Vincent Pun
MSDS 450 - Section 55 - Solo 2 Assignment - Discrete Choice Experiment
02/08/2021

INTRODUCTION

STC Technologies Company (STC), a reputable company in the televisions and audio systems space. Recently, STC announced its plans to release a computer tablet in the first half of 2015. Being new to this market, STC's goal will be to go to market with just one tablet configuration that is most popular with their target customers.

To gain understanding of customer preferences, Obee Juan, STC's product development manager, has collaborated with Neverending Marketing Insights, a global marketing research services provider, to design a choice-based conjoint (CBC) survey that will be sent to 424 respondents in the form of an online questionnaire.

Each respondent is asked for their preferences on 108 different product choices, and these are presented three at a time (total of 36 sets of choices). These configurations are created using combinations of attributes (design choices) that Obee Juan believes customers may find important when purchasing tablets, and they are: screen size, RAM, processor speed, retail unit price, and brand. Attributes in this study also contain multiple levels; for example, the three levels regarding retail price points are $199, $299, and $399. Modeling estimates and predictions are performed using this survey respondent data, and we are able to measure the degree of impact for every feature regarding product design preference.

DATA PREPARATION

Sources:

1. **stc-cbc-respondents-v3.RData (resp.data.v3)** - Contains the 424 individual responses to the conjoint survey
2. **stc-dc-task-cbr-v3(1).csv (taskV3)** - Outlines the attributes and levels that make up each of the 36 choice sets that are presented to the respondents
3. **Stc-extra-scenarios-v3.csv** - Contains two additional scenarios (product design choices) that are used to make predictions using the designed multinomial logit regression model(s)
4. **efCode.RData** - Contains R functions that will be used to create predictor variables for regression models based on existing attributes and levels

Preparation:

1. **Design Matrix** (X.matrix) - Attributes are prepared through the use of effects coding, which is a technique that allows categorical variables such as Brand, Price, Screen, RAM, and Processor to be used to estimate coefficients for regression modeling. These product combinations are then combined with three columns that represent the

interaction between brand and price, which is calculated using multiplication in this study.

2. **Choice Responses** (ydata) - Extracted STC survey responses for the 36 choice sets. This data subset excludes any missing variables and is converted into a matrix to be used for subsequent analysis.
3. **STC Product Ownership** (zowner) - To interpret how previous STC product ownership affects product preferences, the variable within the response data, vList3, is simplified to only indicate whether (1) or not (0) a participant has ever owned an STC product.
4. **Combined Data Structure Input** (lgtdata) - Data that is used with rhierMnlDP, which is a function of R's bayesm package; a list containing the Design Matrix (X.matrix) and the Choice Responses (ydata) is created for each respondent.

MODEL DESIGN

Two models are created to calculate respondent level estimates of the effects of attribute levels on survey choices. Markov Chain Monte Carlo simulation is used to estimate parameters for both Hierarchical Bayes Multinomial Logit (MNL) regression models.

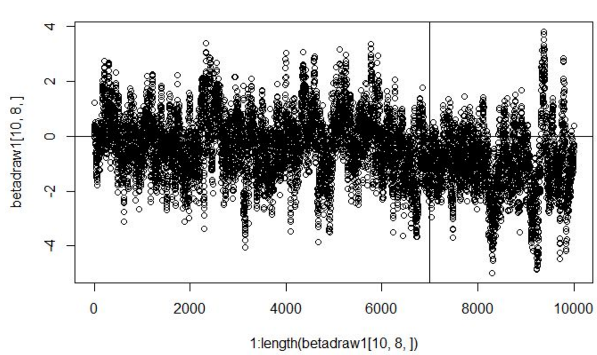
Model 1

Model 1 is fitted using 50,000 iterations, and every 5th sample is kept. In result, a total of 10,000 samples are available for each of the 424 respondents. Given that there is no prior information, default priors from the MCMC algorithm are used.

By visualizing the posterior distributions across ten unique respondents for each of the fourteen beta coefficients, we are able to identify whether these coefficients converge throughout the course of the iterations. On Figure 1, we can see that the simulation begins with a beta coefficient of 0, and there appears to be a high variance in these estimates. A conservative burn-in period is essential to stabilize the MNL regression model, so it can be reliable when making predictions on new data.

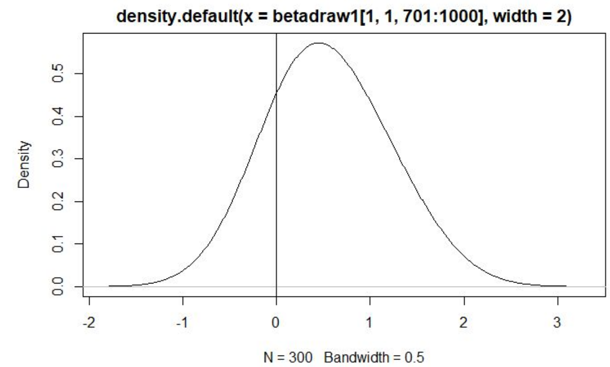
After reviewing the beta coefficients for a random subset of survey respondents, it was determined that the model stabilized around 7,000 iterations. Thus, the final 3,000 out of 10,000 samples across all respondents are used to compute average beta coefficient values that reflect respondent preferences for each attribute. Positive coefficient values are interpreted as having a preference for an attribute, and negative coefficient values represent a negative preference towards an attribute; thus, values close to zero suggest indifference/no preference.

Figure 1: (Right)
Plot of beta coefficient #8 (Price $399) over the iteration number for respondent #10



To visualize individual respondent preferences, we can also view density plots for each of the attribute coefficients. For example, we can see that Respondent #1 prefers tablets that have 7-inch screens.

Figure 2: (Right)
The density plot of beta coefficient #1 (Screen 7) for Respondent #1



Model 2 - Impacts of STC Product Ownership
A second model is fitted using the same number of iterations, retained samples, and cut-off parameters based on burn-in period used in Model 1. The difference in Model 2 is that it includes the “zowner” covariate that indicates whether survey respondents are previous owners of STC products.

ASSESSING MODEL FIT

Name	Confusion Matrix	Accuracy (%)	AUC (%)
Model 1	ydatavec custchoice 1 2 3 1 3682 418 213 2 229 3761 209 3 293 410 6049	88.39	86.62
Model 2 *(Prior STC Product Ownership)*	ydatavec custchoice 1 2 3 1 3678 416 215 2 235 3743 214 3 291 430 6042	88.20	86.65

Goodness of fit for each model is evaluated through the use of both confusion matrices and areas under ROC curves. It is seen that the two models have very similar accuracy and AUC results, which suggests that the covariate describing STC product ownership is not very significant regarding the MNL regression model.

FACTORS AND LIMITATIONS

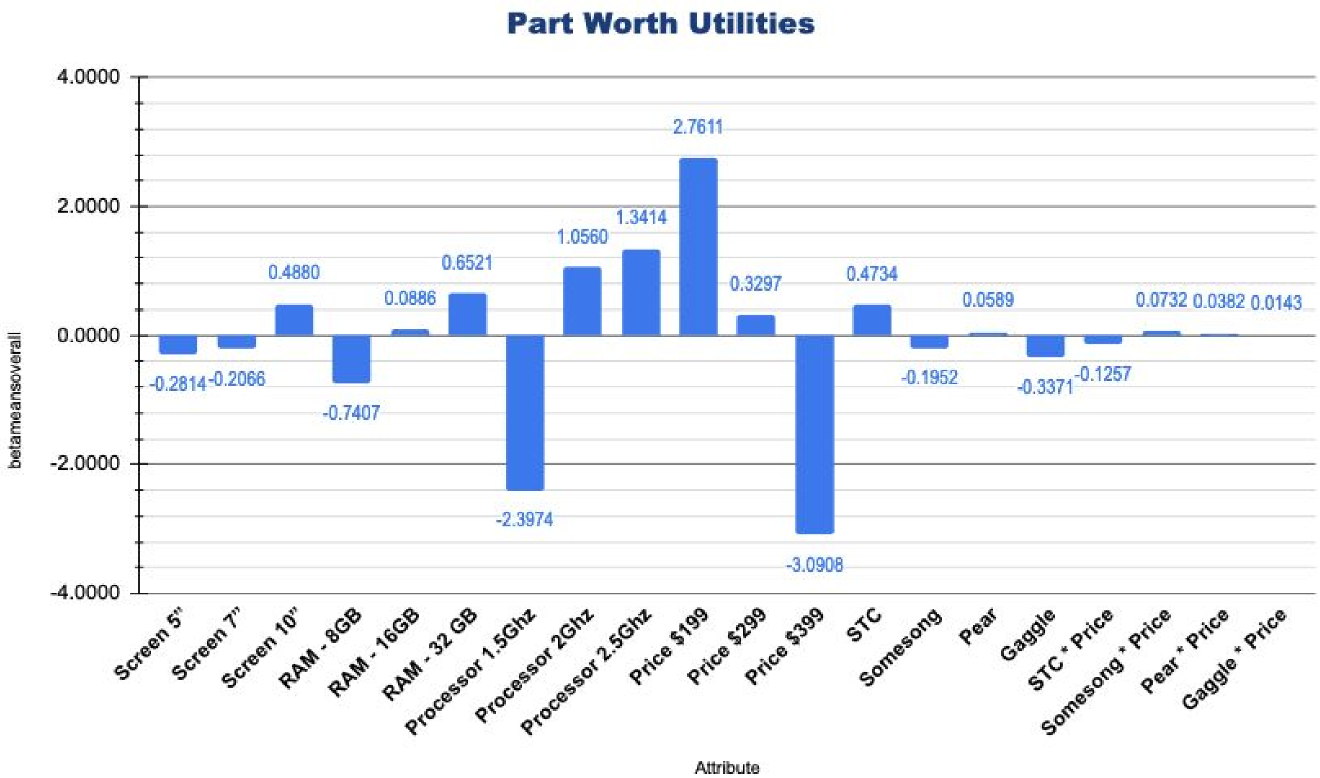
- **Limited Data** - Only 68 out of 424 (16.04%) respondents noted that they previously owned an STC product. This small sample makes it difficult to truly determine whether STC product ownership may impact customer preferences between Model 1 and Model 2.
- **Conjoint Analysis Constraint** - Making important decisions based on market simulator results can often be challenging, for conjoint analysis predictions are often unable to account for real-world characteristics that may affect buying patterns. Further, marketing managers may find challenges in accounting for how purchasing decisions may be

affected by external factors such as advertising and changes in their competitors' product offerings. Also, shares of preferences may not be indicative of actual market shares if simulations do not take into account the quantity purchased by each respondent.

Attribute Preferences

The average values of the betas from all the respondents (betameansoverall) are presented below for model interpretation. However, customer choice predictions are made using the 424 individual respondent MNLR models because it produces more accurate results.

Plotted below are the log odds ratios for all of Model 1's coefficients. These parts worth utilities indicate whether a certain product design choices are preferred (positive value) or not preferred (negative value) and should be interpreted as a summary of Model 1's results. For example, it is clear that tablet configurations with larger screen sizes (10 inch) and low prices ($199) are highly favorable amongst respondents.



#	Attribute	Log Odds (betameansoverall)	Odds Ratio exp(betameansoverall)	Respondent Preference Level
1	Screen 5”	-0.2814	0.7547	Weak Negative
2	Screen 7”	-0.2066	0.8133	Weak Negative
3	Screen 10”	0.4880	1.6290	Moderate Positive
4	RAM - 8GB	-0.7407	0.4768	Moderate Negative
5	RAM - 16GB	0.0886	1.0927	Weak Positive

6	RAM - 32 GB	0.6521	1.9196	Moderate Positive
7	Processor 1.5 GHz	-2.3974	0.0910	**Strong** Negative
8	Processor 2 GHz	1.0560	2.8749	**Strong** Positive
9	Processor 2.5 GHz	1.3414	3.8244	**Strong** Positive
10	Price $199	2.7611	15.8172	Very Strong Positive
11	Price $299	0.3297	1.3905	Weak Positive
12	Price $399	-3.0908	0.0455	**Strong** Negative
13	STC	0.4734	1.6054	Moderate Positive
14	Somesong	-0.1952	0.8227	Weak Negative
15	Pear	0.0589	1.0607	Weak Positive
16	Gaggle	-0.3371	0.7138	Weak Negative

#	Brand Price Interaction	Log Odds betameansoverall	Odds Ratio exp(betameansoverall)	Respondent Preference Level
17	STC*Price	-0.1257	0.8818	Weak Negative
18	Somesong*Price	0.0732	1.0759	Weak Positive
19	Pear*Price	0.0382	1.0390	Weak Positive
20	Gaggle*Price	0.0143	1.0144	Weak Positive

Another way to interpret customer preferences can be to use the odds ratio values, which are the inverse of the log of odds; for example, the odds of preference for 32GB of RAM can be understood as being 1.92 times greater than not preferring this design choice. This is calculated by taking the exponential function of the log of odds.

Management can also make comparisons between attribute levels. While all other attributes are held constant, a comparison is made by calculating the exponential of the difference between log odds values. For example, researchers can calculate that customers prefer to have 2.5GHz processors 33% more than 2 GHz processors (exp(1.3414-1.0560)).

Part of this study is to assess whether price sensitivity varies over brands. Based on the results above, it is clear that price elasticity is low between brands, which implies that customers care more about other factors such as screen size, RAM, price, and processing power when shopping for new tablets.

MODEL PREDICTION (2 Extra Scenarios)

Using Model 1, we are able to predict choices for Obee's two additional choice set scenarios. Predictions are made using the individual respondents' models. As we can see, Option #2 was most preferred for both choice sets, and the common characteristic between both design choices is that they cost $199.

Scenario #	Results	Specifications (Screen / RAM / Processor / Price / Brand)

1	`extra1`
	`1 2 3`
	`93 234 97` Option #2
		10 inch / 32 GB / 2 GHz / $199 / STC

2	`extra2`
	`1 2 3`
	`96 244 84` Option #2
		5 inch / 16 GB / 1.5 GHz / $199 / Gaggle

RECOMMENDATION

#	Specifications (Screen / RAM / Processor / Price / Brand)
1	7 inch / 16 GB / 2.5 GHz / $199 / STC
2	7 inch / 16 GB / 2.5 GHz / $199 / Pear
3	7 inch / 16 GB / 2.5 GHz / $199 / Gaggle
4	**10 inch** / 16 GB / **2.0 GHz** / **$299** / Somesong
5	10 inch / 16 GB / 2.0 GHz / $299 / STC
6	**5 inch** / 16 GB / **2.5 GHz** / **$199** / STC

Using Model 1, we are able to predict customer preferences across the 36 choice sets in the survey. The top three most preferred design choices are very consistent, as all of these tablets have 7-inch displays, 16 GB RAM, 2.5 GHz processors, and cost $199; the only difference is the brand, which is known to have weak impacts on customer choice.

Design choices for the most preferred product (excluding brand) is compared to the runner-up configuration below:

- **Screen Size** - Respondents prefer 10-inch screens **200%** more than 7-inch screens
- **Processor** - Respondents prefer 2.5 GHz processors **33%** more than 2.0 GHz processors
- **Price** - Respondents prefer the $199 price point **1137%** more than the $299 price point

With these results, it is recommended that STC should go to market with just one tablet, which will be priced at $199. While some customers may prefer alternative designs, it is important to understand that the findings in this conjoint study may not be reflective of STC's actual market share in the tablet space. Thus, STC will find it beneficial to conduct subsequent research on tablet preferences after they've released their first product into this competitive market.