Nate Bitting
PREDICT 450 - Marketing Analytics
Solo 2 – Starful Technologies Company (STC): Discrete Choice Experiment

Data Preparation Methodology

The first phase of this discrete choice experiment was to examine the datasets provided by Starful Technologies Company (STC). STC was kind enough to provide documentation of the variables included in the experiment which include the following:

- *STCOwner* – previous owner of an STC product?
- *36 Choice Set* questions with varying attribute options
- *4 Interest questions:* 1) Purchasing a new tablet, 2) Purchasing a new smart phone, 3) Using cloud storage for storing personal digital content, and 4) Taking an online course to improve relevant skills
- *Gen* – gender for each respondent
- *Age* – age ranges of each respondent

The questionnaire choice set questions for this experiment included five attributes as outlined below:

- *Brand* – 4 levels: STC, Somesong, Pear, Gaggle
- *Price* – 3 levels: \$199, \$299, \$399
- *Screen* – 3 levels: 5 inch, 7 inch, 10 inch
- *RAM* – 3 levels: 8 Gb, 16 Gb, 32 Gb
- *Processor* – 3 levels: 1.5 GHz, 2 GHz, 2.5 GHz

The first dataset provided by STC was the task plan for each of the choice set questions. The first data step was to subset the task attribute options from the choice task plan dataset and then convert it into a matrix. We then had to process this matrix to apply effects coding to the categorical features in our task plan dataset. One of the objectives of this analysis is to assess any variation in price sensitivity between brands. To do this we also created a vector called *pricevec* that measures the price for each task in relation to the mean task price. We also then had to create another matrix that includes brands, prices, and interaction terms between brand and price. We combined these two matrices into a single matrix (*X.matrix*) that we used as our predictor variables in the modeling phase. The questionnaire dataset provided by STC contained results from 360 respondents who answered the following questions: *STCOwner, 36 choice sets, 4 interest questions, gender, and age*. We took this data and applied effects coding to create a new matrix for our response variable set, *ydata.* We then created a list, called *lgtdata*, that combined both the choice task plan dataset along with the respondent dataset that can be used for estimating our choice models.

Model Estimation

Now that we have completed the data preparation, we can now enter into the model estimation phase. For this assignment, our objective was to leverage a Hierarchical Bayes multinomial logit model and we have elected to use the *rhierMnlDP()* function from the *bayesm* R package. Before we can begin, we must first create the necessary inputs for the rhierMnlDP() function. First, we need to know how many iterations and how many samples to keep for the Markov Chain Monte Carlo (MCMC) simulation. We decided to use 30,000 iterations and to keep every 30th sample which we stored in a list labeled *mcmctest*. We then have to create our input dataset in a list that includes the choice set size, labeled *p*, and the *lgtdata* list we covered earlier. We are now ready to estimate our first model.

Based on the objectives outlined in the assignment, we have been tasked with estimating two separate models. The first model (Model 1) only includes the choice set responses for all 360 respondents and the second model (Model 2) includes two additional covariates for *STCOwner* and *Gender*.

Model Evaluation

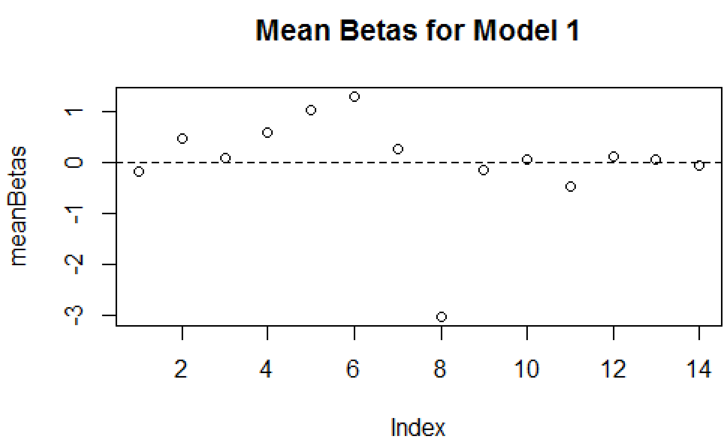
There are several statistical methods to assess the overall model fit for DCE models that include the following:

- Mean Absolute Error (MAE) or Mean Squared Error (MSE)
- Likelihood-related measures – LL, RLH, pseudo R-squared
- Bayes Factor
- Posterior predictive checks

In order to calculate the statistical measures above, we had to first calculate the posterior means of each beta for each subject. The next step is to calculate the X times the means of each subject from the previous step. Using this matrix, we put the subjects into rows with choice sets stacked within subjects and then exponentiating those betas. Lastly, we took the exponentiated betas and divided it by the row sums of the exponentiated betas. These results can be leveraged to calculate the measures listed above to assess overall model fit.

Other means of assessing the models is by doing a simple "sniff test" to see if the beta means make logical sense. Figure 1, below, shows a plot of the beta means for Model 1. Upon reviewing the mean betas for each attribute, we can assume the model is adequate in that it passes the "sniff test" due to the mean betas having logical results.

Figure 1: Model Mean Betas



Index Legend

1) 16 Gb RAM
2) 32 Gb RAM
3) 2 GHz Processor
4) 2.5 GHz Processor
5) 7 inch Screen
6) 10 inch Screen
7) $299
8) $399
9) Somesong Brand
10) Pear Brand
11) Gaggle Brand
12) Somesong by Price
13) Pear by Price
14) Gaggle by Price

We also need to ensure we have enough burn-in which can be assessed by plotting the betadraws and the log likelihood model outputs. Figures 2 and 3 provide the betadraw and log likelihood plots, respectively.

Figure 2: Betadraw Plot

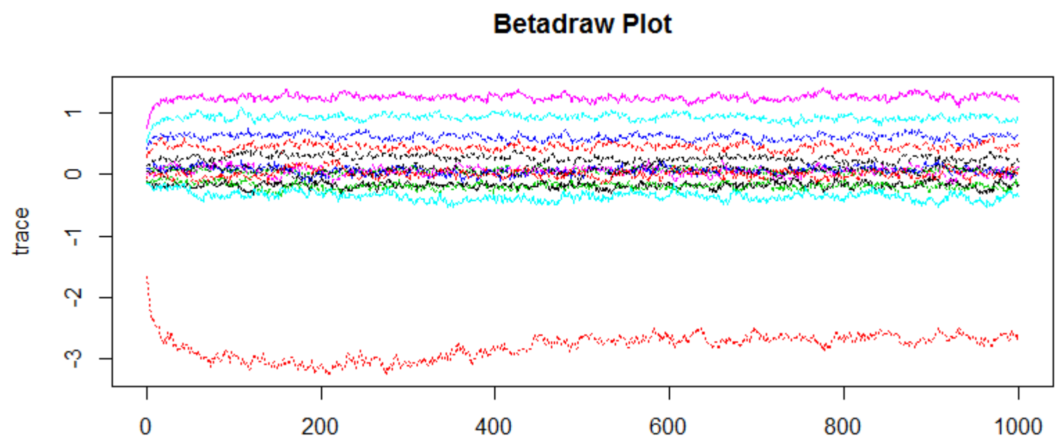
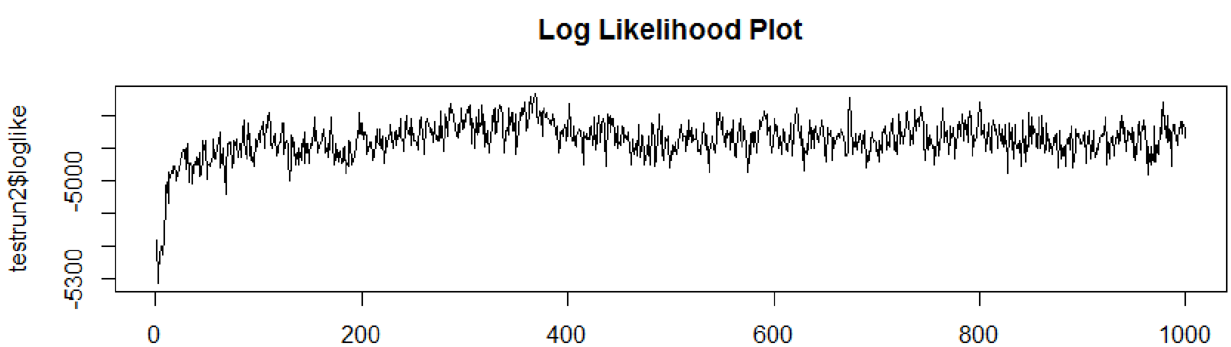


Figure 3: Log Likelihood Plot



Due to the fact that the patterns level out around iteration 500, we can assume we had enough iterations to allow for enough burn-in to ensure adequate model performance.

Model Interpretation

Now that we have assessed model fit, we can now explore what the model results indicate to the choice selections from the STC discrete choice experiment. Based on the mean betas shown in Figures 1 and 2, we can see there are no differences between each model. We need to understand the sensitivity for each attribute in relation to the base level option for each attribute. The results shown in Figure 1 allow us to make some conclusions regarding the choice preferences for each attribute. Below we will provide our analysis for each attribute based on these results.

- **RAM** – it appears respondents preferred 8Gb of RAM over 16 Gb, but preferred 32 Gb of RAM vs 8 Gb.
- **Processor** – respondents had little sensitivity between 1.5 GHz and 2 GHz processor speed, but highly preferred 2.5 GHz over both other options.
- **Screen size** – the results of the mean betas for screen size indicate that respondents prefer larger screens vs smaller screens. It is clear that the mean betas increase as the size of the screen increases with 10-inch screens having the highest preference.
- **Price** – it appears the $299 price is acceptable to respondents, but the mean betas of the $399 price drop dramatically.
- **Brand** – both the Somesong and Gaggle brands are not viewed as equitable choices when compared to the STC brand, with Gaggle being the worst. Pear, on the other hand, is slightly favored over the STC brand.
- **Brand by Price** – when comparing the brands by mean price compared to STC, we can see that Somesong and Pear have slightly higher mean betas where Gaggle is slightly lower. It does not appear that price sensitivity varies between brands as they are relatively close to zero.

To assess whether price sensitivity varies by brand, we first leverages the beta matrix for the last 300 of the 1,000 samples for 12^{th}, 13^{th}, and 14^{th} betas in the betadraw array for Somesong by Price, Pear by Price and Gaggle by Price. We then used a custom function to calculate where zero falls relative to the distribution of each row of the matrix. We applied this function against the matrix and stored it in a vector called *zp*. We created a vector of zeros with the same number of rows as the matrix and labeled each observation with a 1 where the *zp* result was less than 0.05 or greater than 0.95. Leveraging this vector, we subsetted the original beta matrix for the respondents that met the criteria of our *zp* vector.

By applying this methodology to the 12^{th}, 13^{th}, and 14^{th} betas, we can see how many respondents whose betas are likely to be different than zero. Table 1 below provides us with the results of the procedure described above.

Table 1: Price Sensitivity by Brand

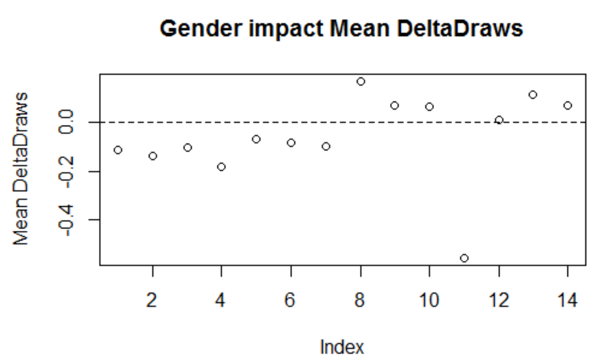
Betas	# of Respondents whose betas are likely to be different than zero	% of total respondents
Somesong by Price	17	5%
Pear by Price	5	1%
Gaggle by Price	6	2%

Based on these results in Table 1, it is safe to conclude that price sensitivity does not vary by brand due to the limited number of respondents whose betas are likely to be different than zero.

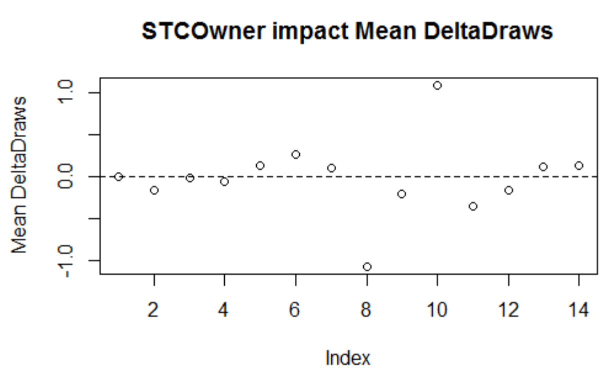
Gender and STC Ownership Impact

The next objective we have been tasked with is to assess how adding additional covariate impact the effects of the attributes on preferences. Figure 4 illustrates plots of the mean betas for all attributes in relation to *Gender* and *STCOwner*.

Figure 4: Gender and STCOwner impact on attribute preferences



Index Legend
1) 16 Gb RAM
2) 32 Gb RAM
3) 2 GHz Processor
4) 2.5 GHz Processor
5) 7 inch Screen
6) 10 inch Screen
7) $299
8) $399
9) Somesong Brand
10) Pear Brand
11) Gaggle Brand
12) Somesong by Price
13) Pear by Price
14) Gaggle by Price



Index Legend
1) 16 Gb RAM
2) 32 Gb RAM
3) 2 GHz Processor
4) 2.5 GHz Processor
5) 7 inch Screen
6) 10 inch Screen
7) $299
8) $399
9) Somesong Brand
10) Pear Brand
11) Gaggle Brand
12) Somesong by Price
13) Pear by Price
14) Gaggle by Price

We can see that males prefer the base options for RAM, processor speed and screen size; prefer the non-STC brands; and have slight price sensitivity variation with Pear and Gaggle brands. Those respondents who have owned an STC device previously appeared to show more loyalty for STC products with relation to the mean betas for all attributes with exception to the Pear brand. For some reason, they are indifferent to the technical attribute levels, but have clear favorability towards the Pear brand over the STC brand.

Part-Worth Utility Calculations

One of the objectives of the assignment was to explain how to calculate the part-worth utilities for each attribute. To do this, one can simply count the number of times an attribute was selected divided by the number of times it was available to be selected in one of the task plan questions. This provides a proportion of how many times that attribute level was selected. These proportions are what we call the part-worth utilities, which can be used to measure the relative importance of attributes.

New Scenario Choice Predictions and Preference Share Estimates

The final task of the assignment is to predict choices and estimate preference shares for the alternatives in each of Obee's four additional scenarios provided. These four scenarios provide us with alternative task questions to provide to the customer and we'd like to estimate the probability of choice for each new scenario task question. To do this, we elected to leverage a simple approach leveraging the expected values. The first step is to create the *X.matrix* using the additional scenarios provided and processing them through our effects coding process (as we did with our original task plan). We then need to calculate the posterior means for each beta for each subject by using the betas from our HB model (this results in a 360 x 14 matrix). The next step is to do some basic matrix algebra by multiplying *X.matrix* (from four additional scenarios) by the transpose of the mean betas matrix (this results in a matrix of 108 x 360). Using this matrix, we can create a "stacked" matrix of the choice sets (this results in a 12,960 x 3 matrix) and exponentiate the stacked matrix. Finally, we take the exponentiated stacked matrix and divide it by the row sums of that same stacked matrix. This provides us with a matrix containing the choice probabilities using the four additional scenarios.

The next piece of this objective was to estimate the share preferences by performing a sensitivity analysis that can be used by the product management organization to understand what features or attributes should be included in the product. We can calculate these by simply taking the conjoint part-worth utilities and perform a sensitivity analysis to estimate the differences in share based on selecting one attribute level over another. This sensitivity analysis enables us to understand how selecting one attribute level over another can improve (or worsen) the product's overall preference based on the results of the conjoint experiment.

Recommendation

The good news is that we concluded that price sensitivity did not vary between brands, which indicates STC is in a good position to enter the tablet market. We also now have a better understanding of the sensitivity of various attribute levels, and STC is in a better position to know what tablet makes the most sense to take to market. As we covered in the model interpretation section earlier, we now can leverage the mean betas to assess preference when compared to the base level attributes. Below contains a table with the optimal attribute levels for the STC tablet.

Table 2: Optimal Attribute Levels for STC Tablet

Attribute	Optimal Attribute Level
RAM	32 Gb
Processor	2.5 GHz
Screen	10 inch
Price	$299

We feel these attribute levels will put STC in the best position to be successful in entering the tablet market. Given the fact that we saw previous STC product owners had an affinity towards the Pear brand, it may be in STC's best interest to have a marketing campaign comparing themselves to the Pear brand. Another consideration STC should take into account is that their male customers seem satisfied with the base levels for all attributes and do not like the Gaggle brand. STC should use this information to their advantage during the creation of their tablet marketing campaign.