Sabuj Chandra Bhowmick

A Conjoint Analysis Applied to Consumers Choice of Tablets.

Master's thesis

NTNU
Norwegian University of
Science and Technology

A Conjoint Analysis Applied to Consumers Choice of Tablets

Sabuj Chandra Bhowmick

Trondheim, June 2019

NTNU
Norwegian University of
Science and Technology

Contents

1 Declaration

I declare that this Master thesis is written independently under the supervision of John Sølve Tyssedal and accordance with mathematical sciences department at the Norwegian University of Science ans Technology.

Yours sincerely,

Sabuj Chandra Bhowmick

2 Acknowledgements

My gratitude to Almighty God, the most Merciful and Beneficent, knows no bound who gave me the opportunity to pursue the degree offered by esteemed Norwegian University of Science and Technology (NTNU) and foremost thank goes to Department of Mathematical Sciences, Norwegian University of Science and Technology (NTNU) for providing me a platform for my study.

I would like to express my heartiest gratitude to my supervisor John Sølve Tyssedal for his wholehearted support and guidance throughout the entire Master's thesis and introducing me to the potential research site. I would like to appreciate his immense knowledge, enthusiasm, and patience. The thesis would not have been possible without his cooperation and encouragement.

Finally, I would like to express my gratitude to my family members and friends who have always been constant source of inspiration and support for the completion of this study and having faith on my abilities.

3 Abstract

Conjoint analysis is increasingly used in the applied consumers choice of products and services. This method is widely applied in many market research projects. In this thesis, we used the conjoint analysis applied to consumers choice of tablets as a case study.

We wanted to know what are the main factors of choosing a tablet for buying. Design of experiment is used in this thesis as a principal tool for finding the features of a tablet. A 2^{5-1} fractional factorial design with two levels is the main study design used in this thesis. R programming language is used for the analysis in this thesis,using both self coding and built-in functions in R.

Three different responses are used in this thesis, which are mainly mean, variance and logarithm of variance. We compare all three models with different responses and choose the best alternative among them.

4 Introduction

Choice design or conjoint analysis is a process of studying consumers' choice process and determining the best option for the consumers. This method was first introduced in 1971 for market research tools[3]. After the introduction of this method, it has been widely applied in many market research projects. It is must as a help in marketing decisions for taking the best possible option for the design of new products, and so on.

These are the far reaching purpose of this method. It also enables us to solve many real life and hypothetical problems in a broad statistical way.

4.1 Thesis Outline

This thesis is designed so that the readers will have the necessary theoretical background before further discussion. Chapter 2 gives the theoretical overview of full and fractional factorial design at various levels. Chapter 2 also delivers the other theories of linear regression model, weighted linear regression model, logistic regression model, and principal component analysis. The proposed methods and experiments are described in the chapter 3. Results of the experiments are presented in the chapter 4. Finally, conclusion is presented in the chapter 5.

5 Background

In modern competitive market it is very important to know the product and service attributes. The attributes are the features of the products or services. Once the the key attributes are determined then it is easier to develop product that keeps the company ahead of others. Conjoint analysis in this case is a very important tool for determining the key attributes of a desired product and service in the early phase of product development. In conjoint analysis attributes are mainly the key features of a product. Key features of a mobile phone will be touch screen, music support, digital camera, internal memory, long battery life, internet connectivity and many more. However,it is possible to include more attributes but not to increase the respondent fatigue and economic consideration [4].

The most important thing in the conjoint analysis is finding the key features for the product. To know the feasible and most important features of a product respondents can give their judgment on a product profile. To do this the experimenter can adopt three different types of strategies[4].

S1:Build a physical prototype and let them respondent interact with them.

S2:Build a virtual prototype and let them interact in the virtual environment.
S3:Use the already existing product in the market.

Developing physical prototype at the beginning of the business is costly and if the product does not meet the market demand the whole things will be waste of money. Virtual products describe only a virtual reality that include some of the characteristic of the physical product but not the real product. It is a good idea of saving time and resources. However, these two situation is not feasible at the early stage of product development. So, the use of profile that is already existing in the market could save resource and time and money.

5.1 Principal methods of Conjoint Analysis

Literature suggests that there exist four main types of conjoint analysis: Traditional method or full profiles, Choice-based conjoint analysis, adaptive conjoint analysis, and self-explained conjoint analysis[5]. Many other writers stated these methods as full profile, paired comparison,trade-off matrices and combination of these methods.

- Full profile method

- Choice based method

- paired comparison method

- Trade off method

All four methods have their own inheritance merits and demerits. Based on application complexity and cost consideration, this thesis focused only on Full profile method as the primary method of conjoint analysis. The detailed discussion of Full profile method is described in the next section.

5.2 Full profile method

Traditional conjoint analysis consider the preferences or judgments for profiles of any hypothetical products where profiles are entire set of attributes denoted as full profiles [3]. In conjoint analysis full profile methods is the widely applicable method in the real world [6]. In conjoint analysis, attributes are the characteristics, consumers consider when evaluating products or services.For example, consumers shopping for mobile devices will consider: operating system, ram, storage , battery size, screen size etc. These attributes can be presented in

different degree called the attribute levels. For example, the attribute of screen size in different levels, 4 inch, 4.5 inch, 5 inch, 6 inch etc.

5.2.1 Defining Profiles

For any hypothetical product in conjoint analysis sets of attributes are used to describe the product. These are called the profiles or full profiles[3]. In conjoint study selection of attributes and levels are the mains steps to design hypothetical profiles.

5.2.2 Types of Attributes

In conjoint analysis we evaluate a product or service with it's characteristics and those characteristics are knows as attributes. The attributes can be divided into two categories namely, categorical or quantitative. Generally, nominal or verbal description is used in categorical attributes and quantitative attribute is measured in the ratio or interval scale. Levels of categorical attributes are re-coded into dummy variables and in quantitative attribute numbers describes the levels.

5.2.3 Part worth Functions

In conjoint analysis individual's overall preference of a set of choice alternatives is decomposed into separate and compatible utility values corresponding to each attribute[5]. These functions are named part worth functions. They can be linear or non-linear mainly determined by the researcher. A linear function is also called a vector model for which utility increases or decreases with the numerical value of the attributes while a non-linear function is called a point model. The point model can be either positive or negative.

5.2.4 Statistical design for Full Profile methods

Full factorial designs, fractional factorial designs, orthogonal arrays, and incomplete block designs are the main statistical designs used in full profile conjoint analysis.

The full profile method has many advantages in the real life applications.

- Information can be easily achievable with minimal effort.

- Interactions between different attributes can be observed and measured.

- It is also possible to display all the attributes at the same time.

There are some disadvantages in full profile methods such as,

- They provide much information at the same time and that makes the task more complicated to handle.

5.3   Choice Based Method

In the recent years choice based conjoint analysis has also gained much popularity side by side with the full profile method in the market research community[1].In choice based conjoint studies an individual makes a choice among a set of choice alternative, each of which is usually described by a set of attributes. Many choice sets are delivered to each individual[5].Then these choice data are analyzed using a choice model. This procedure is known as choice based conjoint analysis. The method has an alternative name, which is stated as choice methods.There are many advantages of choice based conjoint analysis.

- It considers simulated purchased decision or choices, which is comparatively more easy and realistic than rating or ranking based method.

- Statistical methods can be used for analyzing choice conjoint data at the gross level or at the contingent level.

- It has the capability to value brand based attributes.

- It has the capability to measure competitive effects on choice.

- It also has the capability to enumerate price sensitivity to price differences.

- It is very flexible to use the estimated model to forecast the real marketplace choice.

The main disadvantage of this study is that the design of the choice based conjoint study if very complex.

5.4   Experimental design

In the scientific and engineering world experimental design is an important statistical tool for improving the product development process. The most use of experimental design is in develop of new products and development of processes. Experimental design helps this process in efficient and effective way by choosing the best product features and minimizing the product manufacturing cost.

The design of experiment in this thesis will be the principal tool for finding the best product feature of future product development. In the early phase of product development use of experimental design could lead to improved process yield, reduced overall cost, reduced development time, reduce the variability and bring to the ideal requirements[2]. The design of experiment has extensive use in engineering, marketing, market research, agricultural research and other business operations.

6 Theory

6.1 Factorial Design

In general, many real life experiments involve the study of factors with two or more levels. Usually, two or more than two factors are used in such experiments. The efficient way of studying many factors are factorial design with two levels [2]. For simplicity, we consider two factors A and B each having two levels "low" and "high", which are usually denoted by "-"and "+" respectively. In the factorial design variables are called factors. In general, k factors with 2 levels give 2^k possible combinations of experiments.

Table 1: standard 2^2 experiment with notation

Experiment no.	A	B	AB	level code	Response
1	-1	-1	1	1	y_1
2	1	-1	-1	a	y_2
3	-1	1	-1	b	y_3
4	1	1	1	ab	y_4

The effect of A is called the main effect of A, which is defined by the average response when A is on its high level minus the mean response where A is on its low level. Similarly true for the main effect of B. These two factor may also have an interaction between them and this is termed the interaction effect of AB. We can also illustrate the concept of main effect and interaction effect in the regression model representation.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon \tag{1}$$

Here, y is the response , β's are the parameters whose values are unknown and are to be determined.Where ϵ is multi-normally distributed with mean 0 and a diagonal co-variance matrix $N(0, \sigma^2 I)$. x_1 is the variable that represents factor A and x_2 is the variable that represents factor B. x_1 x_2 represents the interaction between A and B. In general regression model can be expressed as

$$Y = X\beta + \epsilon \tag{2}$$

When the column of the design matrix $1, x_1, x_2, ..., x_k$, are orthogonal, then the coefficients in the vector form are given by:

$$\hat{\beta} = (X^\mathsf{T} X)^{-1} X^\mathsf{T} Y \tag{3}$$

Where,

$$\{X^\mathsf{T} X\}_{jk} = \sum_{i=1}^{n} (X_{ij} X_{ik}) = \begin{cases} 0 & j \neq k \\ n & j = k \end{cases}$$

$$\hat{\beta} = \begin{bmatrix} \frac{1}{n} & 0 & . & 0 \\ 0 & \frac{1}{n} & . & 0 \\ . & . & . & . \\ 0 & 0 & . & \frac{1}{n} \end{bmatrix} \begin{bmatrix} X \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ . \\ y_n \end{bmatrix} = \frac{1}{n} \sum_{i=1}^{n} X_{ij} y_i \tag{5}$$

$\hat{\beta}$ is also multi-normally distributed. Since the intercept is included in X, $\hat{\beta}_0 = \bar{y}$,the average of all y_i

6.1.1 Full factorial designs at two levels:2^k design

The 2^k full factorial design consists of k factors with two levels. The combination of factor levels are known as treatment or runs. So,a 2^k factorial design consists of 2^k runs. We consider a three factors two level full factorial design which is termed as 2^3 factorial design. The three factors in this case are A,B and C with two levels "-" as low and "+" as high level. the corresponding design matrix is displayed in the table 2.

Table 2: Design matrix for 2^3 factorial design

A	B	C	AB	AC	BC	ABC	level code	Response
-	-	-	+	+	+	-	1	y_1
+	-	-	-	-	+	+	a	y_2
-	+	-	-	+	-	+	b	y_3
+	+	-	+	-	-	-	ab	y_4
-	-	+	+	-	-	+	c	y_5
+	-	+	-	+	-	-	ac	y_6
-	+	+	-	-	+	-	bc	y_7
+	+	+	+	+	+	+	abc	y_8

The estimated main and interaction effect become: Main effect of A is

$$\hat{A} = \frac{y_2 + y_4 + y_6 + y_8}{4} - \frac{y_1 + y_3 + y_5 + y_7}{4} \tag{6}$$

Main effect of B is

$$\hat{B} = \frac{y_3 + y_4 + y_7 + y_8}{4} - \frac{y_1 + y_2 + y_5 + y_6}{4} \tag{7}$$

Main effect of C is

$$\hat{C} = \frac{y_5 + y_6 + y_7 + y_8}{4} - \frac{y_1 + y_2 + y_3 + y_4}{4} \tag{8}$$

Interaction effect of AB is

$$\hat{AB} = \frac{y_1 + y_4 + y_5 + y_8}{4} - \frac{y_2 + y_3 + y_6 + y_7}{4} \tag{9}$$

Interaction effect of AC is

$$\hat{AC} = \frac{y_1 + y_3 + y_6 + y_8}{4} - \frac{y_2 + y_4 + y_5 + y_7}{4} \tag{10}$$

Interaction effect of BC is

$$\hat{BC} = \frac{y_1 + y_2 + y_7 + y_8}{4} - \frac{y_3 + y_4 + y_5 + y_6}{4} \tag{11}$$

Interaction effect of ABC is

$$\hat{ABC} = \frac{y_2 + y_3 + y_5 + y_8}{4} - \frac{y_1 + y_4 + y_6 + y_7}{4} \tag{12}$$

6.1.2 Full Factorial Designs at three levels: 3^k

In many situation, there are more than two levels and other design have two level factorial design can be used. A 3 level design with 3 factors in 27 runs is given in table 3. This design is the representation of a multi-way layout experiment. If we denote the three levels 0,1,2 the design matrix in the three factors A,B and C becomes:

Table 3: Design matrix and data for a 3^3 design

A	B	C	AB	AB^2	AC	AC^2	BC	BC^2	ABC	ABC^2	AB^2C	AB^2C^2
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	1	2	1	2	1	2	1	2
0	0	2	0	0	2	1	2	1	2	1	2	1
0	1	0	1	2	0	0	1	1	1	1	2	2
0	1	1	1	2	1	2	2	0	2	0	0	1
0	1	2	1	2	2	1	0	2	0	2	1	0
0	2	0	2	1	0	0	2	2	2	2	1	1
0	2	1	2	1	1	2	0	1	0	1	2	0
0	2	2	2	1	2	1	1	0	1	0	0	2
1	0	0	1	1	1	1	0	0	1	1	1	1
1	0	1	1	1	2	0	1	2	2	0	2	0
1	0	2	1	1	0	2	2	1	0	2	0	2
1	1	0	2	0	1	1	1	1	2	2	0	0
1	1	1	2	0	2	0	2	0	0	1	1	2
1	1	2	2	0	0	2	0	2	1	0	2	1
1	2	0	0	2	1	1	2	2	0	0	2	2
1	2	1	0	2	2	0	0	1	1	2	0	1
1	2	2	0	2	0	2	1	0	2	1	1	0
2	0	0	2	2	2	2	0	0	2	2	2	2
2	0	1	2	2	0	1	1	2	0	1	0	1
2	0	2	2	2	1	0	2	1	1	0	1	0
2	1	0	0	1	2	2	1	1	0	0	1	1
2	1	1	0	1	0	1	2	0	1	2	2	0
2	1	2	0	1	1	0	0	2	2	1	0	2
2	2	0	1	0	2	2	2	2	1	1	0	0
2	2	1	1	0	0	1	0	1	2	0	1	2
2	2	2	1	0	1	0	1	0	0	2	2	1

6.1.3 Fractional Factorial design

Generator

In the 2^{3-1} experiments column sign of AB and column sign of C are same i.e.
C=AB This relation is called the generator of the design. C=AB, C^2=I=ABC
that means I=ABC is the defining relation of the experiment.

Resolution

Fractional factorial designs may have the ability to separate the main effects from the lower order interactions. This ability is expressed through called the resolution of the design. Design is of resolution R a p factor effect is aliased with an effect containing less than $R - p$ factors. So, main effects are aliased with $R - 1$ factor interaction, and two-factor interactions are aliased with $R - 2$ factor interactions.

Resolution III

In this case main effects are aliased with the two factor interactions. Such an experiment is 2^{3-1}, I=ABC.

Resolution IV

In this case main effects are aliased with the three factor interactions and two factor interactions are aliased with the two factor interactions. Such an experiment is 2^{4-1}, I=ABCD.

Resolution V

In this case main effects are aliased with the four factor interactions and two factor interactions are aliased with the three factor interactions. Such an experiment is 2^{5-1}, I=ABCDE.

The full 2^5 factorial design has 32 runs, which takes a lot of resources and time. In choice experiments we often need fewer runs which requires fewer resources and less time, while still gaining the important information. In this regard we may consider a half fraction of a 2^5 design, a 2^{5-1} design which has 16 runs. In a 2^{5-1} experiment $I = ABCDE$ is the defining relation of the experiment. In 2^{5-1} experiment main effects are aliased with a four-factor interactions (e.g) $[A] \rightarrow A + BCDE$ and two-factor interactions are aliased with a three-factor interaction (e.g) $[AB] \rightarrow AB + CDE$. Thus the design is of resolution V.Multiplying the defining relation by the effects gives us which effects that are aliased.

```
A=A( I )=A(ABCDE)=BCDE
B=B( I )=B(ABCDE)=ACDE
C=C( I )=C(ABCDE)=ABDE
D=D( I )=D(ABCDE)=ABCE
E=E( I )=E(ABCDE)=ABCD
AB=CDE
AC=BDE
AD=SCE
BC=ADE
BD=ACE
CD=ABE
ABC=DE
ABD=CE
BCD=AE
ACD=BE
```

Table 4: 2^{5-1} fractional factorial design

Run	A	B	C	D	E
1	-1	-1	-1	-1	1
2	1	-1	-1	-1	-1
3	-1	1	-1	-1	-1
4	1	1	-1	-1	1
5	-1	-1	1	-1	-1
6	1	-1	1	-1	1
7	-1	1	1	-1	1
8	1	1	1	-1	-1
9	-1	-1	-1	1	-1
10	1	-1	-1	1	1
11	-1	1	-1	1	1
12	1	1	-1	1	-1
13	-1	-1	1	1	1
14	1	-1	1	1	-1
15	-1	1	1	1	-1
16	1	1	1	1	1

6.2 Weighted Linear Regression Model

In classical linear regression model we have the following model.

$$Y_{n \times 1} = X_{n \times p} \beta_{p \times 1} + \epsilon_{n \times 1} \tag{13}$$

Where,

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & . & . & x_{1p} \\ 1 & x_{21} & x_{22} & . & . & x_{2p} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & x_{n1} & x_{n2} & . & . & x_{np} \end{pmatrix} \times \begin{pmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_0 \\ \epsilon_1 \\ . \\ . \\ \epsilon_n \end{pmatrix} \tag{14}$$

x_{ij} is the value of $(i = 1, 2, ..., n)$ ith respondent in the jth attribute($j = 1, 2, ..., p$). For some cases we use known multiplicative coefficient in the design matrix X. With the multiplicative coefficient the design matrix becomes,

$$X_{new} = \begin{pmatrix} 1 & \alpha_{11}x_{11} & \alpha_{12}x_{12} & . & . & \alpha_{1p}x_{1p} \\ 1 & \alpha_{21}x_{21} & \alpha_{22}x_{22} & . & . & \alpha_{2p}x_{2p} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & \alpha_{n1}x_{n1} & \alpha_{n2}x_{n2} & . & . & \alpha_{np}x_{np} \end{pmatrix} \tag{15}$$

Then the model with multiplicative factor becomes,

$$Y = X_{new}\beta_{new} + \epsilon \tag{16}$$

The new multiplicative coefficients β_{new} can be related to β where, $_{ij}$ is different for each attribute and same for all respondent,

$$\beta_{new} = \begin{bmatrix} \beta_0, & \frac{\beta_1}{\alpha_1}, & ...., & \frac{\beta_p}{\alpha_p} \end{bmatrix}^T \tag{17}$$

Here we introduce the relative weights in the model by considering,

$$\alpha_j = \frac{1}{w_j} \tag{18}$$

Then the new coefficient becomes,

$$\beta_{new} = w\beta \tag{19}$$

6.3 Weighted ordinal Logistic Regression

In rank or rating based conjoint analysis when data are obtained in a Likert scale then ordinal logistic regression is more suitable for modeling. In general we are considering sample of binary data $y_1, y_2, \ldots, y_n$ from the population. Here, $y_i=1$, if subject i has a certain property otherwise, $y_i=0$. In general we consider,

$$P(x) = E(y/x) = P(y = 1/x) \tag{20}$$

Here we want to model a relationship between $p(x)$ and x, which can be expressed in terms of logistic regression model, which is as follows

$$P(x) = \frac{exp(\beta_0 + \beta_1 x)}{1 + exp(\beta_0 + \beta_1 x)} \tag{21}$$

This model can also be expressed as odds,

$$\frac{P(x)}{1 - P(x)} = exp(\beta_0 + \beta_1 x) \tag{22}$$

or,

$$log\left(\frac{P(x)}{1 - P(x)}\right) = \beta_0 + \beta_1 x \tag{23}$$

Which is called logit model.

6.4 Principal Component Analysis

The principal component analysis (PCA) is based on the variance covariance structure of a set of variables by a small number of linear combinations of these variables. The main purposes of PCA are (i) Data reduction and $ii)$ interpretation.

We can understand how the variables contribute to the variation in the data through PCA. For the total system of variability we need P components, in most of the cases much of these variability can be explained by the few K principal components. This method is cost effective and less time consuming.

6.4.1 Population Principal Components

Principal components can be expressed as linear combination of the P random variables $X_1,X_2,......,X_P$. As we will show principal components mainly depends on the covariance matrix Σ or correlation matrix ρ of $X_1,X_2,......,X_P$. Let the random vector,

$$X' = \begin{bmatrix} X_1, & X_2, & X_3,....., X_p \end{bmatrix} \tag{24}$$

have the covariance matrix Σ with their eigenvalues $\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_p$. we consider the following linear combinations,

$$Y_1 = a_1{}'X = a_{11}X_1 + a_{12}X_2 + ... + a_{1p}X_p \tag{25}$$

$$Y_2 = a_2{}'X = a_{21}X_1 + a_{22}X_2 + ... + a_{2p}X_p \tag{26}$$

$$Y_p = a_p{}'X = a_{p1}X_1 + a_{p2}X_2 + ... + a_{pp}X_p \tag{27}$$

We know from the linear algebra for any linear combination of $Z = CX$ have mean $\mu_z = E(Z) = E(CX) = C\mu_x$ and covariance matrix $\Sigma_Z = Cov(Z) = Cov(CX) = C\Sigma_x C'$ By applying the above theorem we have,

$$Var(Y_i) = a_i{}'\Sigma a_i \quad i = 1, 2, ..., p \tag{28}$$

$$Cov(Y_i, Y_k) = a_i{}'\Sigma a_k \quad i = 1, 2, ..., p \tag{29}$$

The principal component here are the uncorrelated linear combinations $Y_1,Y_2,Y_3.....,Y_P$. The first principal component is the linear combination with maximum variance. That is $Var(Y_1) = a_1{}'\Sigma a_1$, linear combination $a_1{}'X$ that maximizes $Var(a_1{}'X)$ subject to $a_1{}'a_1 = 1$
The second principal component is,linear combination $a_2{}'X$ that maximizes $Var(a_2{}'X)$ subject to $a_2{}'a_2 = 1$ and $cov(a_1{}'X, a_2{}'X) = 0$.

7 Methods

7.1 General Procedure

7.2 Method overview

The proposed solution for investigating the preference of attributes in this thesis is the fractional factorial design. The fractional factorial design is already described in the theory section. This thesis consider a real life case study of investigating the preference of desired tablet (Ipad, Notepad, Papertab) selection to buy. Ultimately, five product attributes were selected, and for each attribute two levels were considered.

Table 5: Description of the chosen product attributes and their relative levels

Attribute	Description	Low level(0)	High level(1)
A	Price	2000-3000(Kr)	3000-5000(Kr)
B	Multitasking	No	Yes
C	Battery Life	10 hours	>10hours
D	Memory	≤64 Gb(small)	>64 Gb(Big)
E	Screen Size	≤10 inch(small)	>10 inch(Big)

In this case study, total of 23 university students preference were investigated. A full factorial design consists of 32 product profiles which is expensive and time consuming for conjoint analysis experimentation. To solve this problem a fractional factorial design 2^{5-1} with resolution V was consider consisting of 16 profiles. The experimental design is presented in the following table.

Table 6: Experimental design for the evaluation

Run	Price(Kr.)	Multitasking	Battery Life(Hours)	Memory(Gb)	Size(Inch)	Score(1 to 10)
1	2000-3000	No	≤10	≤64	>10	
2	3000-5000	No	≤10	≤64	≤10	
3	2000-3000	Yes	≤10	≤64	≤10	
4	3000-5000	Yes	≤10	≤64	>10	
5	2000-3000	No	>10	≤64	≤10	
6	3000-5000	No	>10	≤64	>10	
7	2000-3000	Yes	>10	≤64	>10	
8	3000-5000	Yes	>10	≤64	≤10	
9	2000-3000	No	≤10	>64	≤10	
10	3000-5000	No	≤10	>64	>10	
11	2000-3000	Yes	≤10	>64	>10	
12	3000-5000	Yes	≤10	>64	≤10	
13	2000-3000	No	>10	>64	>10	
14	3000-5000	No	>10	>64	≤10	
15	2000-3000	Yes	>10	>64	≤10	
16	3000-5000	Yes	>10	>64	>10	

All respondent were given this 16 profile choices and asked them to score the profile between 1 to 10. Higher the score is the most likely to prefer or buy and lower the score is less likely to prefer or buy.

7.3 The relation Model

To find the relationship between respondent preference and desired product attributes, data will be analyzed through fractional factorial design and ordinary regression model. The output for the response variable 'preference score' is presented in the following table. Each score is associate with the corresponding profile preference. For, instance score one is associated with the first profile. The score is ranges between 1 to 10.

Table 7: Respondent's evaluation for profile preference

Respondent	Score 1	Score 2	Score 3	Score 4	Score 5	Score 6	Score 7	Score 8
1	2	2	3	2	6	3	7	8
2	2	2	5	7	2	3	6	8
3	2	2	2	2	4	3	7	4
4	6	6	7	7	6	6	7	7
5	3	7	4	4	5	7	4	4
6	5	1	1	1	1	2	5	3
7	1	2	8	6	5	4	7	7
8	2	5	2	6	8	4	2	3
9	6	6	7	8	4	5	7	8
10	1	1	2	2	1	1	5	4
11	1	1	1	1	1	1	10	1
12	3	2	6	4	5	2	3	8
13	5	9	7	6	7	7	5	6
14	5	3	7	6	6	5	6	7
15	3	3	6	6	5	4	6	6
16	1	1	2	1	3	4	7	6
17	1	1	8	9	2	1	10	3
18	1	1	4	1	2	1	5	1
19	10	6	6	7	8	8	7	7
20	4	5	5	4	4	7	8	8
21	2	2	5	5	3	3	6	6
22	3	3	5	6	4	6	7	8
23	1	1	4	4	3	3	8.5	8.5

Table 8: Respondent's evaluation for profile preference

Respondent	Score 9	Score 10	Score 11	Score 12	Score 13	Score 14	Score 15	Score 16
1	5	4	5	6	2	4	10	9
2	4	5	6	6	5	4	9	7
3	2	2	3	3	3	4	7	9
4	6	6	7	7	6	6	9	7
5	4	3	4	3	5	7	8	3
6	4	4	8	7	5	2	7	10
7	2	3	8	9	6	5	8	9
8	7	6	8	1	5	4	3	1
9	5	5	7	8	3	5	9	8
10	1	1	2	5	1	1	10	8
11	1	1	9	8	1	1	10	10
12	2	6	5	6	7	5	9	5
13	2	6	5	6	7	5	9	5
14	5	6	6	6	8	6	9	7
15	5	7	7	6	8	7	9	8
16	5	5	8	7	9	8	10	10
17	2	1	7	6	4	1	5	2
18	1	1	3	1	1	1	10	5
19	7	8	6	8	9	8	9	7
20	5	4	4	4	5	4	8	7
21	3	4	7	6	4	4	7	6.5
22	3	8	6	4	5	5	9	10
23	8	8	9	9	9	9	10	10

7.4   Analysis tools and packages

R programming language is used in this thesis paper for analysis and computation purpose. Various built-in R packages used in this thesis paper. R package $FrF2$ is part of those packages. This package is good for fractional factorial design with two level factors. This package gives the all necessary analysis of two level factor with main effect, interaction effect along with the plots and summary statistics.

7.5   Experiment

In this thesis we used the individual preference of buying the desired tablet (IPad, Notepad, Paper tab) based on the choice attributes. Preference score is considered as the response variable in this experiment. This thesis consider the three responses in three different measurements. In the first response mean preference score is considered as the response variable, where as variance is considered as the second response variable and finally logarithm of variance is considered as the response variable. To compare between the response we use the two sample t test for difference in the means for each pair groups. Where the test statistic is

$$t = \frac{\bar{X}_{16} - \bar{X}_{15}}{s\sqrt{\frac{1}{n_{16}} + \frac{1}{n_{15}}}} \tag{30}$$

t distributed with $n_1 + n_2 - 2$ degrees of freedom. The pooled estimate of the variance is given by

$$s^2 = \frac{(n_{16} - 1)s_1^2 + (n_{15} - 1)s_2^2}{n_{16} + n_{15} - 2} \tag{31}$$

Provided to that the level combinations 15 and 16 give the highest score. Level combination 16 has a high variance level combination 15 has the lowest and seems preferable. The confidence interval is (-0.111,0.429) with p-value 0.050697 for the expected response at the level combination 15. Test of the expected response at 15 is larger than level combination 16.

Table 9: Set up for the 2^{5-1} fractional factorial design with 16 runs and three different responses.

Run	A	B	C	D	E	Y_1(mean score)	Y_2(variance)	Y_3(Log of Variance)
1	-1	-1	-1	-1	1	3.04	0.223	-0.651
2	1	-1	-1	-1	-1	3.13	0.238	-0.623
3	-1	1	-1	-1	-1	4.65	0.216	-0.666
4	1	1	-1	-1	1	4.57	0.264	-0.578
5	-1	-1	1	-1	-1	4.13	0.195	-0.710
6	1	-1	1	-1	1	3.91	0.201	-0.696
7	-1	1	1	-1	1	6.33	0.159	-0.796
8	1	1	1	-1	-1	5.72	0.239	-0.619
9	-1	-1	-1	1	-1	4.04	0.184	-0.736
10	1	-1	-1	1	1	4.52	0.228	-0.641
11	-1	1	-1	1	1	6.17	0.165	-0.783
12	1	1	-1	1	-1	5.74	0.214	-0.668
13	-1	-1	1	1	1	5.09	0.276	-0.558
14	1	-1	1	1	-1	4.74	0.258	-0.588
15	-1	1	1	1	-1	8.39	0.129	-0.888
16	1	1	1	1	1	7.15	0.288	-0.541

8   Results of the experiment

The experiment involves five factors (A) Price(low,high),(B) Multitasking(no,yes), (C) Battery Life(10 hours,>10 hours),(D) Memory(small,big), and (E) Screen Size(small,big). This experiment based on 2^{5-1} design with 16 runs and defining relation is $I = ABCDE$. In this design main effect is aliased four-factor interaction and two factor interaction is aliased with three factor interaction.

The result of the experiment with response mean score is presented in the table 10.

Table 10: Effects, Regression Coefficients and Sum of Squares for the experiment with response mean score.

Variable	Regression Coefficient	Estimated Effect	Sum of Squares
Overall Average	5.0825		
A	-0.1475	-0.29	0.348
B	1.0075	2.01	16.241
C	0.600	1.20	5.760
D	0.6475	1.30	6.708
E	0.0150	0.30	0.004
AB	-0.1475	-0.30	0.348
AC	-0.155	-0.31	0.384
AD	-0.0450	-0.09	0.032
AE	0.0875	0.18	0.123
BC	0.207	0.41	0.689
BD	0.1250	0.25	0.250
BE	-0.055	-0.10	0.040
CD	0.0125	0.02	0.002
CE	-0.077	-0.15	0.096
DE	-0.0125	-0.02	0.002

The table 10 contains effect estimates, regression coefficients, sum of squares form the 15 effects from this experiment. The figure 1 shows the normal probability plot of the effects. Figure 2 shows the normal probability plot of the residuals and plot of residual versus predicted.
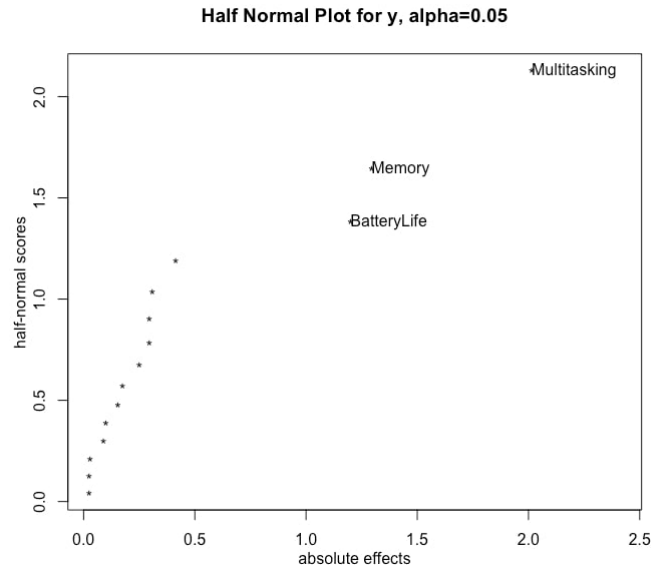
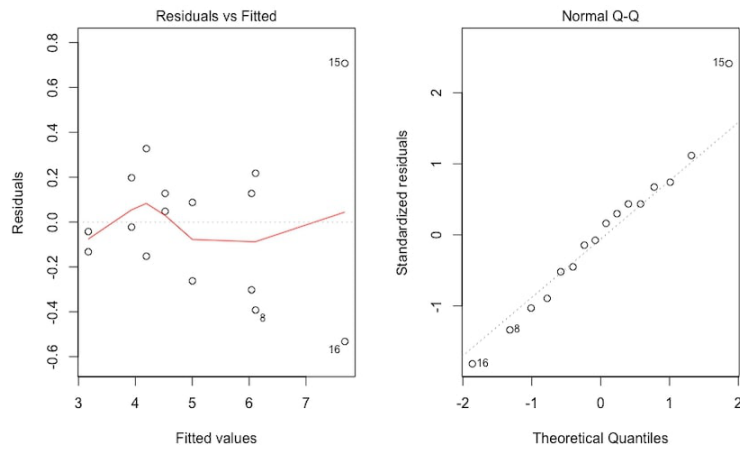Figure 1: Normal Probability Plot of effects

Figure 2: Plot of residuals versus predicted and Normal probability plot of the residuals

E and A has very low effect on the mean score. Main effects of B, C, and D are important effect on the mean score. These main effect also aliased with the four factor interaction, these effects are actually B+ACDE, C+ABDE, and D+ABCE. Since, it is clear from the result that three factor and higher interactions are negligible, we see only B, C, and D are the important effects.

Table 11: Analysis of Variance for the reduced model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F value	P-value
B	16.214	1	16.241	84.04	<0.0001
C	5.760	1	5.760	29.80	<0.0001
D	6.708	1	6.708	34.71	<0.0001
Residuals	2.319	12	0.193		
Total	28.6828	15			

The table 11 shows the reduced model with the main effects of Multitasking, Memory and Battery life. Table 11 summaries the analysis of the variance for this experiment. The model sum of squares is $SS_{model} = SS_B + SS_C + SS_D = 28.6828$, which accounts for the most of the variability in the mean score. Figure 2 shows the Plot of residuals versus predicted values and Normal probability plot of the residuals. Both plots are satisfactory.

Table 12: Regression output for the reduced model

Coefficients	Estimates	Standard Error	t-value	P-value
Intercept	5.0825	0.1099	46.245	6.82e-15
Multitasking	1.0075	0.1099	9.167	9.09e-07
Battery Life	0.6000	0.1099	5.459	0.000145
Memory	0.6475	0.1099	5.892	7.35e-05

Table 12 shows the regression output with coefficients, estimates, t-value and p-value for the reduced model. The effect on the score increases by 1.0075 units for multitasking from going from -1 to 1. For Battery life it is increased by 0.60 units where as for Memory it increases by 0.64 units form going from -1 to 1.

Table 13: Effects, Regression Coefficients and Sum of Squares for the experiment where response is the variance.

Variable	Regression Coefficient	Estimated Effect	Sum of Squares
Overall Average	0.2175		
A	0.0243	0.05	0.009251
B	-0.008	-0.02	0.001029
C	0.0008	0.00	0.000013
D	0.00034	0.00	0.000002
E	0.008	0.02	0.001085
AB	0.018	0.04	0.005214
AC	0.004	0.01	0.000293
AD	0.005	0.01	0.000448
AE	-0.004	-0.01	0.000292
BC	-0.006	-0.01	0.000601
BD	-0.0107	-0.02	0.001852
BE	0.00138	0.00	0.000031
CD	0.0191	0.04	0.005850
CE	0.004	0.01	0.000349
DE	0.013	0.03	0.002842

The above table contains effect estimates, regression coefficients, sum of squares form the 15 effects from this experiment. The figure 3 shows the normal probability plot of the effects. Figure 4 shows the normal probability plot of the residuals and plot of residual versus predicted.
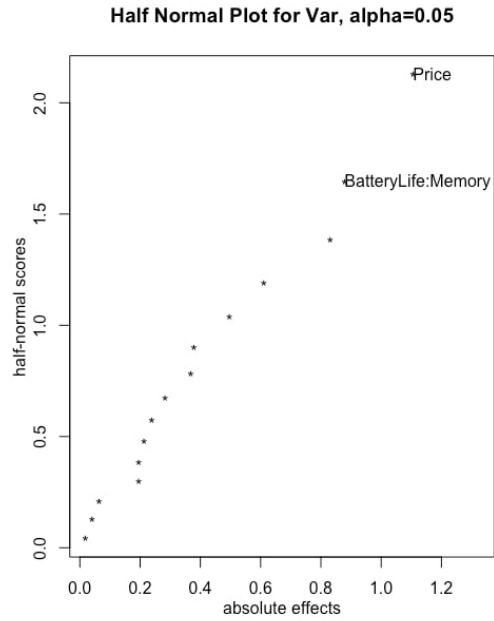
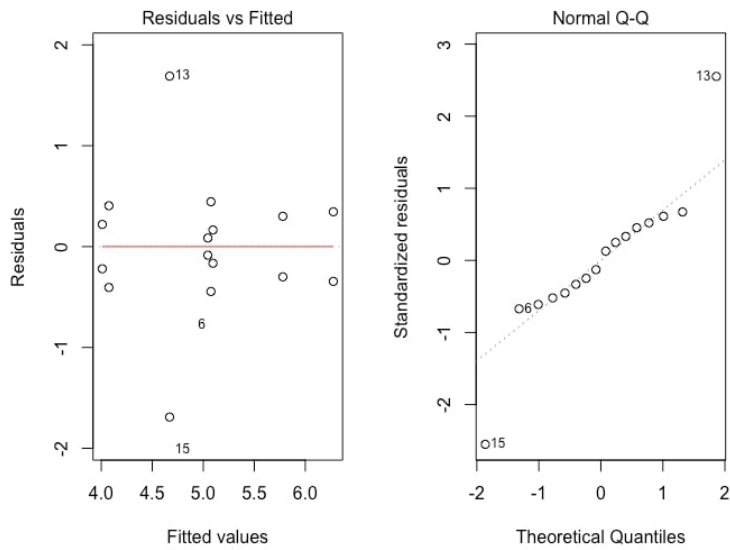Figure 3: Normal Probability Plot of effects

Figure 4: Plot of residuals versus predicted and Normal probability plot of the residuals

It is possible to reduce the variance with a low price(A) or having C:D on a low level. The main effect A and the interaction between C and D are large in this experiment.

Table 14: Analysis of Variance for the reduced model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F value	P-value
A	0.009251	1	0.009251	7.250	0.0209
C	0.000013	1	0.000013	0.010	0.9220
D	0.000002	1	0.000002	0.002	0.9698
CD	0.005850	1	0.005850	4.584	0.05552
Residuals	0.014037	11	0.001276		
Total	0.0151163	15			

The table 14 shows the reduced model with the main effects of Price, Memory and Battery life. Table 14 summaries the analysis of the variance for this experiment. The model sum of squares is $SS_{model} = SS_A + SS_C + SS_D + SS_{CD} = 0.0151163$, which accounts for the most of the variability. Figure 4 shows the Plot of residuals versus predicted values and Normal probability plot of the residuals. Both plots are satisfactory.

The result of the experiment with response log of variance of the score is presented in the table 12.

Table 15: Effects, Regression Coefficients and Sum of Squares for the experiment with response log of variance.

Variable	Regression Coefficient	Estimated Effect	Sum of Squares
Overall Average	0.690		
A	0.0512	0.10	0.042
B	-0.0212	-0.04	0.007
C	-0.004	-0.01	0.0002
D	-0.005	-0.01	0.0004
E	0.0163	0.03	0.00423
AB	0.0375	0.08	0.0225
AC	0.0125	0.03	0.0025
AD	0.0137	0.03	0.003
AE	-0.01	-0.02	0.0016
BC	-0.015	-0.03	0.0036
BD	-0.024	-0.05	0.009
BE	0.003	0.00	0.0001
CD	0.034	0.07	3.0182
CE	0.0138	0.02	0.0025
DE	0.0286	0.06	0.013

The above table 15 contains effect estimates, regression coefficients, sum of squares form the 15 effects from this experiment.The figure 5 shows the normal probability plot of the effects. This plot shows that price has the most significant effect. Figure 6 shows the normal probability plot of the residuals and plot of residual versus predicted. We can also find the alternative analysis of estimated variances with following formula.

$$Var(ef\hat{f}ect) = \frac{\sum variances}{64} \tag{32}$$

and $\frac{effect}{sd(effect)}$ follows $N(0,1)$

$$Var(\hat{A}) = \frac{var(y_1) + ... + var(y_{16})}{64} \tag{33}$$
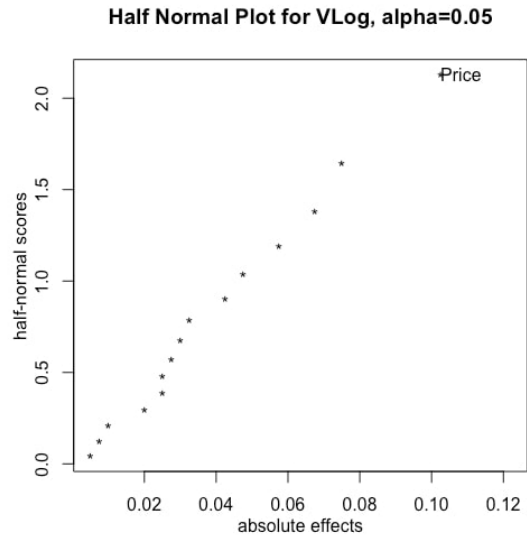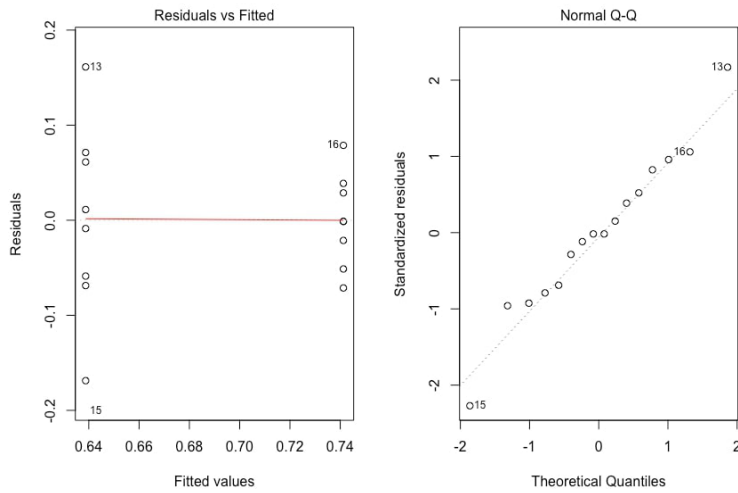
Figure 5: Normal Probability Plot of effects

Figure 6: Plot of residuals versus predicted and Normal probability plot of the residuals

Here we consider the log of variance as a response. Log gives a better approximation to the normal distribution. Now it is factor A that is larger than comes AB.

33

The reduced model only considering price as a main factor in the experiment.

Table 16: Analysis of Variance for the reduced model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F value	P-value
A	0.04346	1	0.04346	6.776	0.0209
Residuals	0.08979	14	0.00641		
Total	0.04346	15			

The table 16 shows the reduced model with the main effects of Price. Table 16 summaries the analysis of the variance for this experiment. Low price is desirable here in this experiment. Figure 4 shows the Plot of residuals versus predicted values and Normal probability plot of the residuals. Both plots are satisfactory.

9 Conclusion

This thesis is based on a case study of consumers choice of tablets considering three responses mean, variance and logarithm of variance. The conclusion is based on the result obtained from the analysis which is presented in the result section. Three responses produced distinct factors and factor combinations that are the most significant in the model.

In the first experiment main effects B(Multitasking), C(Battery Life) and D (Memory) are found to be the most significant effects on the mean score. Among these main effects B (multitasking) is the most significant one. In the reduced model all three main effects show the significant effect.

Where as price together with the interaction between battery life and memory are found to be the most significant in the second experiment with response variance. Which clearly shows in the normal-plot and also in the analysis. Finally, price is the most significant main effect in the third experiment with response logarithm of variance.

In the conclusion we can say that the third experiment with logarithm of variance as a response is the most desirable experiment. Since logarithm gives the better approximation to the normal distribution. So, low price is the most desirable.

The study has some limitations. This is a case study and result of this study can not generalize to the wider population.

References

[1] Wayne S DeSarbo, Venkatram Ramaswamy, and Steven H Cohen. Market segmentation with choice-based conjoint analysis. *Marketing Letters*, 6(2):137–147, 1995.

[2] Montgomery DouglasC. Design and analysis of experiments. douglas c. montgomery, 2009.

[3] Paul E Green and Vithala R Rao. Conjoint measurement for quantifying judgmental data. *Journal of Marketing research*, pages 355–363, 1971.

[4] Anders Gustafsson, Fredrik Ekdahl, and Bo Bergman. Conjoint analysis: a useful tool in the design process. *Total Quality Management*, 10(3):327–343, 1999.

[5] Vithala R Rao et al. *Applied conjoint analysis*. Springer, 2014.

[6] Dick R Wittink and Philippe Cattin. Commercial use of conjoint analysis: An update. *The Journal of Marketing*, pages 91–96, 1989.

NTNU

Norwegian University of
Science and Technology