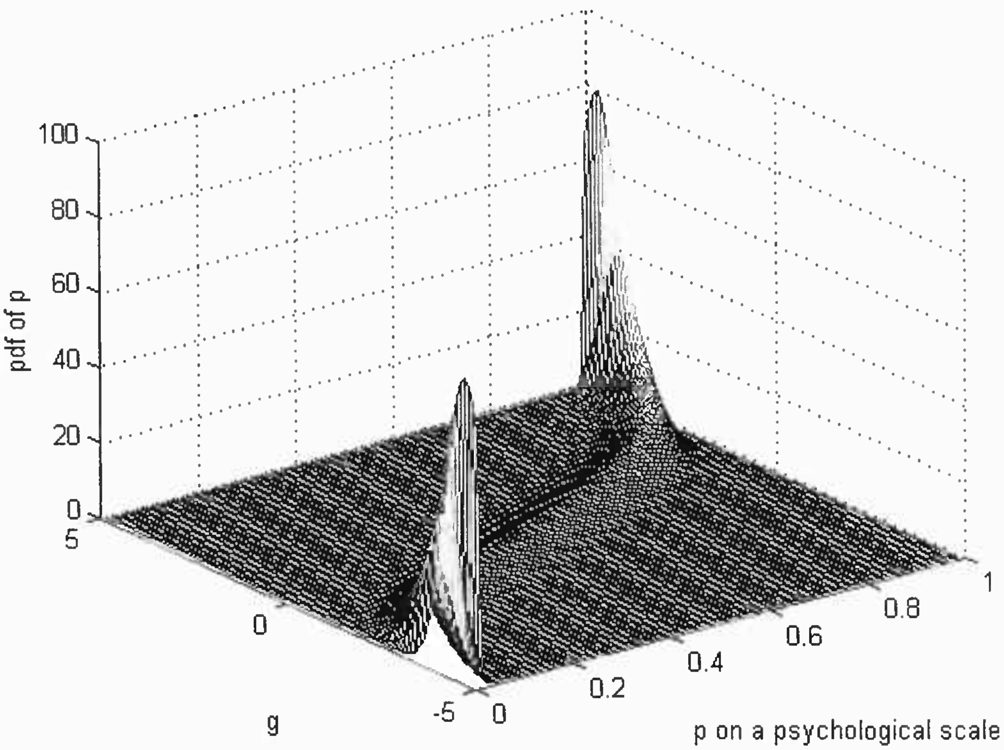
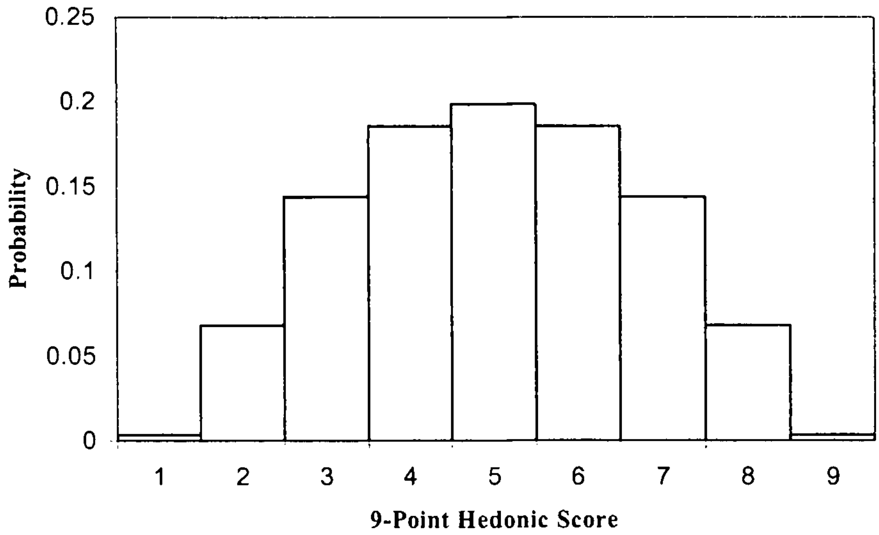


Figure 3.15 Theoretical results for the DL model, $\sigma_R^2 = 0.4$, $\sigma_E^2 = 0.35$, $g = 0.00$.



Empirical Results

The empirical results showed that the DL model was sufficient to simulate sensory data. It could generate data with various types of distributions, e.g., bi-modal (Figure 3.16 and 3.17), roughly uniform (Figure 3.18 and 3.19), and unimodel distribution (Figure 3.20).

As the values of the Empirical Logit model inputs (σ_R and σ_E) decreased, the DL model could have the minimum dispersion under the condition $g = 0.00$ (mean of distribution is located in the center of a psychological scale ranging from 0 to 1). For example, Figure 3.21 showed that at $g = 0.00$, $\sigma_R = 0.10$, $\sigma_E = 0.10$, or $var = 0.02$, the variation of the distribution was around 0.12 on a 9-point hedonic scale.

Unfortunately, the LB and LBB models were not very flexible to generate affected sensory data when respondents were all internally consistent in response patterns. They were better used to simulate responses over-dispersed instead of less-dispersed. 10000 data were generated separately by the LB and LBB model with the inputs: $\sigma_R = 0.0001$ (low variation of the random respondent effect), $g = 0.00$ (Figure 3.22) and $\gamma = 0.00001$ (dispersion coefficient, which means each respondents repeat themselves consistently at each time) (Figure 3.23). The observed variances were 1.98 and 2.03 respectively, which means no matter how consistent the responses were given by individuals, the minimum variance of the

Figure 3.16 The distribution of 10000 simulated sensory consumer responses by the DL model (σ_R = 10.00, σ_E = 10.00, g = 0.00, n = 9). The observed mean = 4.99, variance = 14.68.

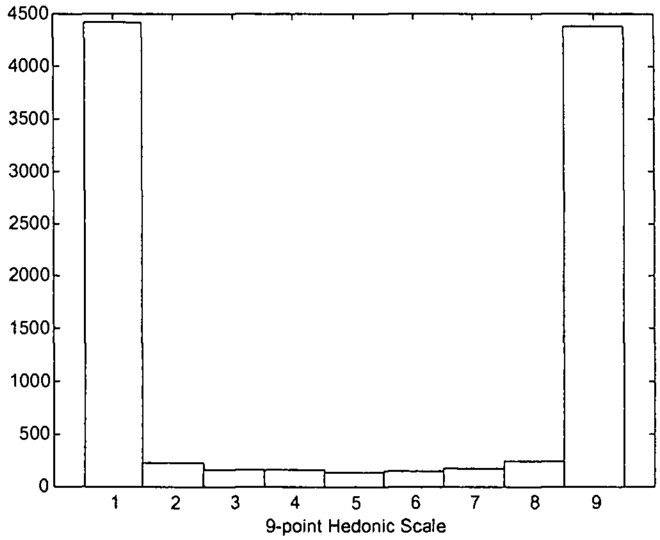


Figure 3.17 The distribution of 10000 simulated sensory consumer responses by the DL model (σ_R = 0.50, σ_E = 1.66, g = 0.00, n = 9). The observed mean = 4.98, variance = 6.86.

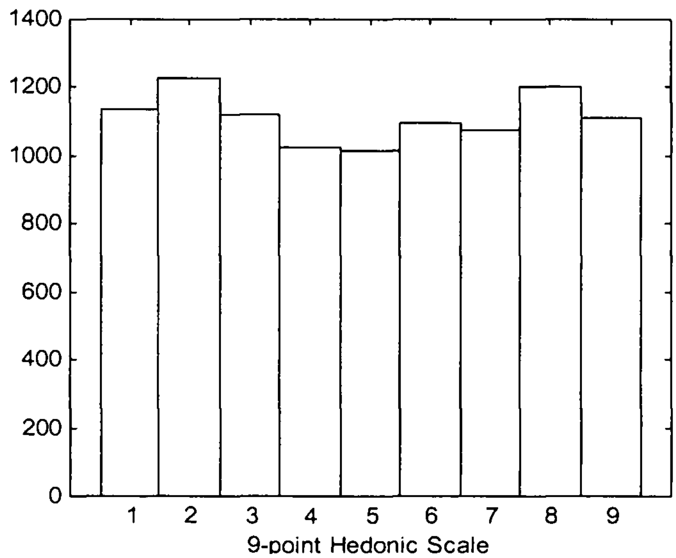


Figure 3.18 The distribution of 10000 simulated sensory consumer responses by the DL model ($\sigma_R = 0.50$, $\sigma_E = 1.60$, $g = 0.00$, $n = 9$). The observed mean = 5.00, variance = 6.63.

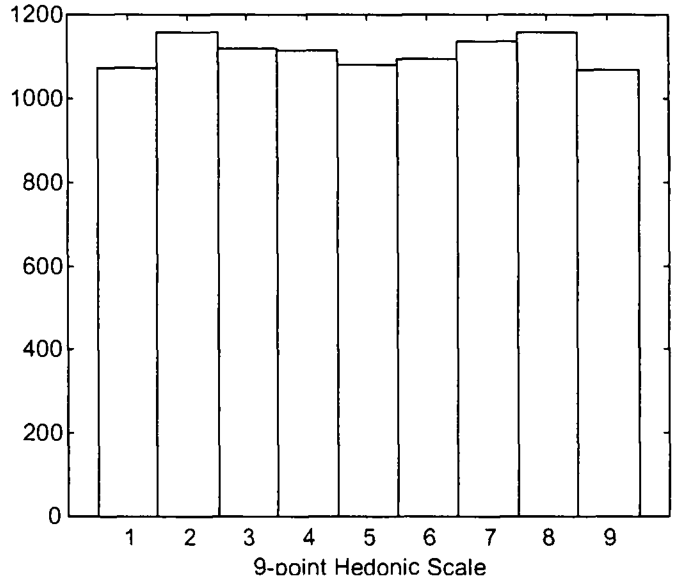


Figure 3.19 The distribution of 10000 simulated sensory consumer responses by the DL model ($\sigma_R = 0.65$, $\sigma_E = 1.40$, $g = 0.00$, $n = 9$). The observed mean = 5.02, variance = 6.18.

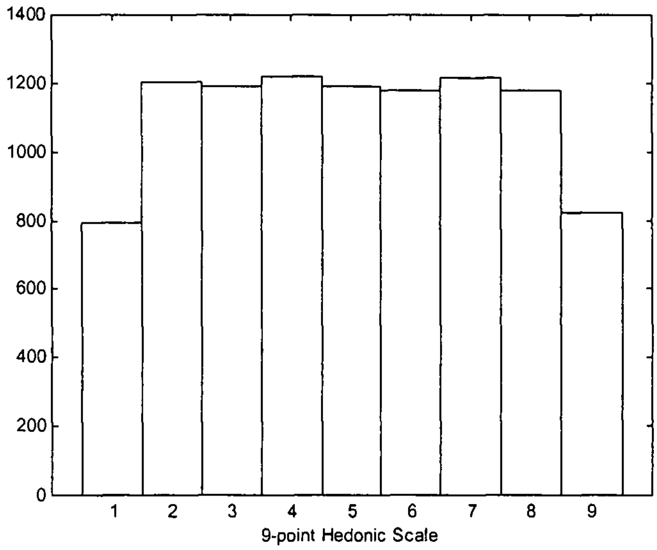


Figure 3.20 The distribution of 10000 simulated sensory consumer responses by the DL model ($\sigma_R = 0.30$, $\sigma_E = 0.40$, $g = 0.00$, $n = 9$). The observed mean = 4.98, variance = 1.21.

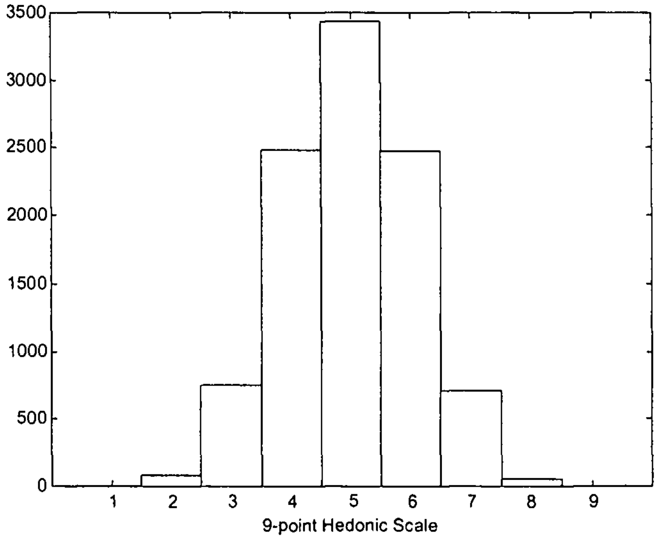


Figure 3.21 The distribution of 10000 simulated sensory consumer responses by the DL model ($\sigma_R = 0.10$, $\sigma_E = 0.10$, $g = 0.00$, $n = 9$). The observed mean = 5.00, variance = 0.12.

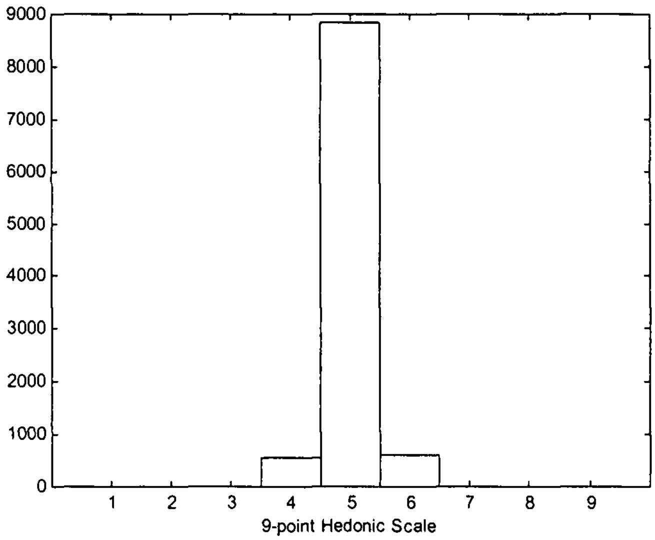


Figure 3.22 The distribution of 10000 simulated sensory consumer responses by the LB model ($\sigma_R = 0.0001$, $g = 0.00$, $n = 8$). The observed mean = 5.02, variance = 1.98.

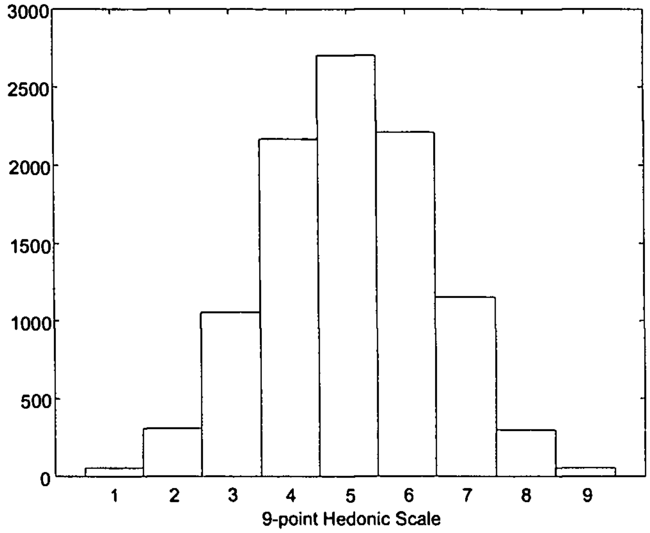
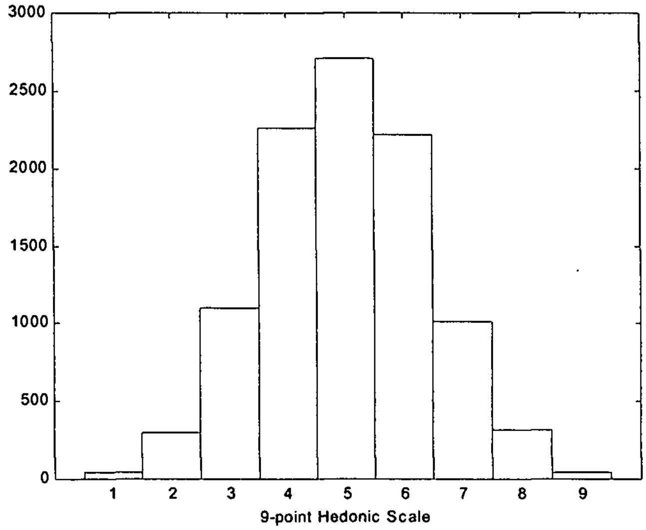


Figure 3.23 The distribution of 10000 simulated sensory consumer responses by the LBB model ($\sigma_R = 0.0001$, $g = 0.00$, $\gamma = 0.00001$, $n = 8$). The observed mean = 4.98, variance = 2.03.



LB or LBB model was roughly about 2.00. This is because the variance of the BB distribution:

$$V(Y) = np\,(1-p)[1+(n-1)\gamma]$$

and the inflation factor $[1+(n-1)\gamma_{jk}]$ is always greater than or equal to 1. Therefore, the LBB model always has an equal or even larger variance than the LB model.

Comparing with LB and LBB models, the LHG and LBHG models were relatively flexible and could be used to simulate data sets with lower variances. The observed variance of 10000 simulated responses by the LHG model was 0.23 with $\sigma_R = 0.0001$, $g = 0.00$, and $M = 500$ (Figure 3.24). Similarly, the observed variance of 10000 simulated responses by the LBHG model was 0.24 with an additional input: $\gamma = 0.00001$ (Figure 3.25). However, these models – LB, LBB, LHG and LBHG required more complex simulation algorithms and therefore longer run times on MATLAB™ used in this study.

Finally, the DL model was chosen as the simulation model to mimic sensory cross-cultural consumer responses for the next step. Figure 3.26 illustrated the distributions of 10000 simulated data in the central tendency case with observed means equaled to 5.0 ($\sigma_R^2 = 0.40$, $\sigma_E^2 = 0.35$, $g = 0.00$). Case A represented a situation where responses were using the whole scale ($k_1 = k_2 = 0$) (inputs of the truncation parameter: $k_1 = k_2 = 0$), while case B, C and D used a smaller range of the scale (2 – 8, 3 – 7 and 4 – 6 respectively). Case D simulated an extreme case that responses were highly centered.

Figure 3.24 The distribution of 10000 simulated sensory consumer responses by the LHG model ($\sigma_R = 0.0001$, $g = 0.00$, $M = 500$, $n = 8$). The observed mean = 5.00, variance = 0.23.

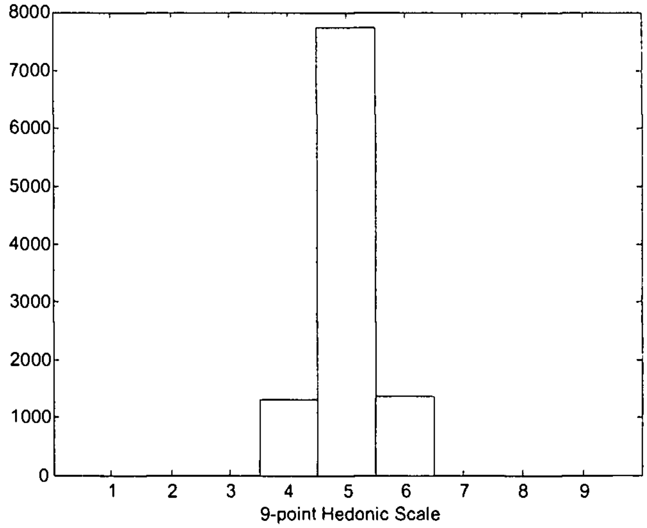


Figure 3.25 The distribution of 10000 simulated sensory consumer responses by the LBHG model ($\sigma_R = 0.0001$, $g = 0.00$, $\gamma = 0.00001$, $M = 500$, $n = 8$). The observed mean = 5.05, variance = 0.24.

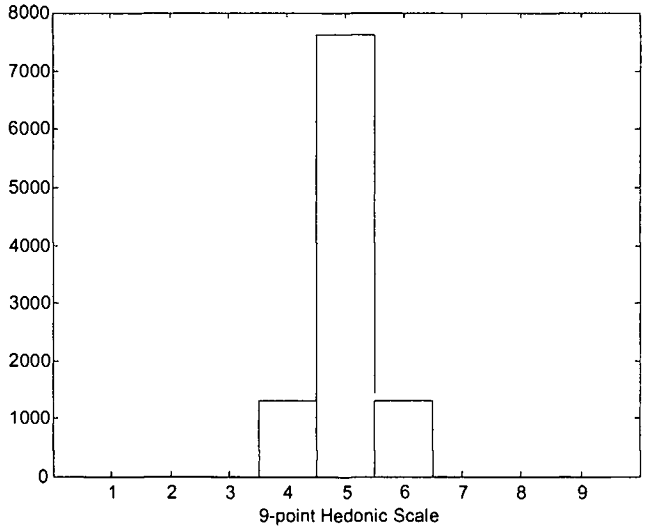


Figure 3.26 Central Tendency Case: the distribution of 10000 simulated sensory responses by DL model ($\sigma_R^2 = 0.35$, $\sigma_E^2 = 0.40$, $g = 0.00$, $n = 9$).

Case A: $k_1 = 0$, $k_2 = 0$, observed mean = 5.0, variance = 2.9;
B: $k_1 = 2$, $k_2 = 1$, observed mean = 5.0, variance = 1.8;
C: $k_1 = 4$, $k_2 = 2$, observed mean = 5.0, variance = 1.0;
D: $k_1 = 6$, $k_2 = 3$, observed mean = 5.0, variance = 0.4.

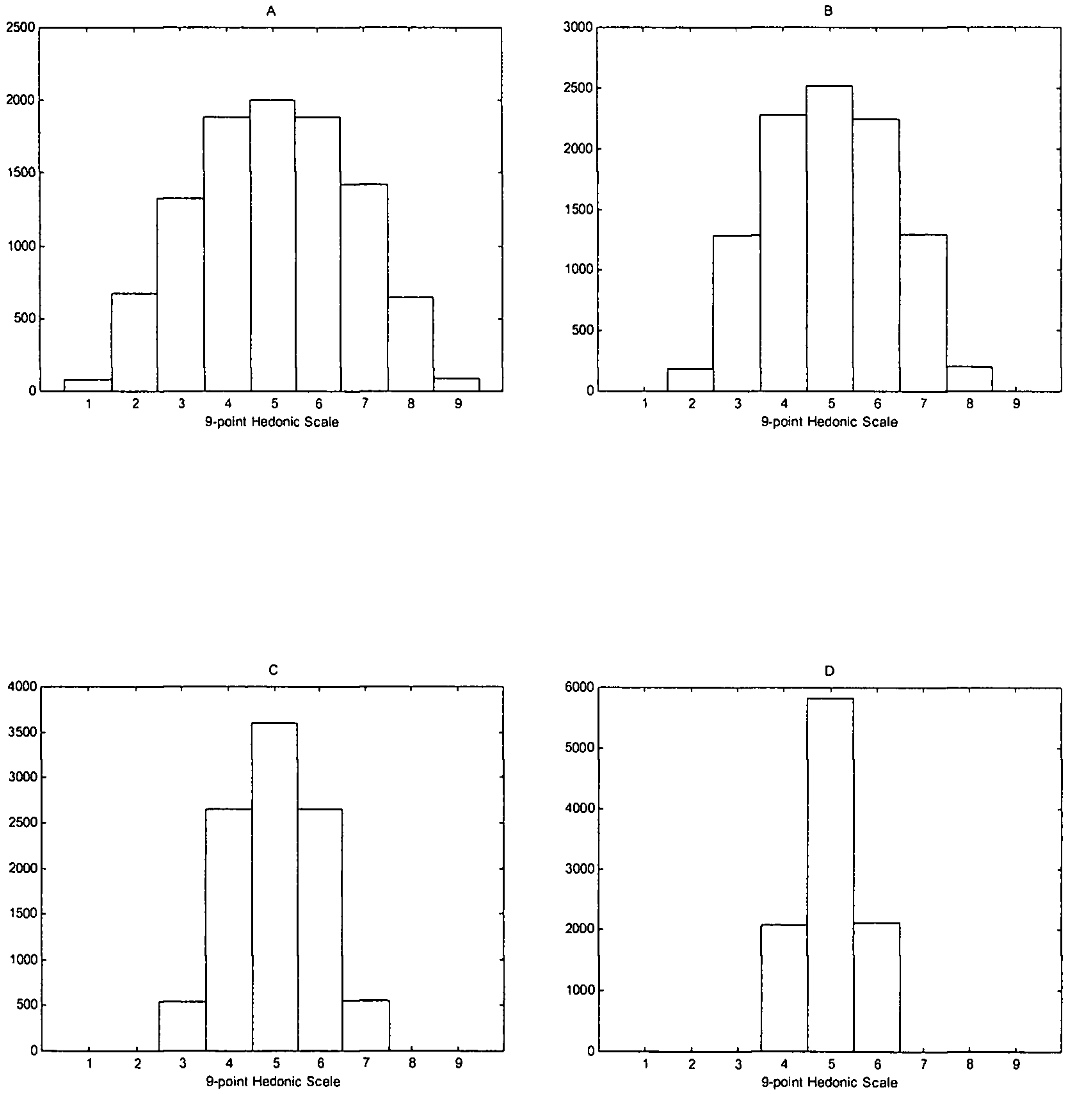
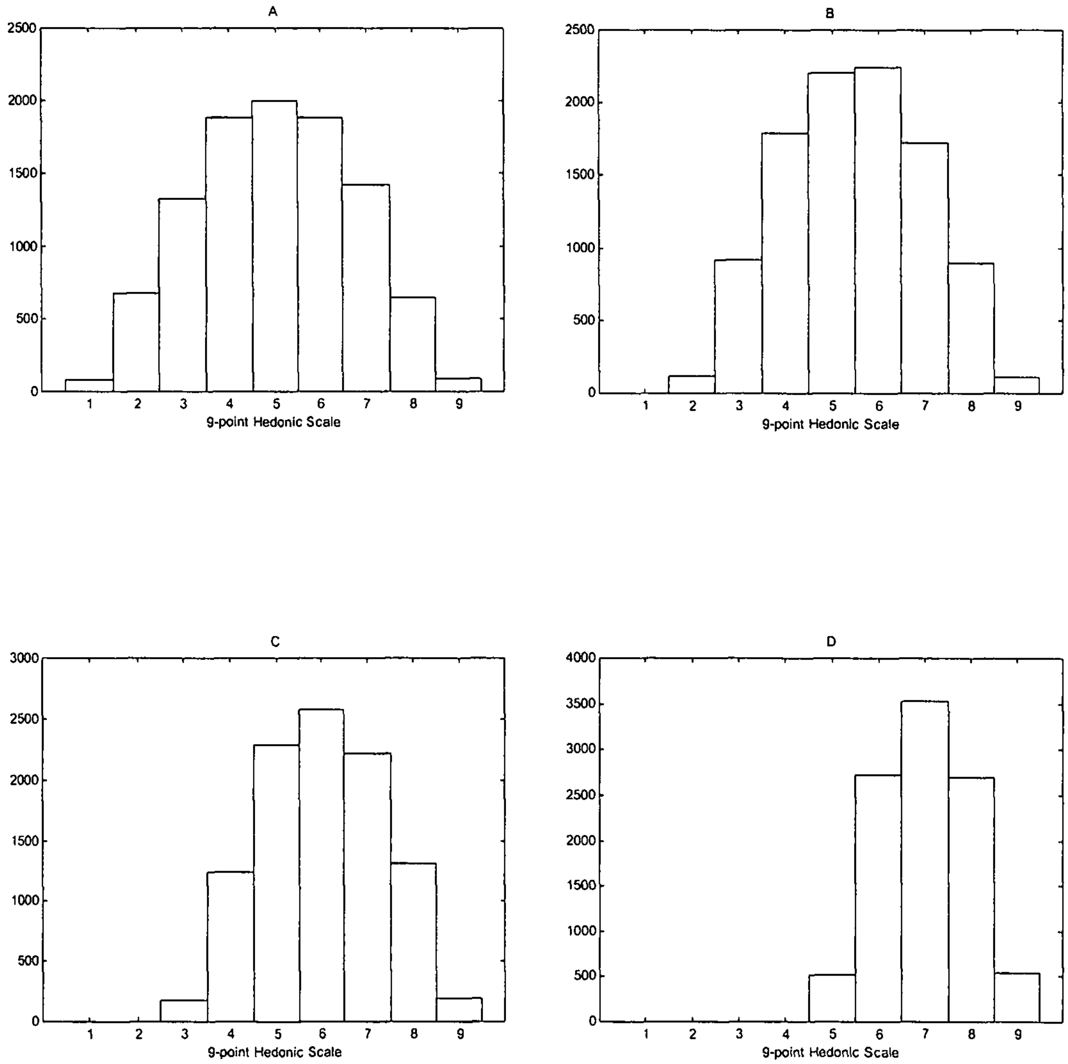


Figure 3.27 demonstrated the dislike avoidance case. Case A still represented a situation where responses were using the whole scale ($k_1 = k_2 = 0$). From case B to D, by increasing the input values of the truncation parameter k_1 and k_2, the distribution of simulated data shifted toward the right end of the scale to mimic more extreme responses.

Figure 3.27 Dislike Avoidance Case: the distribution of 10000 simulated sensory responses by DL model ($\sigma_R^2 = 0.35$, $\sigma_E^2 = 0.40$, $g = 0.00$, $n = 9$).

Case A: $k_1 = k_2 = 0$, observed mean = 5.0, variance = 2.9;
B: $k_1 = k_2 = 1$, observed mean = 5.5, variance = 2.3;
C: $k_1 = k_2 = 2$, observed mean = 6.0, variance = 1.8;
D: $k_1 = k_2 = 4$, observed mean = 7.0, variance = 1.0.



References

Amerine, M.A., Pangborn, R.M., and Roessler, E.B. (1965). Principles of Sensory Evaluation of Food. Academic Press Inc., New York.

Ayabe-Kanamura, S., Schicker, I., Laska, M., Hudson, R., Distel, H., Kobayakawa, T., and Saito, S. (1998). Differences in perception of everyday odors: a Japanese-German cross-cultural study. Chem. Senses 23: 31-38.

Bang, J.W., Schumacker, R.E., and Schlieve, P.L. (1998). Random-number generator validity in simulation studies: an investigation of normality. Educational and Psychological Measurement 58: 430.

Bertino, M. and Chan, M.M. (1986). Taste perception and diet in individuals with Chinese and European ethnic backgrounds. Chem. Senses 11: 229-241.

Bower, B. (1999). Cultures of reason: thinking styles may take Eastern and Western routes. Science News, vol.157, No. 4: 56-58.

Cardello, A.V. (1996). The role of the human senses in food acceptance, in H.L.Meisleman, & H.J.H.MacFie, Food Choice Acceptance and Consumption (pp. 1- 82), Blackie Academic & Professional, London.

Casella, G. and Berger, R.L. (1990). Statistical Inference. Duxbury Press, Wadsworth, Inc., Belmont, CA.

Chatfield, C. and Goodhardt, G.J. (1970). The beta-binomial model for consumer purchasing behavior. Appl. Statist. 19: 240-250.

Christensen, R. (1990). Log-linear Models. Springer-Verlag New York, Inc., New York, NY.

Cooper, R.A. and Weekes, A.J. (1983). Data, Models, and Statistical Analysis. Barnes and Noble Books, Totowa, NJ.

Cox, D.N., Clark, M.R. and Mialon, V.S. (2001). A cross-cultural methodological study of the uses of two common hedonic response scales. Food Qual. and Pref. 12: 119-131.

Edwards, L.K. (1993). Applied Analysis of Variance in Behavioral Science (Chapter 1. Analysis of variance overview). Marcel Dekker, Inc., New York, NY.

Ennis, D.M. and Bi, J. (1998). The beta-binomial model: Accounting for inter-trial variation in replicated difference and preference tests. J. Sensory Studies 13: 389-412.

Ennis, D.M. and Bi, J. (1998). A thurstonian variant of the beta-binomial model for replicated difference tests. J. Sensory Studies 13: 461-466.

Everitt, B.S. (1998). The Cambridge Dictionary of Statistics. Cambridge University Press, Cambridge, U.K.

Gacula, M.C. (1993). Design and Analysis of Sensory Optimization. Food and Nutrition Press, Inc, Trumbull.

Gay, C. and Mead, R. (1992). A statistical appraisal of the problem of sensory measurement. J. Sensory Studies 7: 205-228.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (1997). Bayesian Data Analysis. Chapman & Hall/CRC, Florida.

Glass, G.V., Peckham, P.D. and Sanders, J.R. (1972). Consequences of failure to meet assumptions underlying the fixed analysis of variance and covariance. Review of Educational Research 42: 237-288.

Griffiths, D.A. (1973). Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of a disease. Biometrics 29: 637-648.

Johnson, N.L. (1949). Systems of frequency curves generated by methods of translation. Biometrica 36: 149-176.

Jones, L.V., Peryam, D.R. and Thurstone, L.L. (1955). Development of a scale for measuring soldier's food preferences, Food Research, 20, 515-520.

Kelly, P.A. (1999). Overview of computer intensive statistical inference procedures. http://garnet.acns.fsu.edu/~pkelly/resampling.html.

Kleinman, J.C. (1973). Proportions with extraneous variance: single and independent samples. J. American Statistical Association 68 (341): 46-54.

Kuehl, R.O. (2000). Design of Experiments: Statistical Principles of Research Design and Analysis, 2nd Edition, Brooks / Cole, Pacific Grove, CA.

Lawless, H.T. and Malone, G.J. (1986). A comparison of rating scales: sensitivity, replicates and relative measurement. J. Sensory Studies 1: 155-174.

Lawless, H.T. and Heymann, H. (1998). Sensory evaluation of food, NY, Chapman Hall.

Lawless, H.T. (1997). Commentary on random vs. fixed effects for panelists. Food Qual. and Pref. 9: 163-164.

Lundahl, D.S. and McDaniel, M.R. (1988). The panelist effect – fixed or random? J. Sensory Studies 3: 113-121.

Lundahl, D.S. and McDaniel, M.R. (1990). Use of contrasts for the evaluation of panel inconsistency. J. Sensory Studies 5: 265-277.

Lundahl, D.S. and McDaniel, M.R. (1991). Influence of panel inconsistency on the outcome of sensory evaluations from descriptive panels. J. Sensory Studies 6: 145-157.

Lundahl, D.S. (1992). Influence of special consumer groups as a subset of respondents on the outcomes of consumer acceptance tests. Product Testing with Consumers for Research Guidance: Special Consumer Groups, 2nd Vol., American Society for Testing and Materials, Philadelphia.

Lundahl, D.S. (1997). Thurstonian models – an answer to Gridgeman's Paradox? http://camo.com/Publications/WPapers/thurstonian.html

Madden, L.V. and Hughes, G. (1994). BBD-Computer software for fitting the beta-binomial distribution to disease incidence data. Plant Dis. 78: 536-540.

Mage, D.T. (1980). An explicit solution for S_B parameters using four percentile points. Technometrics, vol. 22, No. 2.

Mattes, R. and Lawless, H.T. (1985). An adjustment error in optimization of taste intensity. Appetite 6: 103-114.

MATLAB™ User's Guide (1997). The MathWorks, Inc. Natick, MA.

Morrison, D.G. (1979). Purchase intentions and purchasing behavior. J. Marketing 43: 65-74.

Moskowitz, H.R. (1983). Product Testing and Sensory Evaluation of Foods, Moskowitz / Jacobs, Inc., Scarsdale, New York and Tasc Group, Inc., Evanston, Illinois.

Næs, T. (1990). Handling individual differences between assessors in sensory profiling. Food Qual. and Pref. 2: 187-199.

Næs, T. and Langsrud, Ø. (1998). Fixed or random assessors in sensory profiling? Food Qual. and Pref. 9: 145-152.

O'Mahony, M. (1982). Some assumptions and difficulties with common statistics for sensory analysis. Food Tech. 36: 75-82.

O'Mahony, M. (1986). Sensory Evaluation of Food, Marcel Dekker, Inc., New York.

Otake, M. and Prentice, R.L. (1984). The analysis of chromosomally aberrant cells based on beta-binomial distribution. Radiation Res. 98: 456-470.

Prescott, J., Laining, D., Bell, G., Yoshida, M., Gillmore, R., Allen, S., Yamazaki, K., and Ishii, R. (1992). Hedonic responses to taste solutions: a cross-cultural study of Japanese and Australians. Chem. Senses 17: 801-809.

Prescott, J., Bell, G., Gillmore, R., Yoshida, M., O'Sullivan, M., Korac, S., Allen, S., and Yamazaki, K. (1997). Cross-cultural comparisons of Japanese and Australian responses to manipulations of sweetness in foods. Food Qual. and Pref. 8: 45-55.

Prescott, J. (1998). Comparisons of taste perceptions and preferences of Japanese and Australian consumers: Overview and implications for cross-cultural sensory research. Food Qual. and Pref. 9: 393-402.

Ramsey, R.L. and Schafer, D.W. (1997). The statistical sleuth: a course in methods of data analysis. Wadsworth Publishing Company, Belmont, CA.

Rust, R.T. and Klompmaker, J.E. (1981). Improving the estimation procedure for the beta binomial TV exposure model. J. Marketing Res. 18: 442-448.

Sabavala, D.J. and Morrison, D.G. (1977). Television show loyalty: a beta-binomial model using recall data. J. Advertising Res. 17: 35-43.

Slifker, J.F. and Shapiro, S.S. (1980). The Johnson System: selection and parameter estimation. Technometrics 22: 239-246.

Skellam, J.G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as a variable between sets of trials. J. Roy. Statist. Soc. Ser. B 10: 257-265.

Spiegel, M.R. (1975). Theory and Problems of Probability and Statistics, McGraw-Hill, Inc.

Thurstone, L. L. (1927). A law of comparative judgment. Psychological Review 34: 273-286,

Wackerly, D.D., Mendenhall III, W. and Scheaffer, R.L. (1996). Mathematical Statistics with Applications, Duxbury Press, Wadsworth Publishing Company, Belmont, CA.

Wilcox, R. (1993). Applied Analysis of Variance in Behavioral Science (edited by Edwards, L.K.): Chapter 9. Robustness in ANOVA. Marcel Dekker, Inc., New York, NY.

Wilkinson, C. and Yuksel, D. (1997). Modeling differences between panelists in use of measurement scales. J. Sensory Studies 12: 55-68.

Williams, D.A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. Biometrics 31: 949-952.

Yamamoto, E. and Yanagimoto, T. (1994). Statistical methods for the β-Binomial model in teratology. Environmental Health Perspectives Supplements 102, Suppl 1: 25-31.

Yeh, L.L. and Lundahl, D.S. (1999). Use of hedonic scale among Chinese, Koreans, and Thais. M.S. Thesis. Oregon State University.

CHAPTER 4

A SIMULATION ASSESSMENT OF MULTICULTURAL DATA ANALYSIS STRATEGIES

Shuo Wang, Dave Plaehn, David Lundahl and Mina McDaniel

To be submitted to the Journal of Sensory Studies

May 9, 2003

Abstract

The objective of this study is based upon the generalized simulation tool – Discrete Empirical Logit model previously developed in Chapter 3, to evaluate the robustness of ANOVA and its empirical power on detecting product difference under different cross-cultural cases.

The study simulated a simplified situation – two segments (100 respondents within each segment) and two testing products. First, for each case, 10000 data sets which represent cross-cultural responses were randomly generated by the Discrete Empirical Logit (DL) model. The simulation study was focused on six cross-cultural cases – differences in scale usage (central tendency and dislike avoidance), product preference (crossover), location variation, acceptance variation and segment effect. Second, the data sets were analyzed using both ANOVA reduced and full model, and their empirical powers were compared.

Results revealed that when respondents perform consistently normal (*MSE* $\approx$ 1), ANOVA was very robust in the central tendency and dislike avoidance case; at worst it lost less than 15% in power of detecting product differences. On the other hand, highly inconsistent performance of respondents could result in significant loss in ANOVA power. No significant difference in power was detected between the reduced and full model except in the crossover case. The full model implied a significant effect of segment by product interaction where segments differed in product preferences.

Introduction

In cross-cultural sensory studies, respondents from different countries or cultures are procedurally asked to rate the overall liking, flavor liking, texture liking for certain products. Judges are generally considered random rather than fixed effect in cross-cultural studies (Lundahl and McDaniel, 1988). Usually these respondents are untrained consumers and their responses are collected as ratings on a numerical scale, such as the balanced nine-point category scale (hedonic scale) introduced by the food research section of the U.S. Army Quartermaster Corps in the 1950s (Jones, Peryam and Thurstone, 1955). For example, the word categories of product liking or disliking used for the 9-point hedonic scale are from "dislike extremely" (score 1) to "like extremely" (score 9) with a neutral category "neither like nor dislike" (score 5) at the center of the scale (Table 4.1). Data are analyzed by statistic methods, usually analysis of variance (ANOVA). For the descriptive or affective panels, scaled responses were analyzed to evaluate sensitivity, scale location, agreement, reproductively and acceptability of panelists.

There are some complex issues that arise in the field of cross-cultural sensory evaluation. Responses from different cultures may combine with culture habits, social, historical, economic, and religious reasons. Cross-cultural differences in the use of a measurement scale may exist both in location (where on the scale the scores tend to be located) and in dispersion (the range of the scale used) (Wilkinson and Yuksel, 1996).

Table 4.1 Example of juice product questionnaires.

Panelist number: Country: Sample:

After taste the sample, please answer (check X) the following questions:

1. How do you like this product overall?

Dislike Extremely	Dislike Very Much	Dislike Moderately	Dislike Slightly	Neither Like Nor Dislike	Like Slightly	Like Moderately	Like Very Much	Like Extremely
☐	☐	☐	☐	☐	☐	☐	☐	☐
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)

2. How do you like the sweetness of this product?

Dislike Extremely	Dislike Very Much	Dislike Moderately	Dislike Slightly	Neither Like Nor Dislike	Like Slightly	Like Moderately	Like Very Much	Like Extremely
☐	☐	☐	☐	☐	☐	☐	☐	☐
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)

3. What is the sweetness intensity of this product?

Not Sweet At All	None To Slightly	Slightly Sweet	Slightly to Moderately	Moderately Sweet	Moderately To Very Much	Very Much Sweet	Very Mucl To Extremely	Extremely Sweet
☐	☐	☐	☐	☐	☐	☐	☐	☐
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)

4. How did this product meet your expectations in overall liking?

Extremely Worse	Much Worse	Moderately Worse	Slightly Worse	The Same	Slightly Better	Moderately Better	Much Better	Extremely Better
☐	☐	☐	☐	☐	☐	☐	☐	☐
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)

Previous studies have shown that cultures may influence measurement scale usage, For example, Bertino and Chan (1986) investigated relationships between taste perception and diet in individuals with Chinese and European ethnic backgrounds. Besides the taste difference between the two cultures, the study suggested that it was possible that the Chinese, in an attempt to be polite, assigned higher pleasantness ratings to all taste stimuli. Yeh and Lundahl (1999) compared 9-point hedonic scale usage between consumers from the US and Pacific Rim cultures. The hedonic scale was translated directly from English to their respective languages. Evidence showed that consumers from the Pacific Rim culture used the 9-point hedonic scale differently from American respondents. They tend to agree with each other and not score extremely low or high. For instance, Thais hide their dislike feelings by not using the lower part of the scale.

Central tendency and dislike avoidance are two common cases that respondents from different cultures, while sharing a common perception of the differences among products, use a consistently smaller portion of the scale range, scoring more or less. Central tendency is observed in scoring when respondents tend to agree with each other and seldom use the extreme values of a scale. While in the dislike avoidance case, respondents have more positive attitudes toward products and avoid using the lower part of the scale (e.g., scores below 5 on a 9-point hedonic scale). Both cases result in skewed data distributions with less variance. This different scale usage case is illustrated in Figure 4.1 and 4.2. In this example, 70 Indonesians and 51 Chinese responded in a sports drink study (Chung,

Figure 4.1 The distribution of 70 Indonesian overall liking scores on sweetness of one sports drink product.

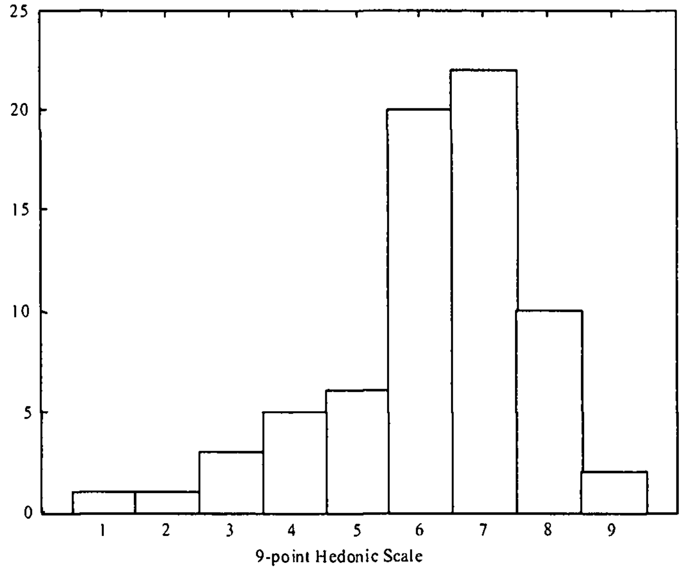
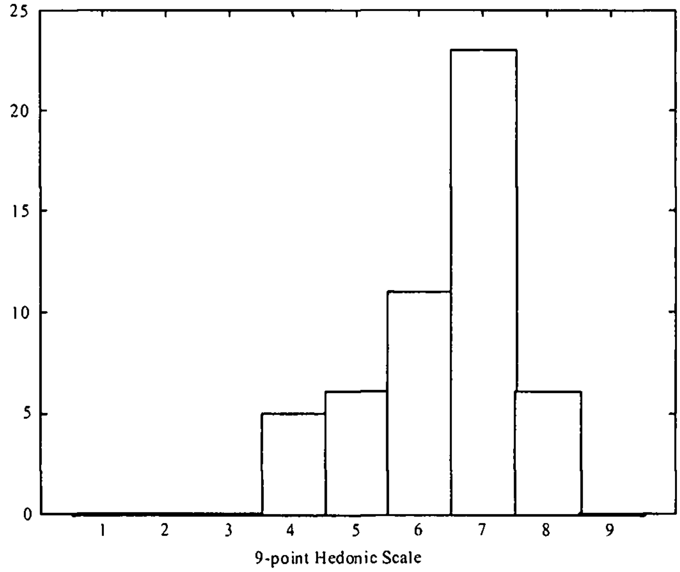


Figure 4.2 The distribution of 51 Chinese overall liking scores on sweetness of one sports drink product.



1999). Apparently the 70 Indonesians were using the whole range of the 9-point hedonic scale while the 51 Chinese were using a small portion of the scale, from 4 to 8.

In addition to the central tendency and dislike avoidance case, location variation within segment case reflects the within cultural heterogeneity in scaling location. Also, since an individual's taste may vary from day to day. Acceptance variation within segment reflects the within cultural heterogeneity in product preference. It is the variation in preferences based upon individual replication level and is associated with the mean square error (*MSE*) from the ANOVA model.

Further more, Bertino and Chan, 1986; Prescott et al., 1992, 1997 and 1998; Ayabe-Kanamura et al., 1998; Yeh and Lundahl, 1999 have showed cross-cultural difference may exist in taste preferences. When cultures differ, the patterns in responses can lead to a crossover case. For example, two segments hold opposite taste preference over products. Without segmenting, differences among products might be diminished and results in a Type II error (not detecting a real product difference).

Most descriptive experiments produce data matrices appropriate for ANOVA. Products, judges, and replications are the usual factors, with additional factors depending upon the variables manipulated among products, such as ingredient levels or processing treatments (Lawless, 1996). The crossover effect or the interaction between product and the segment can be partitioned to examine the product effect.

In sensory studies, one of the applications of using ANOVA method is to detect product differences. Usually analysts advocate segmenting cross-cultural consumer populations in a conservative way because combining data from different cultures may create the problem of violating the ANOVA underlying assumptions and risk in losing its testing power. However, ANOVA is regarded as being robust to certain departures from the assumptions. This is a critical problem in the field of sensory evaluation that cannot be solved easily by any other demonstrated way except by simulations.

Simulation studies are frequently used when investigating expected measurement outcomes and examining expected statistical outcomes (Bang et al., 1998). Simulation has a merit that samples can often be generated from a probability distribution, even when the density function cannot be explicitly integrated. Monte Carlo simulation is a common approach to examine expected statistical outcomes by using random-number generators.

A few simulation studies have been applied in the sensory field. For example, Næs and Langsrud (1998) advocated the use of the mixed ANOVA models to reduce the interaction effect by removing the scaling effect mathematically. They applied Monte Carlo simulation and power function to compute Type I errors and power of the F-tests for a number of cases with different use of the scale. Ennis and Bi (1998) conducted Monte Carlo experiments to investigate the behavior of difference or preference tests with over-dispersed binomial data based on the binomial model and the beta-binomial model.

Lundahl (1992) used simulation strategy to determine the influence of special consumer groups as a small subset (10 to 30%) of respondents on the outcomes of consumer acceptance tests. He concluded that a segment by treatment interaction with crossover patterns or minority acceptance can contribute to errors of type II or I, and large crossovers from 10% special consumers or small crossovers from 20% to 30% differently responding ("special") consumers can significantly increase type II error.

In this study, we aimed to use simulation to evaluate the robustness and testing power of the reduced and full ANOVA models under different cross-cultural situations. To accomplish this, the Discrete Empirical Logit (DL) model which was designed in the previous chapter was applied to simulate random cross-cultural responses.

Methods

Data Structure

Data sets were created to simulate the following simplified situation:

- 2 segments – reference segment and comparison segment
- 100 respondents in each segment
- Each respondent measures 2 products – A and B without replication
- 3 levels of product difference – none, moderate and large.

The simulation algorithm is as follows:

1. Generate 10000 data sets for each case;
2. Run ANOVA models on each data set;
3. Calculate the empirical power of ANOVA – percentage of times when a significant product difference was detected.

Simulation Model – DL Model

The Discrete Empirical Logit (DL) model was chosen as a simulator from the previously five developed models to generate sensory consumer 9-point hedonic data. It was preceded in two stages. First, input parameters ($\beta_0, \beta_1, \beta_2, \beta_3$) and noise ($R_j$, $E_{i(j)k}$) to the Empirical Logit model:

$$logit(p_{i(j)kl}) = \beta_0 + R_{i(j)} + \beta_1 X_j + \beta_2 X_k + \beta_3 X_j X_k + E_{i(j)kl} \tag{4.1}$$

where

$$logit(p_{i(j)kl}) = log\frac{p_{i(j)kl}}{1 - p_{i(j)kl}} \tag{4.2}$$

Let

$$g(X_j, X_k \mid \beta_0, \beta_1, \beta_2, \beta_3) = \beta_0 + \beta_1 X_j + \beta_2 X_k + \beta_3 X_j X_k \tag{4.3}$$

then the Equation 4.1 can be written as

$$logit(p_{i(j)kl}) = R_{i(j)} + g(X_j, X_k \mid \beta_0, \beta_1, \beta_2, \beta_3) + E_{i(j)kl} \tag{4.4}$$

$i = 1, 2, \ldots, 100$ (respondent), $j = 1, 2$ (segment), $k = 1, 2$ (product)

where β_0 is the grand mean, β_1, β_2 and β_3 are the parameters of three fixed effects: segment, product, segment by product interaction, respectively. X_j, X_k are the explanatory variables that represent factor segment, product respectively, and X_jX_k represents the interaction between X_j and X_k. X_j, X_k both have two levels. For example

$$\begin{cases} X_j = 0 \text{ for segment 1} \\ X_j = 1 \text{ for segment 2} \end{cases} \qquad \begin{cases} X_k = 0 \text{ for product A} \\ X_k = 1 \text{ for product B} \end{cases}$$

$R_{i(j)}$ represents a random location effect for the i^{th} respondent within j^{th} segment, which reflects the variation of mean score location of respondents on a scale. $E_{i(j)kl}$ represents within segment consistency, which related to the acceptance variation within segment based upon the individual preference. In this simulation study, both $R_{i(j)}$ and $E_{i(j)kl}$ are normally distributed, $R_{jk} \sim N(0, \sigma_R^2)$ and $E_{i(j)kl} \sim N(0, \sigma_E^2)$. In addition, σ_R and σ_E are independent of product and segment. Next step is to transform $logit(p_{i(j)kl})$ to p-value by the equation

$$p_{i(j)kl} = \frac{exp[logit(p_{i(j)kl})]}{1 + exp[logit(p_{i(j)kl})]} \qquad (4.5)$$

On the second stage, input the p-value to the binning model:

$$Y_{i(j)kl} = Floor[(n - k_1)p_{i(j)kl}] + k_2 + 1 \qquad (4.6)$$

and calculate the random discrete categorical responses, where n is the number of response scale categories. In this study, $n = 9$ (number of scale categories), MATLAB™ command "*floor*" means *round* the elements to the nearest integers towards minus infinity, k_1 and k_2 are truncation parameters:

- In the central tendency case, the 9-point hedonic scale was double truncated by k_1 and k_2 where k_1 is an even integer and

$$k_2 = k_1 / 2 \tag{4.7}$$

That is, if k_1 is chosen as 2, 4, 6, then k_2 is equal to 1, 2, 3, and the scores of response $Y_{i(j)kl}$ will vary from 2 to 8, 3 to 7, 4 to 6 respectively.

- In the dislike avoidance case, the scale was one side truncated where k_1 is an integer and

$$k_2 = k_1 \tag{4.8}$$

If k_1 is chosen as 1, 2, 4, then k_2 is equal to 1, 2, 4, and the scores will be allowed to range from 2 to 9, 3 to 9, 5 to 9 respectively.

Reference Segment

For the reference segment, expected powers of detecting none, moderate (0.2 for mean scores on a 9-point hedonic scale) and large product differences (0.4 for mean scores) are 5.0% (Type I error rate), 50.0% and 90.0% (Table 4.2) respectively. (Note: the three product difference levels were defined by the associated expected powers – no product difference was associated with 5.0% Type I error rate, moderate difference was associated with 50.0% ANOVA power, large difference was associated with 90.0% ANOVA power). The input parameters were specified as $\sigma_R^2 = 0.4$, $\sigma_E^2 = 0.35$, $k_1 = k_2 = 0$. As a result, the mean square of respondent ($MS_{respondent}$) was 4 and the resulting mean square error (MSE) was 1.

Table 4.2 The expected ANOVA results for the reference segment at 3 product difference levels – none, moderate and large (significant level = 0.05, $\sigma_R^2 = 0.4$, $\sigma_E^2 = 0.35$, $k_1 = k_2 = 0$).

Product Difference Level	*Expected Mean Scores*		*Expected Power of ANOVA*	$MS_{respondent}$	*MSE*
	Product A	*Product B*			
None[1]	5.0	5.0	5.0%	4	1
Moderate[2]	5.0	5.2	50.0%	4	1
Large[3]	5.0	5.4	90.0%	4	1

Note: [1] $\beta_0 = 0$, $\beta_1 = 0$, $\beta_2 = 0$, $\beta_3 = 0$;
[2] $\beta_0 = 0$, $\beta_1 = 0$, $\beta_2 = 0.12$, $\beta_3 = 0$;
[3] $\beta_0 = 0$, $\beta_1 = 0$, $\beta_2 = 0.2$, $\beta_3 = 0$.

Comparison Segment

The comparison segment was modified in six cases:

Case 1 – Segments (reference segment and comparison segment) both agree in product preference and scale usage;

Case 2 – Segments agree in product preference, but disagree in scale usage, e.g. central tendency and dislike avoidance situation (varying the value of truncation parameters k_1 and k_2 in Equation 4.7 and 4.8);

Case 3 – Segments agree in scale usage, but disagree in product preference, e.g. crossover case (varying the value of β_3 parameter in Equation 4.3);

Case 4 – Variation in scaling usage within the comparison segment (adding noise by increasing the value of R_{jk});

Case 5 – Variation in acceptance (among respondents) within the comparison segment (varying the value of $E_{i(j)k}$);

Case 6 – The segment location effect (segments differ in scale location usage, varying the value of β_1 parameter).

ANOVA Models

Analysis of variance was performed to examine the segment, product and the consistency of responses within each segment. The ANOVA table accounting for this consumer segmentation was presented in Table 4.3. This expanded

ANOVA was named as the "full model". It included six sources of variation: segment, respondent within segment, product, segment by product interaction, and error; while the two-way ANOVA – "reduced model" (Table 4.4) only included product, respondent and error. Segment and product were considered as fixed factors, so was the segment by product interaction. Respondent within segment was considered as a random effect. The empirical power of ANOVA was calculated for both the full and reduced models.

Results and Discussion

Case 1 – Agreement in Scale Usage and Preference

Case 1 simulated a situation where no difference between the two groups. That is, both segments used the whole scale (1-9) with the same product preference (Table 4.5). By keeping the other variables as constants ($\sigma_R^2 = 0.4$, $\sigma_E^2 = 0.35$, $k_1 = k_2 = 0$), the ANOVA powers were 5.3%, 50.1% and 90.3% respectively (α-level = 0.05) at three product difference levels – none ($\beta_2 = 0$), moderate ($\beta_2 = 0.12$) and large ($\beta_2 = 0.2$). The results were very close to the theoretical expected value (5.0%, 50.0% and 90% for the three product difference levels) and were further used as the baseline situation when comparing with the other cases. In this case, the $MS_{respondent}$ was 4.4 and MSE was 1.4. Therefore, results were as expected suggesting Case 1 was a good reference case.

Table 4.3 ANOVA full model.

Source of Variation	*DF*	*SS*	*MS*	*F-Value*
Segment	1	$SS_{segment}$	$MS_{segment}$	$\frac{MS_{segment}}{MS_{respondent(segment)}}$
Respondent (segment)	198	$SS_{respondent(segment)}$	$MS_{respondent(segment)}$	
Product	1	$SS_{product}$	$MS_{product}$	$\frac{MS_{product}}{MSE}$
Segment by Product	1	$SS_{segment \times product}$	$MS_{segment \times product}$	$\frac{MS_{segment \times product}}{MSE}$
Error	198	SSE	MSE	

Table 4.4 ANOVA reduced model.

Source of Variation	*DF*	*SS*	*MS*	*F-Value*
Product	1	$SS_{product}$	$MS_{product}$	$\frac{MS_{product}}{MSE}$
Respondent	199	$SS_{respondent}$	$MS_{respondent}$	$\frac{MS_{respondent}}{MSE}$
Error	199	SSE	MSE	

Table 4.5 Case 1 – Agreement in scale usage and product preference: results of ANOVA at 3 product difference levels – none, moderate and large, based on 10000 simulations ($\sigma_R^2 = 0.4$, $\sigma_E^2 = 0.35$, $k_1 = k_2 = 0$).

Significant Level	*Product Difference Level*	*Actual Type I Error Rate*		*Segment by Product (Full Model)*	*MS of respondent*	*MSE*	*MS of Segment*
		Reduced Model	*Full Model*				
0.05	None[1]	5.3%	5.3%	5.3%	4.4	1.4	4.4
	Moderate[2]	50.0%	50.1%	5.1%	4.4	1.4	4.4
	Large[3]	90.3%	90.3%	4.9%	4.4	1.4	4.4
0.1	None	10.1%	10.1%	10.2%	4.4	1.4	4.4
	Moderate	62.4%	62.4%	10.1%	4.4	1.4	4.4
	Large	94.6%	94.6%	10.0%	4.4	1.4	4.4

Note: [1] $\beta_0 = 0$, $\beta_1 = 0$, $\beta_2 = 0$, $\beta_3 = 0$;
[2] $\beta_0 = 0$, $\beta_1 = 0$, $\beta_2 = 0.12$, $\beta_3 = 0$;
[3] $\beta_0 = 0$, $\beta_1 = 0$, $\beta_2 = 0.2$, $\beta_3 = 0$.

Case 2 – Agreement in Preference and Disagreement in Scale Usage

In Case 2, the two segments shared the same product preference but the comparison segment performed central tendency or dislike avoidance on the use of scale. In this situation, other parameters were kept as constants, while the input values of the truncation parameters k_1 and k_2 varied.

In the central tendency case at α-level 0.05, combining the responses from the two segments, the ANOVA power dropped roughly less than 10% at the moderate or large product difference levels. For example, for the large product difference, using scale range 2-8, 3-7 and 4-6, the power of the reduced model dropped about 0.1%, 3% and 8% respectively (Table 4.6 and Figure 4.3).

In the dislike avoidance case at α-level 0.05, without segmentation of the responses leads to less than 3% drop in power. For example, for the large product difference, using scale range 2-9, 3-9 and 5-9, the power of the reduced model dropped 0%, 0.1% and less then 3% respectively (Table 4. 6 and Figure 4.4).

In both situations, 50% of total respondents that from the comparison segment used a portion of the scale instead of the whole. This seems like "a scale compressed from one side or both sides". Imagine two roughly normal distributions of product A and B in the center of the scale with certain location difference. The scores are scattered around the mean value with a given variance. While the end of the scale moves towards the two distributions, both data on the left or right side of the scale will get less and less room to spread. This scale compression forces

Table 4.6 Case 2 (to be continued) – Agreement in product preference, but disagreement in scale usage: power of ANOVA at 3 product difference levels – none, moderate and large, based on 10000 simulations ($\sigma_R^2 = 0.4$, $\sigma_E^2 = 0.35$).

Product Difference Level & Simulation Cases		*Scale Usage*	$MS_{respondent}$	$MS_{respondent(segment)}$	MSE	$MS_{segment}$
None[1]	Central Tendency	2--8	2.7	2.7	0.9	2.6
		3--7	1.4	1.4	0.5	1.4
		4--6	0.6	0.6	0.2	0.6
	Dislike Avoidence	2--9	3.8	3.5	1.0	53.7
		3--9	3.6	2.7	0.8	202.1
		5--9	5.4	1.4	0.4	800.7
Moderate[2]	Central Tendency	2--8	2.7	2.7	0.9	2.8
		3--7	1.4	1.4	0.5	1.9
		4--6	0.6	0.6	0.2	1.7
	Dislike Avoidence	2--9	3.8	3.5	1.0	50.8
		3--9	3.6	2.7	0.8	192.0
		5--9	5.2	1.4	0.4	760.2
Large[3]	Central Tendency	2--8	2.7	2.7	0.8	3.3
		3--7	1.4	1.4	0.5	2.8
		4--6	0.6	0.6	0.2	3.8
	Dislike Avoidence	2--9	3.6	3.4	1.0	49.8
		3--9	3.6	2.7	0.8	186.3
		5--9	4.7	1.4	0.4	735.3

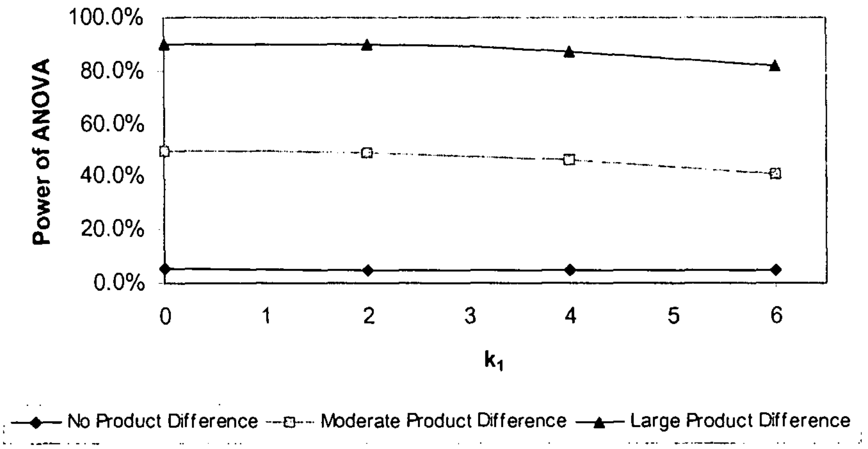
Note: [1] No product difference: $\beta_0 = 0$, $\beta_1 = 0$, $\beta_2 = 0$, $\beta_3 = 0$;
[2] Moderate product difference: $\beta_0 = 0$, $\beta_1 = 0$, $\beta_2 = 0.12$, $\beta_3 = 0$;
[3] Large product difference: $\beta_0 = 0$, $\beta_1 = 0$, $\beta_2 = 0.2$, $\beta_3 = 0$.

Table 4.6 Case 2 (continued)

Significant Level	*Product Difference Level & Simulation Cases*		*Scale Usage*	*Empirical Power of ANOVA* Reduced Model	*Empirical Power of ANOVA* Full Model	*Segment*Product (in Full Model)*
0.05	None[1]	Central Tendency	2--8	5.0%	5.0%	5.0%
			3--7	5.0%	5.1%	5.0%
			4--6	4.9%	5.0%	5.0%
		Dislike Avoidence	2--9	4.8%	4.8%	5.1%
			3--9	5.0%	5.0%	5.0%
			5--9	5.0%	5.0%	5.1%
0.05	Moderate[2]	Central Tendency	2--8	49.0%	49.0%	5.5%
			3--7	46.4%	46.5%	7.8%
			4--6	40.7%	40.9%	13.4%
		Dislike Avoidence	2--9	49.5%	49.6%	5.0%
			3--9	49.0%	49.0%	5.5%
			5--9	46.3%	46.3%	7.9%
0.05	Large[3]	Central Tendency	2--8	90.2%	90.2%	6.8%
			3--7	87.2%	87.2%	14.3%
			4--6	81.9%	81.9%	29.4%
		Dislike Avoidence	2--9	90.3%	90.3%	5.8%
			3--9	90.2%	90.2%	6.8%
			5--9	87.8%	87.8%	13.8%
0.1	None	Central Tendency	2--8	9.9%	9.8%	10.0%
			3--7	10.0%	10.1%	10.2%
			4--6	9.5%	9.7%	9.9%
		Dislike Avoidence	2--9	9.8%	9.8%	9.9%
			3--9	9.9%	9.8%	10.0%
			5--9	10.1%	10.1%	10.1%
0.1	Moderate	Central Tendency	2--8	61.8%	61.7%	10.8%
			3--7	59.5%	59.5%	14.4%
			4--6	53.5%	53.5%	21.9%
		Dislike Avoidence	2--9	62.2%	62.2%	10.1%
			3--9	61.8%	61.7%	10.8%
			5--9	58.5%	58.5%	14.7%
0.1	Large	Central Tendency	2--8	94.4%	94.4%	12.4%
			3--7	92.5%	92.5%	22.8%
			4--6	89.3%	89.3%	41.3%
		Dislike Avoidence	2--9	94.7%	94.7%	10.9%
			3--9	94.4%	94.4%	12.4%
			5--9	93.1%	93.1%	22.5%

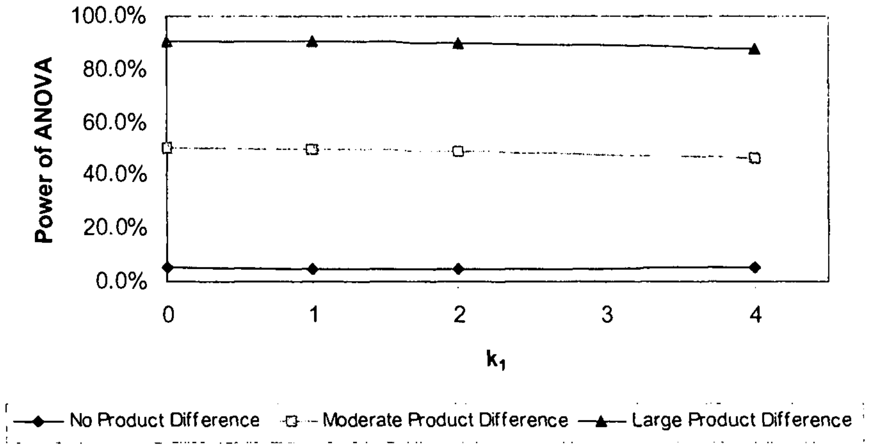
Note: [1] No product difference: $\beta_0 = 0$, $\beta_1 = 0$, $\beta_2 = 0$, $\beta_3 = 0$;
[2] Moderate product difference: $\beta_0 = 0$, $\beta_1 = 0$, $\beta_2 = 0.12$, $\beta_3 = 0$;
[3] Large product difference: $\beta_0 = 0$, $\beta_1 = 0$, $\beta_2 = 0.2$, $\beta_3 = 0$.

Figure 4.3 Case 2.1 – Central tendency within the comparison segment: power of ANOVA at three product difference levels – none, moderate & large, based on 10000 simulation runs (significant level = 0.05)



Note: $k_l = 0$: which the scale usage is from 1 to 9 (whole 9-point hedonic scale);
$k_l = 2$: which the scale usage is from 2 to 8;
$k_l = 4$: which the scale usage is from 3 to 7;
$k_l = 6$: which the scale usage is from 4 to 6.

Figure 4.4 Case 2.2 – Dislike avoidance within the comparison segment: power of ANOVA at three product difference levels – none, moderate & large, based on 10000 simulation runs (significant level = 0.05)



Note: $k_l = 0$: which the scale usage is from 1 to 9 (whole 9-point hedonic scale);
$k_l = 1$: which the scale usage is from 2 to 9;
$k_l = 2$: which the scale usage is from 3 to 9;
$k_l = 4$: which the scale usage is from 5 to 9.

reduction of the dispersion of the distribution, and results in less location differences between the two products. Apparently, this will decrease the power of ANOVA in detecting the product differences.

Meanwhile, the mean square errors (*MSE*) and the variation of the respondents ($MS_{respondent}$ in the ANOVA reduced model) within segment decreased in both cases as the scale truncation got more serious. Unlike in the central tendency case, a largely increased segment effect was revealed in the dislike avoidance case.

At α-level = 0.10, ANOVA had a higher detecting power than α-level = 0.05 (Table 4.6). Results also showed that no significant difference of ANOVA powers was detected between the reduced and full model. This is because when the product by segment interaction effect is small,

$$SS_{product}\,(Full\ model) = SS_{product}\,(Reduced\ model) \tag{4.9}$$

and both have the same degree of freedom, thus

$$MS_{product}\,(Full\ model) = MS_{product}\,(Reduced\ model) \tag{4.10}$$

Since $SSE(Reduced\ model) = SSE + SS_{segment \times product}\,(Full\ model)$ (4.11)

$$df_{SSE} = 199\,(Reduced\ model) \approx df_{SSE} = 198\,(Full\ model) \tag{4.12}$$

If $SS_{segment \times product}$ is small, then

$$SSE\,(Reduced\ model) \approx SSE\,(Full\ model) \tag{4.13}$$

$$MSE\,(Reduced\ model) \approx MSE\,(Full\ model) \tag{4.14}$$

$$\frac{MS_{product}}{MSE}(Reduced\ model) \approx \frac{MS_{product}}{MSE}(Full\ model) \tag{4.15}$$

Therefore, under the simulation conditions tested in this study, both models gave the similar results of F-value and p-value for the product effect.

As a conclusion, the central tendency and dislike avoidance cases are not the driving factors for significant ANOVA power loss. Therefore, as long as segments are in agreement with taste preferences, results from different segments can be combined.

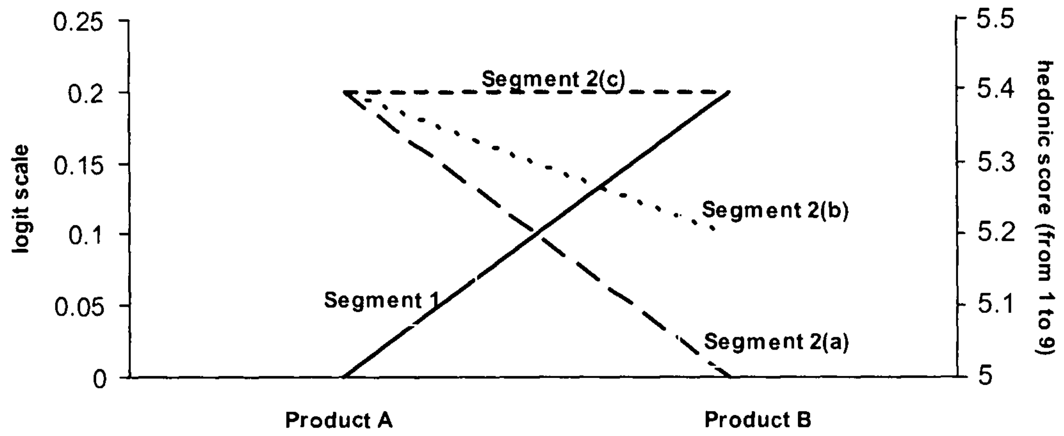
<u>Case 3 – Disagreement in Preference and Agreement in Scale Usage</u>

For simplicity, three situations were used to illustrate the case where the two segments agreed in the scale usage but disagreed in the product preferences (Figure 4.5). The reference segment detected a product difference and scored B over A. On the other hand, the comparison segment scored the product A over B (Situation 1 and 2) or showed no preference (Situation 3). The reduced model failed to detect any product difference. However, the full model revealed a high chance of detecting a significant segment by product interaction effect in all three situations (Table 4.7).

Correspondingly, in Lundahl's paper (1992) of determining the influence of special consumer groups as a small subset of respondents, he concluded that large

Figure 4.5 Case 3 – Agreement in scale usage, but disagreement in product preference: an illustration of three situations where the segment 1 always scores product B (mean 5.4) over A (mean 5.0) and

1. Segment 2(a): segment 2 scores product A (mean 5.4) over B (mean 5.0)
2. Segment 2(b): segment 2 scores product A (mean 5.4) over B (mean 5.2)
3. Segment 2(c): segment 2 has no preference on the two products (both means 5.4)



Note of the three situations:

Situation 1: $\beta_0 = 0$, $\beta_1 = 0.2$, $\beta_2 = 0.2$, $\beta_3 = -0.4$
Situation 2: $\beta_0 = 0$, $\beta_1 = 0.2$, $\beta_2 = 0.2$, $\beta_3 = -0.3$
Situation 3: $\beta_0 = 0$, $\beta_1 = 0.2$, $\beta_2 = 0.2$, $\beta_3 = -0.2$

Table 4.7 Case 3 – Agreement in scale usage, but disagreement in product preference: power of ANOVA at three situations, based on 10000 simulations ($\sigma_R^2 = 0.4$, $\sigma_E^2 = 0.35$, $k_1 = k_2 = 0$).

		Empirical Power of ANOVA		*Segment by*		*MSE*		
Significant Level	*Situation*	*Reduced Model*	*Full Model*	*Product (Full Model)*	*MS of respondent*	*Reduced Model*	*Full Model*	*MS of Segment*
0.05	1	4.6%	9.4%	90.4%	4.4	1.5	1.4	4.4
	2	11.9%	20.0%	67.5%	4.4	1.5	1.4	5.3
	3	37.3%	49.5%	36.7%	4.4	1.4	1.4	8.2
0.1	1	5.2%	10.1%	94.6%	4.4	1.5	1.4	4.4
	2	12.4%	20.6%	78.2%	4.4	1.5	1.4	5.3
	3	37.7%	49.9%	48.6%	4.4	1.4	1.4	8.2

Note of the three situations:

Situation 1: $\beta_0 = 0$, $\beta_1 = 0.2$, $\beta_2 = 0.2$, $\beta_3 = -0.4$

Situation 2: $\beta_0 = 0$, $\beta_1 = 0.2$, $\beta_2 = 0.2$, $\beta_3 = -0.3$

Situation 3: $\beta_0 = 0$, $\beta_1 = 0.2$, $\beta_2 = 0.2$, $\beta_3 = -0.2$

crossovers from 10% special consumers or small crossovers from 20 to 30% special consumers could significantly increase the Type II error.

In our situation, since the two segments had equal sample sizes, we can think that the comparison segment as special consumers group represents 50% of the total respondents. Therefore, crossover response from one of the consumer segments can highly reduce the power of ANOVA in detecting product differences. A significant segment by product interaction effect from the full model implies a great product preference disagreement among the cultural segments. Under this situation, segmentation by different cultures is a necessary step in cross-cultural consumer data analysis. The need for segmentation in the analysis also depends on the type of different consumer response, number of respondents, and degree of disparity between the segments. Failure to design consumer tests to be able to segment in the analysis can result in errors of Type I or II (Lundahl, 1992).

Based on the results from Case 3, we conclude that for cross-cultural sensory consumer data analysis it is important to always evaluate for segment differences and use the ANOVA full model for detecting product differences.

Case 4 – Location Variation within Segment

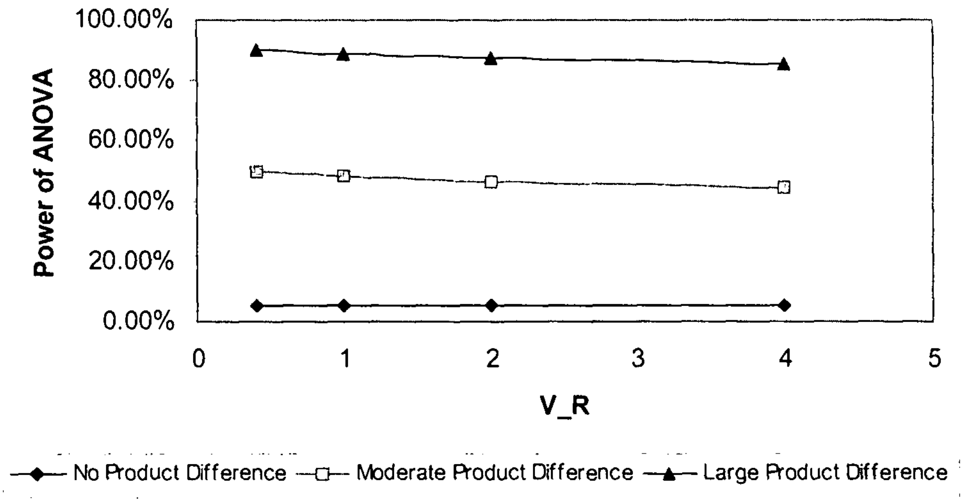
Emboldened by our experience, in the cross-cultural consumer research the mean square of respondents ($MS_{respondent}$) in the ANOVA reduced or full model usually varies in value from 3 to 10 (on a 9-point hedonic scale). Thus, the simulated $MS_{respondent}$ was 7.6, 11.2 and 15.4 corresponding to the scaling location variation (σ_R^2) within the comparison segment varied in 1, 2 and 4 (Figure 4.6 and Table 4.8). As location variation parameter σ_R^2 increased, there was less homogeneity in scaling location within the comparison segment, and thus less power to distinguish the two products and more disagreement between the segments.

Again, there was no significant difference between the results of the two models (ANOVA full vs. reduced model). At α-level = 0.05, the power of ANOVA dropped 6% at worst for the moderate product difference, and 5% at worst for the large product difference (Figure 4.6 and Table 4.8). At α-level = 0.10, the performance of ANOVA was slightly better than at α-level = 0.05. Therefore, we conclude that the location variation within a segment has little effect on the power to detect product differences.

Case 5 – Acceptance Variation within Segment

Based upon our experiences on consumer studies, the acceptance variation within a segment, which reflects the respondents' heterogeneity in product preference, usually varies from value ± 2 on a 9-point hedonic scale. By increasing the respondents' personal random noise on acceptance within the comparison segment (σ_E^2), the *MSE* in the ANOVA results increased from 1.6, 2.0 to 2.6 (Figure 4.7 and Table 4.9). There was no difference between the results of reduced

Figure 4.6 Case 4 – Increasing location variation within the comparison segment: power of ANOVA at moderate and large product difference levels, based on 10000 simulations (σ_E^2 = 0.35, significant level = 0.05), V_R represents σ_R^2.



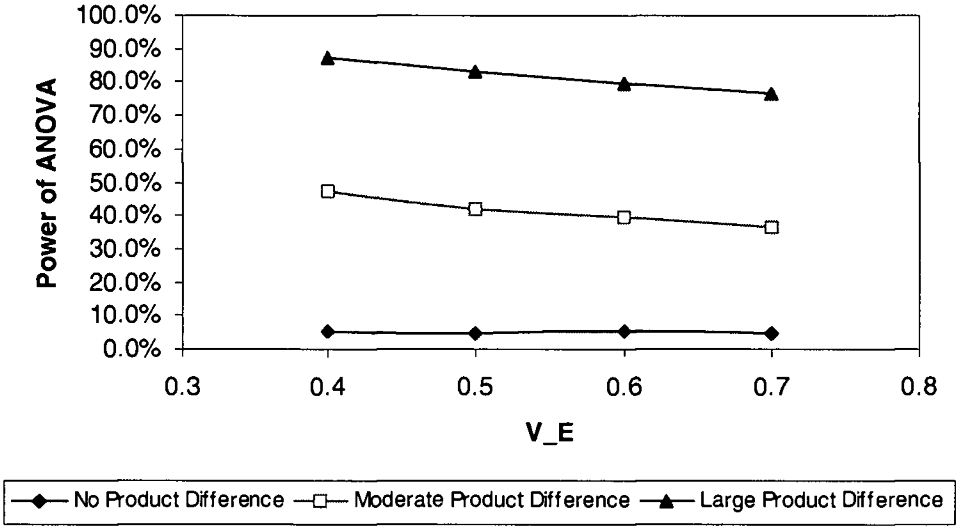
Note: σ_R^2 = 1 for small location variation;
σ_R^2 = 2 for moderate location variation;
σ_R^2 = 4 for large location variation;

Table 4.8 Case 4 – Increasing location variation within the comparison segment: results of ANOVA at moderate and large product difference levels, based on 10000 simulations ($\sigma_E^2 = 0.35$).

Sig. Level	*Product Difference Level*	*Location Variation*	*Empirical Power of ANOVA*		*Segment by Product (Full Model)*	*MS of respondent*	*MSE*	*MS of Segment*
			Reduced	*Full Model*				
0.05	Moderate[4]	Small[1]	48.6%	48.7%	5.0%	7.6	1.2	7.6
		Moderate[2]	46.3%	46.3%	5.3%	11.2	1.0	11.2
		Large[3]	44.1%	43.9%	6.8%	15.4	0.8	16
0.05	Large[5]	Small	88.8%	88.8%	5.6%	7.6	1.2	7.8
		Moderate	87.2%	87.2%	6.4%	11.2	1.0	11.6
		Large	85.4%	85.4%	9.0%	15.4	0.8	16.4
0.1	Moderate	Small	61.3%	61.3%	9.7%	7.6	1.2	7.6
		Moderate	59.0%	59.0%	10.6%	11.2	1.0	11.2
		Large	56.6%	56.6%	12.6%	15.4	0.8	16
0.1	Large	Small	94.1%	94.1%	11.0%	7.6	1.2	7.8
		Moderate	92.8%	92.8%	12.0%	11.2	1.0	11.6
		Large	91.6%	91.6%	15.8%	15.4	0.8	16.4

Note: [1] $\sigma_R^2 = 1$ for small location variation;
[2] $\sigma_R^2 = 2$ for moderate location variation;
[3] $\sigma_R^2 = 4$ for large location variation;
[4] $\beta_0 = 0$, $\beta_1 = 0$, $\beta_2 = 0.12$, $\beta_3 = 0$;
[5] $\beta_0 = 0$, $\beta_1 = 0$, $\beta_2 = 0.2$, $\beta_3 = 0$.

Figure 4.7 Case 5 – Increasing acceptance variation within the comparison segment: power of ANOVA at moderate and large product difference levels, based on 10000 simulations ($\sigma_R^2 = 0.4$, significant level = 0.05), *V_E* represents σ_E^2.



Note: $\sigma_E^2 = 0.4$ for small acceptance variation;
$\sigma_E^2 = 0.5$ for moderate acceptance variation;
$\sigma_E^2 = 0.7$ for large acceptance variation;

Table 4.9 Case 5 – Increasing acceptance variation within the comparison segment: results of ANOVA at moderate and large product difference levels, based on 10000 simulations ($\sigma_R^2 = 0.4$).

Sig. Level	*Product Difference Level*	*Acceptance Variation*	*Empirical Power of ANOVA*		*Segment by Product (Full Model)*	*MSE*	*MS of Respondent*	*MS of Segment*
			Reduced	*Full Model*				
0.05	Moderate[4]	Small[1]	47.5%	47.5%	5.0%	1.6	4.6	4.5
		Moderate[2]	41.9%	41.9%	5.0%	2	4.8	4.8
		Large[3]	36.7%	36.8%	5.3%	2.6	5.2	5.4
0.05	Large[5]	Small	87.7%	87.7%	5.4%	1.6	4.6	4.6
		Moderate	83.2%	83.1%	5.4%	2	4.8	4.8
		Large	76.4%	76.4%	5.1%	2.6	5.2	5.2
0.1	Moderate	Small	60.0%	60.0%	9.8%	1.6	4.6	4.5
		Moderate	54.6%	54.6%	9.9%	2	4.8	4.8
		Large	48.7%	48.7%	10.3%	2.6	5.2	5.4
0.1	Large	Small	93.2%	93.2%	10.6%	1.6	4.6	4.6
		Moderate	90.1%	90.2%	10.1%	2	4.8	4.8
		Large	84.8%	84.8%	10.1%	2.6	5.2	5.2

Note: [1] $\sigma_E^2 = 0.4$ for small acceptance variation;
[2] $\sigma_E^2 = 0.5$ for moderate acceptance variation;
[3] $\sigma_E^2 = 0.7$ for large acceptance variation;
[4] $\beta_0 = 0, \beta_1 = 0, \beta_2 = 0.12, \beta_3 = 0$;
[5] $\beta_0 = 0, \beta_1 = 0, \beta_2 = 0.2, \beta_3 = 0$.

model and full model. At large $\sigma_E^2 = 0.7$ and moderate product difference, power dropped 13%; and for the large product difference, power dropped 14% (α-level = 0.05) (Figure 4.7 and Table 4.9).

As a conclusion, heterogeneity in product acceptance within a segment ($MSE = 2.5 \sim 3$) can have a moderate effect (14% decline) on the ANOVA power to detect product differences.

Case 6 – Segment Location Effect

In Equation 4.3, parameter β_l represents a fixed segment location effect, or the difference on the scaling location between the two segments (e.g. two different cultures) as showed in Figure 4.8. Based on the Thurstonian hypothesis, the responses to sensory stimuli are normally distributed on an internal psychological scale of liking or the logit scale with range ($-\infty$, ∞). However, when the responses are transformed to a 9-point hedonic scale, which has a limited range from 1 to 9, the "end of scale compression" will most likely change the shape of the normal distributions as the differences of location between segment increase (Figure 4.9). The equal amount of product difference shrinks towards the ends of the scale and therefore the power of ANOVA decreases.

As the value of parameter β_l increased from 1 to 3, for the moderate product difference level, the reduced or full model lost 3%, 8%, and 17% for the small, moderate and large segment location effect, respectively; for the large

Figure 4.8 Case 6 – Segment location effect: an illustration of $g(X_j, X_k \mid \beta_0, \beta_1, \beta_2, \beta_3)$ for two products A, B and two segments 1, 2.

$g(X_j = 0, X_k = 0 \mid \beta_0, \beta_1, \beta_2, \beta_3) = \beta_0$ (product A and segment 1)
$g(X_j = 0, X_k = 1 \mid \beta_0, \beta_1, \beta_2, \beta_3) = \beta_0 + \beta_2$ (product B and segment 1)
$g(X_j = 1, X_k = 0 \mid \beta_0, \beta_1, \beta_2, \beta_3) = \beta_0 + \beta_1$ (product A and segment 2)
$g(X_j = 1, X_k = 1 \mid \beta_0, \beta_1, \beta_2, \beta_3) = \beta_0 + \beta_2 + \beta_1$ (product B and segment 2)

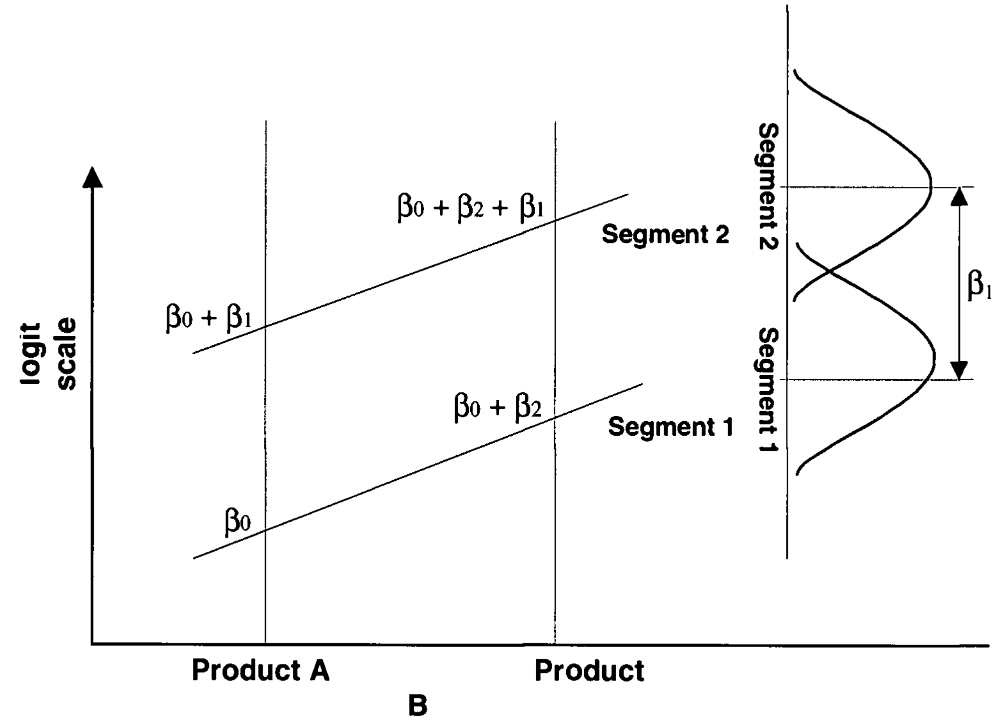
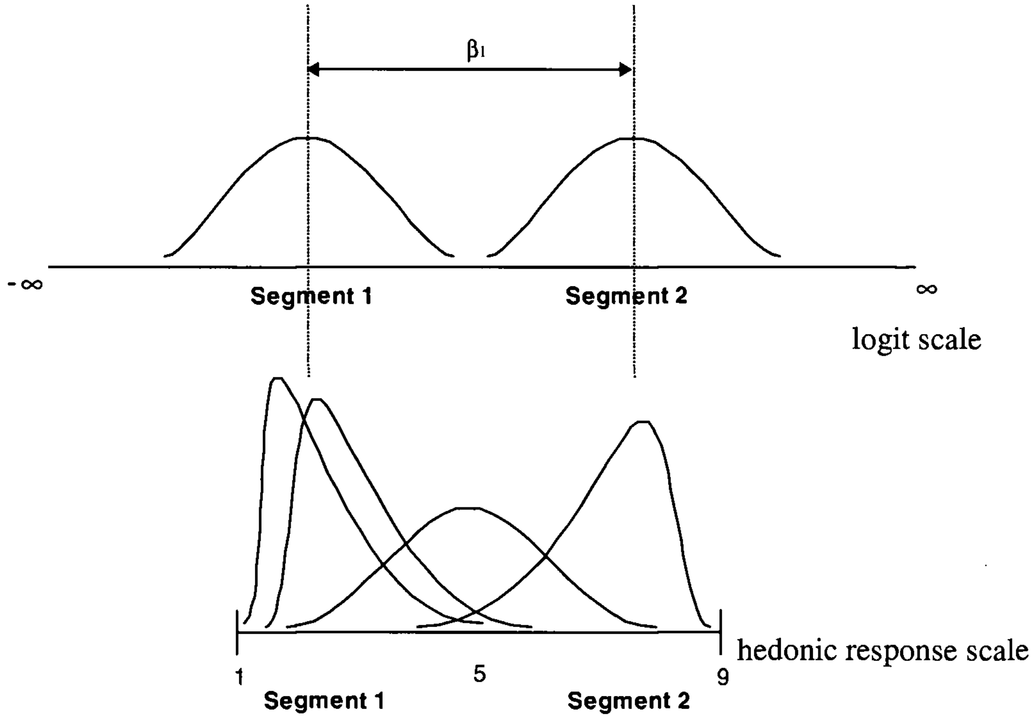


Figure 4.9 Responses are normally distributed on a logit scale ($-\infty, \infty$). However, when they are transformed to a 9-point hedonic scale, "end of scale compression" changes the shape of the distributions.



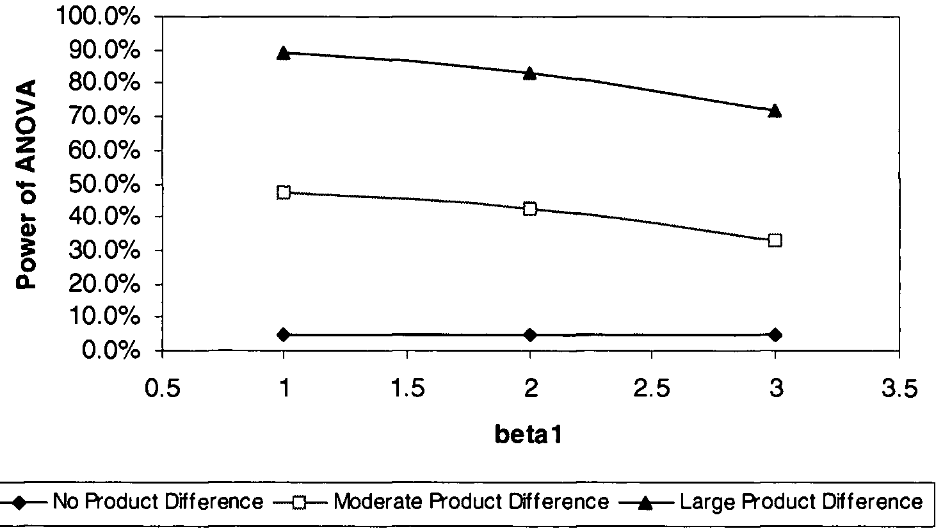
product difference level, the ANOVA model lost 2%, 7%, 18% for three segment location effects, respectively (α-level = 0.05) (Figure 4.10 and Table 4.10).

As a conclusion, when use a hedonic scale or any scale with a limited range, a large segment location effect may result in a situation which the distribution of the responses is heavily compressed towards the ends of the scale, and thus ANOVA has less power in detecting the product differences.

Conclusions

The analysis of variance table summarizes our knowledge about variability in the respondents from the study. Proper interpretation of the analysis will help to understand the statistical simulation model. Five variance components were estimated by the ANOVA full model such as segments, respondents within segment, products, segment by product interaction and random error. The variance component of segments indicates the location for segment differences. The variance component of respondents within segment shows the consistency in location within segments, and it increases as the variance of input parameter $R_{i(j)}$ increases. The variance from the products indicates the product differences. The segment by product interaction reflects the segment differences in product liking. The variance of random error is an estimate of the consistency among individuals within segments for product liking. Also, the variation of individuals from replication to replication to score products is confounded within the random error.

Figure 4.10 Case 6 – Segment location effect: power of ANOVA at moderate and large product difference levels, based on 10000 simulations ($\sigma_R^2 = 0.4$, $\sigma_E^2 = 0.35$, significant level = 0.05), beta1 represents β_1.



Note: beta1 (β_1) = 1 for small segment location effect;
beta1 = 2 for moderate segment location effect;
beta1 = 3 for large segment location effect;

Table 4.10 Case 6 – Segment location effect: results of ANOVA at moderate and large product difference levels, based on 10000 simulations ($\sigma_R^2 = 0.4$, $\sigma_E^2 = 0.35$).

Sig. Level	*Product Difference Level*	*Segment Location Effect*	*Empirical Power of ANOVA*		*MS of Respondent*	*MS of Respondent within Segment*	*MSE*	*MS of Segment*
			Reduced	*Full Model*				
0.05	Moderate[4]	Small[1]	47.2%	47.2%	6.6	3.2	1.0	662.4
		Moderate[2]	42.5%	42.5%	11.0	1.4	0.6	1945.2
		Large[3]	33.2%	33.5%	14.2	0.2	0.2	2786.2
0.05	Large[5]	Small	89.2%	89.2%	6.4	3	1.0	652.8
		Moderate	83.4%	83.4%	10.8	1.2	0.4	1901.8
		Large	72.0%	72.1%	13.8	0.2	0.2	2693.4
0.1	Moderate	Small	60.1%	60.1%	6.6	3.2	1.0	662.4
		Moderate	54.9%	54.9%	11.0	1.4	0.6	1945.2
		Large	45.9%	46.1%	14.2	0.2	0.2	2786.2
0.1	Large	Small	94.1%	94.1%	6.4	3	1.0	652.8
		Moderate	90.2%	90.1%	10.8	1.2	0.4	1901.8
		Large	81.7%	81.8%	13.8	0.2	0.2	2693.4

[1] $\beta_1 = 1$ for small segment location effect;
[2] $\beta_1 = 2$ for moderate segment location effect;
[3] $\beta_1 = 3$ for large segment location effect;
[4] $\beta_0 = 0$, $\beta_2 = 0.12$, $\beta_3 = 0$;
[5] $\beta_0 = 0$, $\beta_2 = 0.2$, $\beta_3 = 0$.

By varying the input value of parameter $E_{i(j)kl}$, we increase/decrease the variation of random error.

This study investigated the robustness of ANOVA to the scale usage differences, which generally lose at worst about 10% in power. In addition, a high *MSE* (around 3) from the ANOVA result indicated a moderate power loss (power drop approximately 14%). On the other hand, disagreement in product preference among segments could also highly affect the ANOVA power and therefore the segmentation of cross-cultural responses is highly recommended.

When a small crossover effect exists, the reduced and full models give similar results in terms of power. However, the full model is preferred in evaluating the cases where segments differ in product preferences.

As a conclusion, when conducting cultural studies, we recommend using the full ANOVA model to detect the interaction between the segment and product effect. From the scope of this study, if segments agree on product differences, the results of different segments can be combined for ANOVA.

In practice, this study demonstrated the value of simulation in sensory research to address difficult issues in sensory data analysis. It provided valuable information on the approximate power of ANOVA techniques. By simulating different cross-cultural situations and specifying input parameters, we can approximate the power of ANOVA. For further research, we can also vary the values of the sample size, location parameter, variances, and number of testing-samples and segments in order to simulate more applicable sensory consumer data and test the empirical power of ANOVA under those situations.

References

Amerine, M.A., Pangborn, R.M., and Roessler, E.B. (1965). Principles of Sensory Evaluation of Food. Academic Press Inc., New York.

Ayabe-Kanamura, S., Schicker, I., Laska, M., Hudson, R., Distel, H., Kobayakawa, T., and Saito, S. (1998). Differences in perception of everyday odors: a Japanese-German cross-cultural study. Chem. Senses 23: 31-38.

Bang, J.W., Schumacker, R.E., and Schlieve, P.L. (1998). Random-number generator validity in simulation studies: an investigation of normality. Educational and Psychological Measurement 58: 430.

Bertino, M. and Chan, M.M. (1986). Taste perception and diet in individuals with Chinese and European ethnic backgrounds. Chem. Senses 11: 229-241.

Casella, G. and Berger, R.L. (1990). Statistical Inference. Duxbury Press, Wadsworth, Inc., Belmont, CA.

Chatfield, C. and Goodhardt, G.J. (1970). The beta-binomial model for consumer purchasing behavior. Appl. Statist. 19: 240-250.

Christensen, R. (1990). Log-linear Models. Springer-Verlag New York, Inc., New York, NY.

Chung, S. (1999). Cross-cultural sweetness preferences for a sports-drink. M.S. Thesis. Oregon State University.

Cooper, R.A. and Weekes, A.J. (1983). Data, Models, and Statistical Analysis. Barnes and Noble Books, Totowa, NJ.

Cox, D.N., Clark, M.R. and Mialon, V.S. (2001). A cross-cultural methodological study of the uses of two common hedonic response scales. Food Qual. and Pref. 12: 119-131.

Edwards, L.K. (1993). Applied Analysis of Variance in Behavioral Science (Chapter 1. Analysis of variance overview). Marcel Dekker, Inc., New York, NY.

Ennis, D.M. (1998). Foundations of sensory science and a vision for the future. Food Technology 52: 78-84.

Ennis, D.M. and Bi, J. (1998). The beta-binomial model: Accounting for inter-trial variation in replicated difference and preference tests. J. Sensory Studies 13: 389-412.

Ennis, D.M. and Bi, J. (1998). A thurstonian variant of the beta-binomial model for replicated difference tests. J. Sensory Studies 13: 461-466.

Everitt, B.S. (1998). The Cambridge Dictionary of Statistics. Cambridge University Press, Cambridge, U.K.

Gacula, M.C. (1993). Design and Analysis of Sensory Optimization. Food and Nutrition Press, Inc, Trumbull.

Gay, C. and Mead, R. (1992). A statistical appraisal of the problem of sensory measurement. J. Sensory Studies 7: 205-228.

Glass, G.V., Peckham, P.D. and Sanders, J.R. (1972). Consequences of failure to meet assumptions underlying the fixed analysis of variance and covariance. Review of Educational Research 42: 237-288.

Griffiths, D.A. (1973). Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of a disease. Biometrics 29: 637-648.

Johnson, N.L. (1949). Systems of frequency curves generated by methods of translation. Biometrica 36: 149-176.

Jones, L.V., Peryam, D.R. and Thurstone, L.L. (1955). Development of a scale for measuring soldier's food preferences, Food Research, 20, 515-520.

Kelly, P.A. (1999). Overview of computer intensive statistical inference procedures. http://garnet.acns.fsu.edu/~pkelly/resampling.html.

Kleinman, J.C. (1973). Proportions with extraneous variance: single and independent samples. J. American Statistical Association 68 (341): 46-54.

Kuehl, R.O. (2000). Design of Experiments: Statistical Principles of Research Design and Analysis, 2nd Edition, Brooks / Cole, Pacific Grove, CA.

Lawless, H.T. and Malone, G.J. (1986). A comparison of rating scales: sensitivity, replicates and relative measurement. J. Sensory Studies 1: 155-174.

Lawless, H.T. (1996). Cognitive Ecology, Chapter 8 – Flavor, Academic Press, Inc., San Diego, CA.

Lawless, H.T. (1997). Commentary on random vs. fixed effects for panelists. Food Qual. and Pref. 9: 163-164.

Lundahl, D.S. and McDaniel, M.R. (1988). The panelist effect – fixed or random? J. Sensory Studies 3: 113-121.

Lundahl, D.S. (1992). Influence of special consumer groups as a subset of respondents on the outcomes of consumer acceptance tests. Product Testing with Consumers for Research Guidance: Special Consumer Groups, 2nd Vol., American Society for Testing and Materials, Philadelphia.

Lundahl, D.S. (1997). Thurstonian models – an answer to Gridgeman's Paradox? http://camo.com/Publications/WPapers/thurstonian.html

Madden, L.V. and Hughes, G. (1994). BBD-Computer software for fitting the beta-binomial distribution to disease incidence data. Plant Dis. 78: 536-540.

Mage, D.T. (1980). An explicit solution for S_B parameters using four percentile points. Technometrics, vol. 22, No. 2.

Mattes, R. and Lawless, H.T. (1985). An adjustment error in optimization of taste intensity. Appetite 6: 103-114.

MATLAB™ User's Guide (1997). The MathWorks, Inc. Natick, MA.

Morrison, D.G. (1979). Purchase intentions and purchasing behavior. J. Marketing 43: 65-74.

Moskowitz, H.R. (1983). Product Testing and Sensory Evaluation of Foods, Moskowitz / Jacobs, Inc., Scarsdale, New York and Tasc Group, Inc., Evanston, Illinois.

Næs, T. (1990). Handling individual differences between assessors in sensory profiling. Food Qual. and Pref. 2: 187-199.

Næs, T. and Langsrud, Ø. (1998). Fixed or random assessors in sensory profiling? Food Qual. and Pref. 9: 145-152.

O'Mahony, M. (1982). Some assumptions and difficulties with common statistics for sensory analysis. Food Tech. 36: 75-82.

O'Mahony, M. (1986). Sensory Evaluation of Food, Marcel Dekker, Inc., New York.

Otake, M. and Prentice, R.L. (1984). The analysis of chromosomally aberrant cells based on beta-binomial distribution. Radiation Res. 98: 456-470.

Prescott, J., Laining, D., Bell, G., Yoshida, M., Gillmore, R., Allen, S., Yamazaki, K., and Ishii, R. (1992). Hedonic responses to taste solutions: a cross-cultural study of Japanese and Australians. Chem. Senses 17: 801-809.

Prescott, J., Bell, G., Gillmore, R., Yoshida, M., O'Sullivan, M., Korac, S., Allen, S., and Yamazaki, K. (1997). Cross-cultural comparisons of Japanese and Australian responses to manipulations of sweetness in foods. Food Qual. and Pref. 8: 45-55.

Prescott, J. (1998). Comparisons of taste perceptions and preferences of Japanese and Australian consumers: Overview and implications for cross-cultural sensory research. Food Qual. and Pref. 9: 393-402.

Ramsey, R.L. and Schafer, D.W. (1997). The statistical sleuth: a course in methods of data analysis. Wadsworth Publishing Company, Belmont, CA.

Rust, R.T. and Klompmaker, J.E. (1981). Improving the estimation procedure for the beta binomial TV exposure model. J. Marketing Res. 18: 442-448.

Sabavala, D.J. and Morrison, D.G. (1977). Television show loyalty: a beta-binomial model using recall data. J. Advertising Res. 17: 35-43.

Slifker, J.F. and Shapiro, S.S. (1980). The Johnson System: selection and parameter estimation. Technometrics 22: 239-246.

Skellam, J.G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as a variable between sets of trials. J. Roy. Statist. Soc. Ser. B 10: 257-265.

Spiegel, M.R. (1975). Theory and Problems of Probability and Statistics, McGraw-Hill, Inc.

Thurstone, L. L. (1927). A law of comparative judgment. Psychological Review 34: 273-286, Wackerly, D.D., Mendenhall III, W. and Scheaffer, R.L. (1996). Mathematical Statistics with Applications, Duxbury Press, Wadsworth Publishing Company, Belmont, CA.

Wilcox, R. (1993). Applied Analysis of Variance in Behavioral Science (edited by Edwards, L.K.): Chapter 9. Robustness in ANOVA. Marcel Dekker, Inc., New York, NY.

Wilkinson, C. and Yuksel, D. (1997). Modeling differences between panelists in use of measurement scales. J. Sensory Studies 12: 55-68.

Williams, D.A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. Biometrics 31: 949-952.

Yamamoto, E. and Yanagimoto, T. (1994). Statistical methods for the β-Binomial model in teratology. Environmental Health Perspectives Supplements 102, Suppl 1: 25-31.

Yeh, L.L. and Lundahl, D.S. (1999). Use of hedonic scale among Chinese, Koreans, and Thais. M.S. Thesis. Oregon State University.

CHAPTER 5

THESIS SUMMARY

In cross-cultural studies, respondents from specific cultures may have different product preferences and scale usage, therefore combining data from different cultures may result in departure from the basic assumptions of analysis of variance (ANOVA) and fail to detect product and culture differences. The result of violations on power of ANOVA is unknown for the food researchers, but is of increasing concern with greater globalization of food markets and opportunities for exporting to different cultures. Therefore, the objectives of this research are by simulating consumer product evaluation data, to evaluate the robustness and testing power of ANOVA under different cross-cultural situations.

This study was conducted in two parts: (1) development of a simulation tool for generating sensory data and (2) applying the simulation tool to evaluate the effects of cross-cultural differences in scale usage and preference on the underlying assumptions of ANOVA.

In the first part, an empirical logit simulation model was selected as best for generating sensory data. This model can include effects such as respondent, product, consumer segment and product by segment interactions. Four underlying distributions (binomial, beta-binomial, hypergeometric, and beta-hypergeometric), and a binning model were applied independently to increase or decrease the dispersion of simulated responses. In addition, a wide range of scale and scale usage effects can be included through the use of scale location and truncation input parameters.

The results showed that the Discrete Empirical Logit was simple, relatively flexible, and capable of producing the designed cases for this simulation study.

In the second part, the Discrete Empirical Logit simulation model was chosen to simulate specified data sets under six different cross-cultural cases – central tendency, dislike avoidance, crossover, location variation, acceptance variation and segment effect. The study simulated a simplified situation with two products, three levels of product differences and two consumer segments – a reference segment and a comparison segment. For each case, 10000 data sets were randomly generated and analyzed by ANOVA using both a reduced model and a full model (including segment and segment by product interaction effects). The empirical power of ANOVA under various situations was calculated as the percentage of times of detecting a significant product difference over the 10000 data sets.

Results recommended using the ANOVA full model whenever cross-cultural data is analyzed for product preferences. The reduced model without segment and segment by product interaction effects can loss significant power to detect product differences when these differences are present. Therefore, these segment effects need to be always included and tested.

Most surprising was the demonstration that ANOVA was concluded to be very robust to cross-cultural differences in scale usage. At worst, only about 18% in power is lost under the most extreme cases evaluated in this simulation study. However, a large noise in preference within a culture can result in a significant loss in power.

As a conclusion, the study recommended using the ANOVA full model for all cross-cultural research. From the scope of this study, results from different cultures can be combined with scale usage effects assumed to have minimal impact on the power to detect differences. Further, this study demonstrated the value of simulation in sensory research to address difficult issues in data analysis.

BIBLIOGRAPHY

Amerine, M.A., Pangborn, R.M., and Roessler, E.B. (1965). Principles of Sensory Evaluation of Food. Academic Press Inc., New York.

Ayabe-Kanamura, S., Schicker, I., Laska, M., Hudson, R., Distel, H., Kobayakawa, T., and Saito, S. (1998). Differences in perception of everyday odors: a Japanese-German cross-cultural study. Chem. Senses 23: 31-38.

Bang, J.W., Schumacker, R.E., and Schlieve, P.L. (1998). Random-number generator validity in simulation studies: an investigation of normality. Educational and Psychological Measurement 58: 430.

Bertino, M. and Chan, M.M. (1986). Taste perception and diet in individuals with Chinese and European ethnic backgrounds. Chem. Senses 11: 229-241.

Bower, B. (1999). Cultures of reason: thinking styles may take Eastern and Western routes. Science News, vol.157, No. 4: 56-58.

Cardello, A.V. (1996). The role of the human senses in food acceptance, in H.L.Meisleman, & H.J.H.MacFie, Food Choice Acceptance and Consumption (pp. 1- 82), Blackie Academic & Professional, London.

Casella, G. and Berger, R.L. (1990). Statistical Inference. Duxbury Press, Wadsworth, Inc., Belmont, CA.

Chatfield, C. and Goodhardt, G.J. (1970). The beta-binomial model for consumer purchasing behavior. Appl. Statist. 19: 240-250.

Christensen, R. (1990). Log-linear Models. Springer-Verlag New York, Inc., New York, NY.

Chung, S. (1999). Cross-cultural sweetness preferences for a sports-drink. M.S. Thesis. Oregon State University.

Cooper, R.A. and Weekes, A.J. (1983). Data, Models, and Statistical Analysis. Barnes and Noble Books, Totowa, NJ.

Cox, D.N., Clark, M.R. and Mialon, V.S. (2001). A cross-cultural methodological study of the uses of two common hedonic response scales. Food Qual. and Pref. 12: 119-131.

Edwards, L.K. (1993). Applied Analysis of Variance in Behavioral Science (Chapter 1. Analysis of variance overview). Marcel Dekker, Inc., New York, NY.

Ennis, D.M. and Bi, J. (1998). The beta-binomial model: Accounting for inter-trial variation in replicated difference and preference tests. J. Sensory Studies 13: 389-412.

Ennis, D.M. and Bi, J. (1998). A thurstonian variant of the beta-binomial model for replicated difference tests. J. Sensory Studies 13: 461-466.

Everitt, B.S. (1998). The Cambridge Dictionary of Statistics. Cambridge University Press, Cambridge, U.K.

Gacula, M.C. (1993). Design and Analysis of Sensory Optimization. Food and Nutrition Press, Inc, Trumbull.

Gay, C. and Mead, R. (1992). A statistical appraisal of the problem of sensory measurement. J. Sensory Studies 7: 205-228.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (1997). Bayesian Data Analysis. Chapman & Hall/CRC, Florida.

Glass, G.V., Peckham, P.D. and Sanders, J.R. (1972). Consequences of failure to meet assumptions underlying the fixed analysis of variance and covariance. Review of Educational Research 42: 237-288.

Griffiths, D.A. (1973). Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of a disease. Biometrics 29: 637-648.

Johnson, N.L. (1949). Systems of frequency curves generated by methods of translation. Biometrica 36: 149-176.

Jones, L.V., Peryam, D.R. and Thurstone, L.L. (1955). Development of a scale for measuring soldier's food preferences, Food Research, 20, 515-520.

Kelly, P.A. (1999). Overview of computer intensive statistical inference procedures. http://garnet.acns.fsu.edu/~pkelly/resampling.html.

Kleinman, J.C. (1973). Proportions with extraneous variance: single and independent samples. J. American Statistical Association 68 (341): 46-54.

Kuehl, R.O. (2000). Design of Experiments: Statistical Principles of Research Design and Analysis, 2nd Edition, Brooks / Cole, Pacific Grove, CA.

Lawless, H.T. and Malone, G.J. (1986). A comparison of rating scales: sensitivity, replicates and relative measurement. J. Sensory Studies 1: 155-174.

Lawless, H.T. and Heymann, H. (1998). Sensory evaluation of food, NY, Chapman Hall.

Lawless, H.T. (1997). Commentary on random vs. fixed effects for panelists. Food Qual. and Pref. 9: 163-164.

Lawless, H.T. (1996). Cognitive Ecology, Chapter 8 – Flavor, Academic Press, Inc., San Diego, CA.

Lundahl, D.S. and McDaniel, M.R. (1988). The panelist effect – fixed or random? J. Sensory Studies 3: 113-121.

Lundahl, D.S. and McDaniel, M.R. (1990). Use of contrasts for the evaluation of panel inconsistency. J. Sensory Studies 5: 265-277.

Lundahl, D.S. and McDaniel, M.R. (1991). Influence of panel inconsistency on the outcome of sensory evaluations from descriptive panels. J. Sensory Studies 6: 145-157.

Lundahl, D.S. (1992). Influence of special consumer groups as a subset of respondents on the outcomes of consumer acceptance tests. Product Testing with Consumers for Research Guidance: Special Consumer Groups, 2nd Vol., American Society for Testing and Materials, Philadelphia.

Lundahl, D.S. (1997). Thurstonian models – an answer to Gridgeman's Paradox? http://camo.com/Publications/WPapers/thurstonian.html

Madden, L.V. and Hughes, G. (1994). BBD-Computer software for fitting the beta-binomial distribution to disease incidence data. Plant Dis. 78: 536-540.

Mage, D.T. (1980). An explicit solution for S_B parameters using four percentile points. Technometrics, vol. 22, No. 2.

Mattes, R. and Lawless, H.T. (1985). An adjustment error in optimization of taste intensity. Appetite 6: 103-114.

MATLAB™ User's Guide (1997). The MathWorks, Inc. Natick, MA.

Morrison, D.G. (1979). Purchase intentions and purchasing behavior. J. Marketing 43: 65-74.

Moskowitz, H.R. (1983). Product Testing and Sensory Evaluation of Foods, Moskowitz / Jacobs, Inc., Scarsdale, New York and Tasc Group, Inc., Evanston, Illinois.

Næs, T. (1990). Handling individual differences between assessors in sensory profiling. Food Qual. and Pref. 2: 187-199.

Næs, T. and Langsrud, Ø. (1998). Fixed or random assessors in sensory profiling? Food Qual. and Pref. 9: 145-152.

O'Mahony, M. (1982). Some assumptions and difficulties with common statistics for sensory analysis. Food Tech. 36: 75-82.

O'Mahony, M. (1986). Sensory Evaluation of Food, Marcel Dekker, Inc., New York.

Otake, M. and Prentice, R.L. (1984). The analysis of chromosomally aberrant cells based on beta-binomial distribution. Radiation Res. 98: 456-470.

Prescott, J., Laining, D., Bell, G., Yoshida, M., Gillmore, R., Allen, S., Yamazaki, K., and Ishii, R. (1992). Hedonic responses to taste solutions: a cross-cultural study of Japanese and Australians. Chem. Senses 17: 801-809.

Prescott, J., Bell, G., Gillmore, R., Yoshida, M., O'Sullivan, M., Korac, S., Allen, S., and Yamazaki, K. (1997). Cross-cultural comparisons of Japanese and Australian responses to manipulations of sweetness in foods. Food Qual. and Pref. 8: 45-55.

Prescott, J. (1998). Comparisons of taste perceptions and preferences of Japanese and Australian consumers: Overview and implications for cross-cultural sensory research. Food Qual. and Pref. 9: 393-402.

Ramsey, R.L. and Schafer, D.W. (1997). The statistical sleuth: a course in methods of data analysis. Wadsworth Publishing Company, Belmont, CA.

Rust, R.T. and Klompmaker, J.E. (1981). Improving the estimation procedure for the beta binomial TV exposure model. J. Marketing Res. 18: 442-448.

Sabavala, D.J. and Morrison, D.G. (1977). Television show loyalty: a beta-binomial model using recall data. J. Advertising Res. 17: 35-43.

Slifker, J.F. and Shapiro, S.S. (1980). The Johnson System: selection and parameter estimation. Technometrics 22: 239-246.

Skellam, J.G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as a variable between sets of trials. J. Roy. Statist. Soc. Ser. B 10: 257-265.

Spiegel, M.R. (1975). Theory and Problems of Probability and Statistics, McGraw-Hill, Inc.

Thurstone, L. L. (1927). A law of comparative judgment. Psychological Review 34: 273-286,

Wackerly, D.D., Mendenhall III, W. and Scheaffer, R.L. (1996). Mathematical Statistics with Applications, Duxbury Press, Wadsworth Publishing Company, Belmont, CA.

Wilcox, R. (1993). Applied Analysis of Variance in Behavioral Science (edited by Edwards, L.K.): Chapter 9. Robustness in ANOVA. Marcel Dekker, Inc., New York, NY.

Wilkinson, C. and Yuksel, D. (1997). Modeling differences between panelists in use of measurement scales. J. Sensory Studies 12: 55-68.

Williams, D.A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. Biometrics 31: 949-952.

Yamamoto, E. and Yanagimoto, T. (1994). Statistical methods for the β-Binomial model in teratology. Environmental Health Perspectives Supplements 102, Suppl 1: 25-31.

Yeh, L.L. and Lundahl, D.S. (1999). Use of hedonic scale among Chinese, Koreans, and Thais. M.S. Thesis. Oregon State University.