Chris Pelkey
PREDICT 450-55
Solo 2

Introduction

Set forth below is a data analysis of the tasks from the choice-based conjoint survey conducted by Neverending Marketing Insights. This analysis was ordered by Obee Juan, Star Technology Company's product development manager. Outlined below is an in-depth analysis of the data's configuration to run HBMNL models, the differences between various models, determining whether price sensitivity is the same over brand choice, prediction over two additional scenarios, description of attribute level part-worths and description of attribute importance. In addition, we will talk about the limitations of this assessment and the product that Star Technology Company should take to market.

Data configuration

Throughout the course of this paper, you should keep this data configuration in mind. The configuration in large part adjusts the outcomes of the analysis. To begin, 424 records of data were loaded into the system reflecting 424 respondents to the survey. These records contain 55 variables representing demographic data like state and household income, a unique identifier code and 36 discrete choice modeling questions.

These modeling choices need to be processed so that they can be fully utilized in this analysis. The best way to do this is to create a design matrix that reflects the attribute levels of the choice task alternatives. This is done by creating dummy variables for each of the categorical predictor variables. This data has one attribute with four levels (brand), and it also has attributes that each have three levels: RAM, screen size, price and CPU. These dummy variables allow the descriptor variables to be accurately modeled.

Next, the price interaction variables are created by making a new variable with price being treated as continuous and centering the mean on 0. The price is then multiplied by the columns that represent the brands. These variables are then added back into the design matrix.

The task choices are then loaded into a separate matrix that will allow us to run the analysis. After this load in we determine also if the respondent has ever owned an STC product before, this will be useful in the final analysis. A basic overview of the task choices can be seen below in Table 1, with the choice for each task that was most selected in bold and red:

	DCM1_1	*DCM1_2*	*DCM1_3*	*DCM1_4*	*DCM1_5*	*DCM1_6*	*DCM1_7*	*DCM1_8*	*DCM1_9*
[1,]	93	**316**	**218**	125	**157**	70	81	65	100
[2,]	**242**	45	180	28	132	83	**182**	85	21
[3,]	89	63	26	**271**	135	**271**	161	**274**	**303**
	DCM1_10	*DCM1_11*	*DCM1_12*	*DCM1_13*	*DCM1_14*	*DCM1_15*	*DCM1_16*	*DCM1_17*	*DCM1_18*
[1,]	25	**271**	192	77	**151**	18	65	22	64
[2,]	**307**	113	**217**	75	148	140	**207**	130	68
[3,]	92	40	15	**272**	125	**266**	152	**272**	**292**
	DCM1_19	*DCM1_20*	*DCM1_21*	*DCM1_22*	*DCM1_23*	*DCM1_24*	*DCM1_25*	*DCM1_26*	*DCM1_27*
[1,]	41	**303**	**225**	114	**183**	37	71	32	109
[2,]	**276**	42	147	25	88	88	**177**	103	19
[3,]	107	79	52	**285**	153	**299**	176	**289**	**296**
	DCM1_28	*DCM1_29*	*DCM1_30*	*DCM1_31*	*DCM1_32*	*DCM1_33*	*DCM1_34*	*DCM1_35*	*DCM1_36*
[1,]	34	**312**	189	88	**164**	27	44	35	86
[2,]	**308**	67	**209**	43	121	114	**207**	116	36
[3,]	82	45	26	**293**	139	**283**	173	**273**	**302**

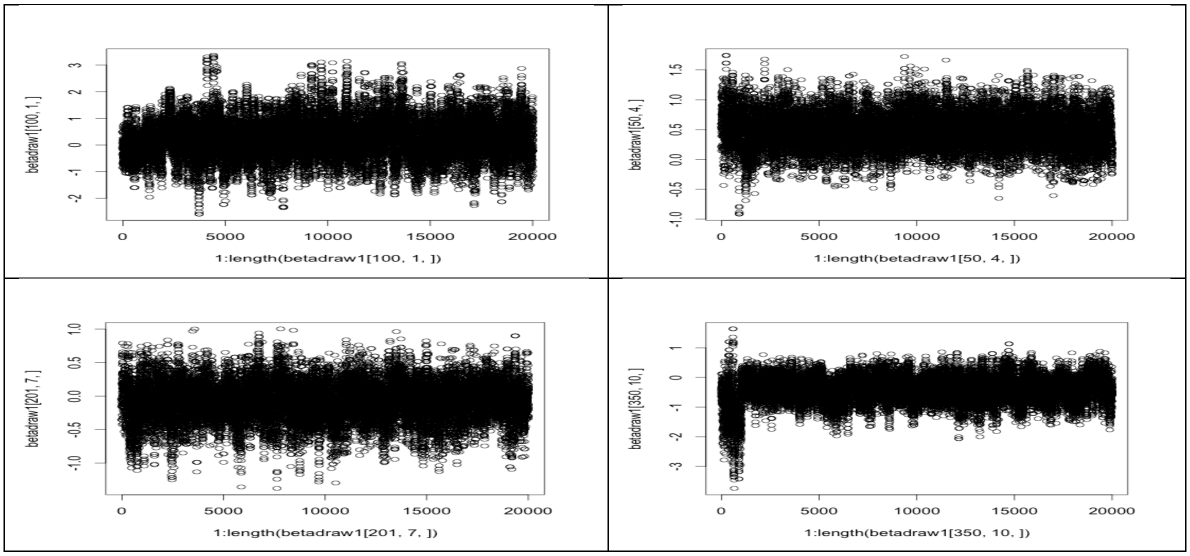
Table 1 – List of choice count for each task

The final piece of closing this data configuration section out is to convert these matrices into a list of lists. This allows the data to be read into the appropriate R package, which will allow us to build the prediction models.

Models and differentiation

MCMC model

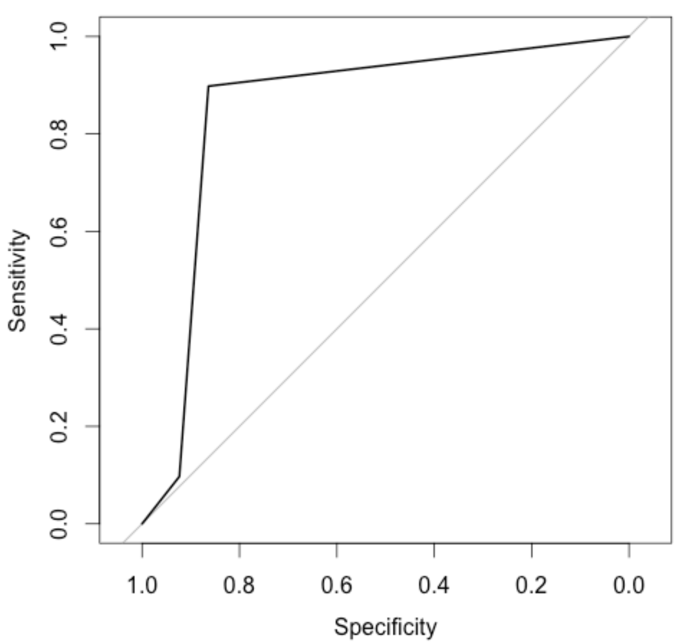
For the analysis here we used 100,000 iterations of the Markov-Chain Monte Carlo simulation and accepted every fifth draw. This allowed for stabilization over time. As can be seen in the figured below, the stabilization begins to occur and around the 70,000 mark, or around draw 14,000.



These simulations are then run through statistical processes to gain insight into what the general feel is for the 6,000 stabilized draws. As can be seen below, there are general log likelihood directions for the data.

1	2	3	4	5	6	7
-0.181	0.452	0.105	0.583	1.059	1.269	0.320
8	9	10	11	12	13	14
-2.974	-0.197	0.047	-0.357	0.068	0.040	-0.014

Processing this model we can understand that there is a solid ROC curve with an area under the curve of 0.8531. This is a fairly good score and means that the model is good. Below you can see the ROC curve, noting that the closer it is to a 90° angle, the better the model fits.



Further, the confusion matrix presented below shows how the model did at accurately predicting choices made. Simply explained, this shows that of the 15,264 possible choices, the model accurately predicted that the respondent would choose the third option 6,025 times and that when it was guessing that the answer would be the third option it was accurate 93% of the time.

	1	2	3	
1	3631	468	216	0.841
2	251	3676	230	0.884
3	322	445	6025	0.887
	0.864	0.801	0.931	

This model did a very good job of predicting the outcomes of the choices of the respondents. It also has a log likelihood of -5921.547.

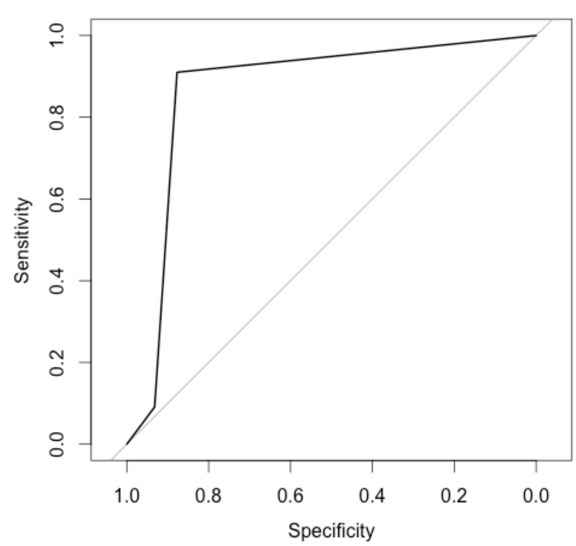
MCMC model with previous ownership

In addition to running a very basic model, a model was run that takes into consideration whether the respondent already owns a tablet. This model

1	2	3	4	5	6	7
-0.045	-0.070	0.075	-0.035	0.190	0.391	-0.023
8	9	10	11	12	13	14
-0.484	-0.269	1.070	-0.083	-0.227	0.062	0.173

These numbers actually differ from the original model. The coefficient for variable 2 for example, is actually a negative when it was a positive in the first model.

The ROC curve for this model seems to be very similar to the original model. The previous ownership model has an area under the curve of 0.8682. This is again good and just slightly beats out the original model.



An examination of the confusion matrix for the previous ownership model shows that it is largely similar to the original model and performs only slightly better.

	1	2	3	
1	3687	413	218	0.854
2	231	3758	213	0.894
3	286	418	6040	0.896
	0.877	0.819	0.933	

The model in comparison to the original model had a slightly worse log likelihood rating of -5823.108.

For all intents and purposes the two models are largely the same. They both had extremely good ROC curves, a nice area under the curve score, good confusion matrices and had good, yet similar log likelihood.

The model is as follows: X1 * -0.04495419 + X2 * -0.07019187 + X3 * 0.0746721 + X4 * -0.03505722 + X5 * 0.19003373 + X6 * 0.39112921 + X7 * -0.02286199 + X8 * -0.48484736 + X9 * -0.2688393 + X10 * 1.0700499 + X11 * -0.08274555 + X12 * -0.2273129 + X13 * 0.06234765 + X14 * 0.17288238.

Outlined below is the summary table from the model.

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]
[1,]	88	346	242	107	190	67	77	59	104	31	307	198
[2,]	251	29	163	13	106	52	191	66	9	306	91	215
[3,]	85	49	19	304	128	305	156	299	311	87	26	11

	[,13]	[,14]	[,15]	[,16]	[,17]	[,18]	[,19]	[,20]	[,21]	[,22]	[,23]	[,24]
[1,]	55	176	17	46	16	56	38	330	260	109	208	31
[2,]	69	125	118	223	116	62	282	27	115	12	55	75
[3,]	300	123	289	155	292	306	104	67	49	303	161	318

	[,25]	[,26]	[,27]	[,28]	[,29]	[,30]	[,31]	[,32]	[,33]	[,34]	[,35]	[,36]
[1,]	50	32	92	36	331	224	72	183	22	28	25	62
[2,]	191	83	10	310	51	178	29	97	87	224	100	26
[3,]	183	309	322	78	42	22	323	144	315	172	299	336

Attribute level descriptions and price sensitivity

It would come as no surprise that the two different models resulted in entirely different mindsets when it comes to tablets. The first table outlined below is the attribute level odds ratio for the various attributes in the model that does not reflect previous ownership of an STC device.

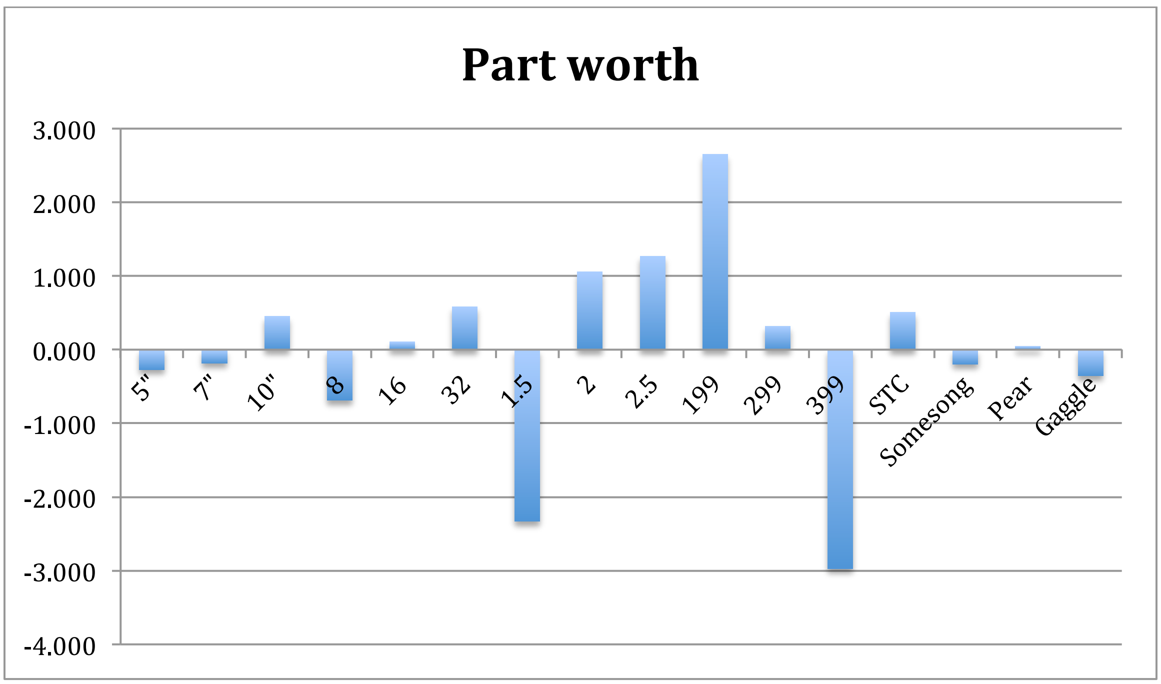
Factor	Attribute	betameans	Odds Ratio
Screen	5"	-0.271	0.763
	7"	-0.181	0.834
	10"	0.452	1.571
RAM	8	-0.688	0.503
	16	0.105	1.111
	32	0.583	1.791
Processor	1.5	-2.328	0.097
	2	1.059	2.883
	2.5	1.269	3.557
Price	199	2.654	14.211
	299	0.32	1.377
	399	-2.974	0.051
Brand	STC	0.507	1.660
	Somesong	-0.197	0.821
	Pear	0.047	1.048
	Gaggle	-0.357	0.700
Brand*Price	STC*Price	-0.094	0.910
	Somesong*Price	0.068	1.070
	Pear*Price	0.04	1.041
	Gaggle*Price	-0.014	0.986

Determining the baseline variables is an easy task once we have the betameans for the dummy variables. For example, in order to determine the betameans for the 5" screen, we simply take the other two betameans, multiply by negative 1 and add them together. So to get the betameans for 5" screen it is -1*(-0.181)+-1*(.452) = 0.181-0.452 = -0.271. In order to find these attribute level part worth, we must simply multiply all values that we have in the group by negative 1 and add them together.

In the table above we can see a whole host of interesting findings. For example there is a very clear winner on pricing. The odds of preference of a price of $199 is 14.21 times greater than that of not preferring. In addition, with all things being equal, the odds of preference of a buying an STC is 1.66 times greater than that of not preferring. When looking at the attribute level of brand, we can see that STC is strongly preferred in the mix.

Purchase	Odds Ratio
STC over Somesong	2.022
STC over Pear	1.584
STC over Gaggle	2.373

The way to determine these odds ratios is by taking the exponential function of the betameans. So, in order to determine the odds ration for the price of $199 for example is OR = $e^{2.654}$ = 14.211. It is an easy enough function to determine. Determining the purchase preference is also fairly easy. It is simply another exponential function, with the betameans of your primary choice minus the secondary choice. For example, the odds ration of picking HTC over Somesong would be OR = $e^{0.507-(-0.197)} = e^{0.704}$ = 2.022.



To get a clear view of what works and what doesn't you can see the above graph. Clear winners are a 10" screen, 32 RAM, 2.5 CPU, $199 price and STC brand.

Looking carefully at price sensitivity, we can see that all things being equal, there is some sensitivity to price across brands. Luckily for STC they are always the preferred brand, but as price goes up, respondents are less likely to choose the STC model. In measuring each of the brands against the STC model, as the price goes up the odds ratio goes down. This should indicate that the price of the model that we put together should remain low, if we want people to purchase STC tablets in the future.

Brand preference	Price	Odds ratio
STC to Somesong	$199	2.377
	$299	2.022
	$399	1.719
STC to Pear	$199	1.811
	$299	1.584
	$399	1.385
STC to Gaggle	$199	2.570
	$299	2.373
	$399	2.190

Looking at the scenario of determining what effect having previously purchased an STC device has on the model we actually get some not great results. Any item with a negative deltameans below indicates that having a previously purchased an STC has a negative effect on the purchasing of that attribute in the future. Having previously purchased an STC means that they are less likely to do so. It does look like having a previous purchase lowers the acceptable screen size to 5", the RAM to 16 and the $299 price is not as big of a shock to the respondent.

Factor	Attribute	deltameans
Screen	5"	0.115
	7"	-0.045
	10"	-0.07
RAM	8	-0.040
	16	0.075
	32	-0.035
Processor	1.5	-0.581
	2	0.190
	2.5	0.391
Price	199	0.508
	299	-0.023
	399	-0.485
Brand	STC	-0.718
	Somesong	-0.269
	Pear	1.070
	Gaggle	-0.083
Brand*Price	STC*Price	-0.008
	Somesong*Price	-0.227

Pear*Price	0.062
Gaggle*Price	0.173

Additional scenarios

As asked for by Obee Juan some additional scenarios were run. We can see the extra scenarios preferences below. For scenario 1, the preference shares are 7.5%, 68.9% and 23.6% respectively. Fore scenario 2 the preference shares are 53.6%, 46.1% and .3%.

	[,1]	[,2]	[,3]
[1,]	0.075	0.689	0.236
[2,]	0.536	0.461	0.003

This means that in scenario 1, the preferred model is 10" screen, 32 RAM, 2.5 CPU, HTC and at a price of $199. For scenario 2, although choice 1 and 2 are fairly close, choice 1 does win. This is 5" screen, 8 RAM, 1.5 CPU, HTC with a price of $199.

The following table shows what the blended model after running the additional scenarios outputs. The blue cells indicate that they are lower, and the whiter they get the higher the percentage is. This is then attached to the output of the counts run under the new model as well.

	[,1]	[,2]	[,3]	[,1]	[,2]	[,3]
[1,]	0.023450019	0.963052099	1.35E-02	10	408	6
[2,]	0.994471706	0.004808963	7.19E-04	422	2	0
[3,]	0.770416717	0.229564815	1.85E-05	327	97	0
[4,]	0.093914394	0.003871144	9.02E-01	40	2	383
[5,]	0.698117362	0.164302424	1.38E-01	296	70	58
[6,]	0.007228307	0.055972863	9.37E-01	3	24	397
[7,]	0.197339687	0.39589328	4.07E-01	84	168	172
[8,]	0.005847887	0.035766447	9.58E-01	2	15	406
[9,]	0.049064627	0.000381338	9.51E-01	21	0	403
[10,]	0.003964469	0.99260701	3.43E-03	2	421	1
[11,]	0.970338752	0.028606718	1.05E-03	411	12	0
[12,]	0.355030285	0.644956928	1.28E-05	151	273	0
[13,]	0.067168462	0.013256532	9.20E-01	28	6	390
[14,]	0.415331409	0.468023506	1.17E-01	176	198	49
[15,]	0.004488916	0.166433129	8.29E-01	2	71	352

	[,1]	[,2]	[,3]	[,1]	[,2]	[,3]
[16,]	0.091666973	0.691521385	2.17E-01	39	293	92
[17,]	0.004715833	0.108458674	8.87E-01	2	46	376
[18,]	0.042993007	0.001256517	9.56E-01	18	1	405
[19,]	0.012085928	0.970483002	1.74E-02	5	411	7
[20,]	0.988858149	0.009349628	1.79E-03	419	4	1
[21,]	0.631841999	0.368120049	3.80E-05	268	156	0
[22,]	0.034960959	0.002675992	9.62E-01	15	1	408
[23,]	0.499572197	0.218327358	2.82E-01	212	93	120
[24,]	0.002585764	0.037181268	9.60E-01	1	16	407
[25,]	0.110443683	0.390743677	4.99E-01	47	166	211
[26,]	0.002696298	0.02908253	9.68E-01	1	12	411
[27,]	0.023007916	0.00031536	9.77E-01	10	0	414
[28,]	0.006596606	0.985552677	7.85E-03	3	418	3
[29,]	0.981270092	0.017262363	1.47E-03	416	7	1
[30,]	0.479833284	0.520142934	2.38E-05	203	221	0
[31,]	0.054071977	0.00790878	9.38E-01	23	3	398
[32,]	0.456416313	0.381159548	1.62E-01	194	162	69
[33,]	0.00380942	0.104671848	8.92E-01	2	44	378
[34,]	0.087120473	0.604910935	3.08E-01	37	256	131
[35,]	0.003297866	0.069809951	9.27E-01	1	30	393
[36,]	0.029195578	0.000785355	9.70E-01	12	0	411

Limitations

As with everything, we should be cognizant of the limitations of this analysis. First and foremost, this analysis is only as good as the data provided. Neverending Marketing Insights probably has given us good data, but we had no control over the collection and may limit the insights given here. Additionally, this data while prescriptive of shopping behaviors will not in anyway be able to alter the choices people make. Further for consideration, while we have been able to decide which model will most likely succeed in the market place, we are unable to give viability to whether that tablet model is price conscious and meets revenue standards. Having a 10” screen, 32 RAM, 2.5 CPU, $199 price may not be feasible.