Christos Kestritsalis

Choice-Based Conjoint Analysis

1. Data Exploration

In order to support decision-making on portable Bluetooth speakers this assignment performs analysis on the choice-based conjoint study data from questionnaire. Specifically, a random sample of size N=400 is used to estimate a choice model, which is later used to estimate market structure and support decision making on segment entry and product introduction.

The dataset consists of 2 tables: choice data and individual data. Dimensionality of choice data is 19200 observations 24 variables and the one of individual data is 400 observations x 30 variables. Choice data consists of index variables for unique index, choice task and alternative of the choice-based conjoint study, dummy variables for choice and none, 4 linear coded variables for attributes battery, weight, price, sound and effect-coded variables for the same 4 attributes. After further analysis, the dataset does not present any null or outlying values.

The individual data consists of unique ID for each respondent, a dummy variable on whether respondent already owns bluetooth speaker, dummy variable whether they intend to buy, numerical variable regarding brand awareness (maximum value of 8), numerical variables regarding 5-point scale on a 1-7 Likert scale on subjective knowledge and product category involvement, relative importance variables for battery, weight, price and sound in percentage %, numerical encoded variables for gender, age group, occupation, education, income along with corresponding text labels for each number, and finally character variable for country of residence.

Taking a closer look into each table's summary, the following observations are made regarding distribution: in choice data, a quarter of respondents have chosen "none" (mean=0.25). Mean value (excluding observations where none=1) of linear-coded battery variable is 11.99 hours, average weight selected 550.05 grams, average price 110.02 €, average sound quality 4.250 stars. These numbers give us an indication of the respondents preferences, taking into account how the questionnaire is structured. Regarding effect-coded variables, it seems "battery_10h" has the only positive average value (mean=0.0005). Similarly, "weight_600g" has only positive mean value at 0.00083, "price_70e" highest mean = 0.00052, "sound_4s" mean = 0.00046, confirming the findings from the linear-coded variables

Regarding individual data, 47.25% of respondents own a bluetooth speaker and 34.5% intend to buy one. On average, respondents are aware of 3.1 brands out of 8. Regarding the subjective knowledge indicators, the highest mean (=4.33, on reversed scale) is found on number 4 ("Compared to most other people, I know less about portable Bluetooth speakers."). Lowest mean was for ("Among my circle of friends, I'm one of the "experts" on portable Bluetooth speakers" - mean=2.91), which reveals a not particularly knowledgeable audience on bluetooth speakers. On product category involvement indicators, highest is on "PII3" (reverse scale): mean =4.407 and lowest "PII_5" mean=4.162, which findings both reveal a lower product involvement and are in line with the findings on subjective knowledge. Regarding stated relative importance variables, from “RelImp_battery”, “RelImp_weight” “RelImp_price”, “RelImp_sound” the mean distribution is respectively: 22.99, 11.47, 28.57, 36.97, which implies that sound is considered the most important attribute on average.

Regarding respondent demographics, 55.25% are male, 41.50% female, and 3.25% non disclosed. Most respondents are between ages of 25-29 (41.25%) followed by ages 18-24 (38.75%), so relatively young, which is what would be expected from a questionnaire answered by a university audience. Occupation variables findings also confirm this with 52% being students and 37.75% employed. Furthermore, 43.75 % are graduates and 33.75% undergraduate and 20.50% are in high school. Most respondents belong in the income category “501-1000 Monthly amount (in €) at the disposal” (32.00%) , followed by “<500” (17.25%), so overall relatively low income. Finally, regarding country of residence, 54.25% have stated “Germany”, followed by 6% Turkey, 5.5% France and 4.5% UK, which is no surprise given the questionnaire is answered from a Berlin university audience.

Summary of the main choice variables can be found in figure 1:

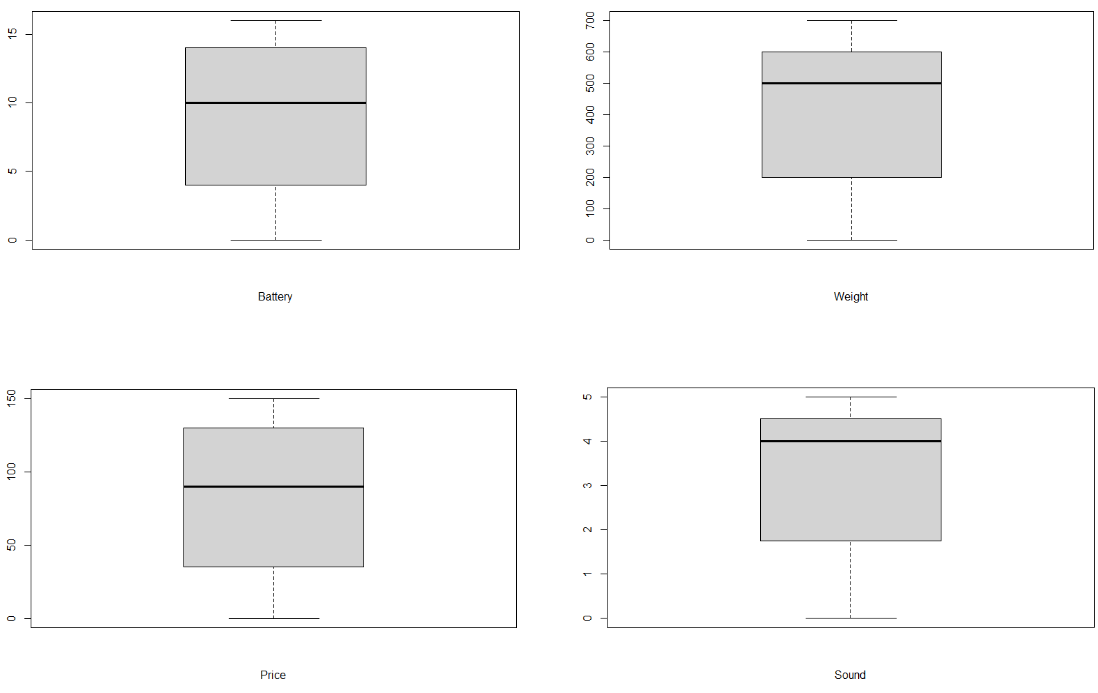

Figure 1: Boxplots of main linear-scaled choice variables

Exploring the relationships between the choice and individual variables in figure 2 it is observed that, as expected, all choice variables are strongly related to each other. The strongest relationship can be observed between sound quality and price, which is an indicator that respondents that consider sound important are also price sensitive. Regarding individual variables it is observed that there is a negative relationship between relative importance of price and relative importance of sound, which implies that respondents that are most interested in sound quality are not so interested about the price and vice-versa. It is also noticed that own and intent to buy variables are negatively correlated, which implies that respondents that already own a bluetooth speaker do not intend to buy a second one. The above mentioned insights that will be useful when deciding on customer segment attractiveness at a later stage.

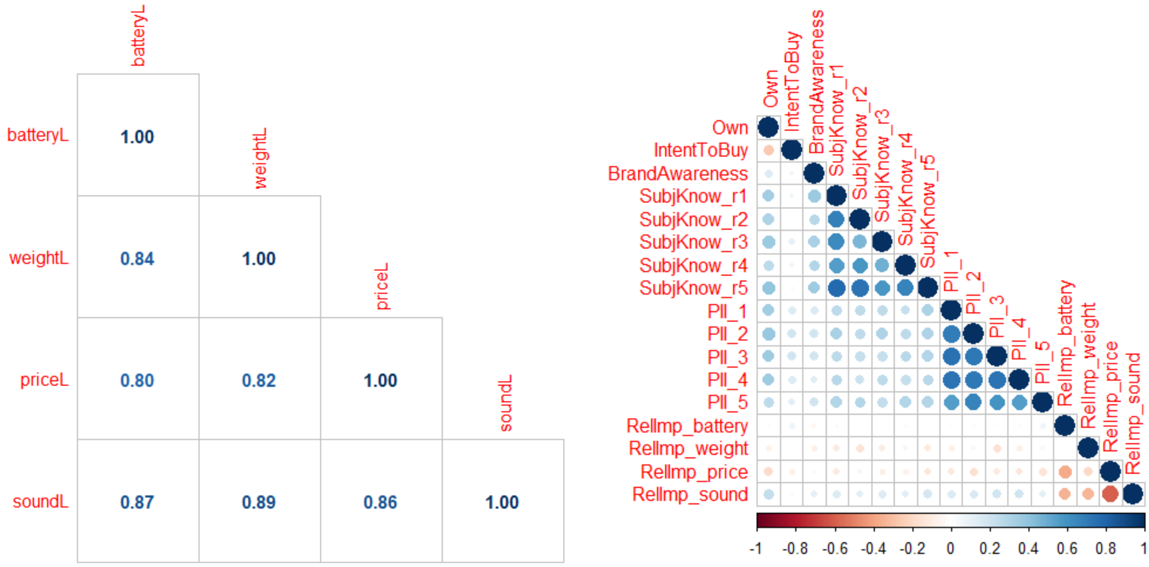

Figure 2: Correlation plot of choice (left) & individual (right) variables

2. Segment Identification and Consumer Preferences

The first step to understanding consumer preferences of the choice-based conjoint study is to estimate a choice model. The choice model will be used to estimate willingness to pay (WTP) and relative importance (RI) measures, on individual and population level, which are important in understanding which attributes drive consumer choice and thus target the customer segments that better represent those attributes.

For the analysis a mixed multinomial logit model (MXL) with linear-coded price and partworth model using effect-coding for rest of attributes is used. MXL allows for unobserved taste heterogeneity. This means that coefficients in the model (preferences) may vary across individuals (also iia no longer holds). Linear effect of price on utility is assumed. For simplicity it is also assumed that sigma is diagonal and error term is iid. In MXL, the utility of individual i derived from product j in choice task t: $U_{ijt} = V_{ijt} + \epsilon_{ijt} + \beta_i x_{ijt}$, with $\beta_i \sim N(\overline{\beta}, \Sigma)$, where V_{ijt} is the vector of product attribute values, β_i the vector of preference estimates, vector of mean preference estimates and Σ covariance matrix of pref. estimates.

Using gmnl in R to estimate the population level MXL model as defined above the following results for different number of draws from the normal distribution are created in table 1:

	R=100	R=500	R=1000		R=100	R=500	R=1000

none	-7.46	-8.58	-8.73 ***	sd.none	3.36	4.36	4.07 ***
priceL	-0.05	-0.05	-0.05 ***	sd.priceL	0.03	0.03	0.03 ***
battery_8h	-1.1	-1.2	-1.22 ***	sd.battery_8h	0.61	0.68	0.76 ***
battery_10h	-0.27	-0.29	-0.28 ***	sd.battery_10h	0.12	0.29	0.32 **
battery_12h	0.13	0.15	0.16 **	sd.battery_12h	0.06	0.07	0.07
battery_14h	0.62	0.64	0.67 ***	sd.battery_14h	0.14	0.01	0.1
weight_400g	0.59	0.63	0.67 ***	sd.weight_400g	0.39	0.53	0.55 ***
weight_500g	0.28	0.32	0.32 ***	sd.weight_500g	0.14	0.27	0.17
weight_600g	-0.18	-0.2	-0.19 ***	sd.weight_600g	0.12	0.17	0.15
sound_3.5s	-2.16	-2.31	-2.32 ***	sd.sound_3.5s	1.15	1.24	1.33 ***
sound_4s	-0.48	-0.53	-0.54 ***	sd.sound_4s	0.5	0.64	0.68 ***
sound_4.5s	0.92	0.97	0.98 ***	sd.sound_4.5s	0.05	0.07	0.14
battery_16h			0.62	AIC	7432.32	7324.75	7304.26
weight_700g			-0.68	BIC	7636.38	7528.80	7508.32
sound_5s			1.73	LL	-3692.16	-3638.38	-3628.13

Table 1: MXL coefficient summary for R number of draws (population level)

It is observed that the model with 1000 draws has highest log-likelihood (LL), thus fits the data best (for the tradeoff of higher computing time). Lower AIC and BIC values further confirm goodness of fit, so consequent analysis will be based on this model.

Choice is distributed almost equally across the three choices (~29% each) with approximately 13% for none. Taking a closer look at the estimated coefficients, it is observed that the attributes that contribute the most to utility (compared to average as the variables are effect-coded) are battery_14h, weight_400g, sound_5s, which is overall what would be expected. The coefficient of battery_16h seems to be slightly smaller than the one of battery_14h, which could indicate that the additional duration in battery life is not valued as highly as one might expect. This could be the case as bluetooth speakers are often used at home, where charging is readily available. So battery duration could contribute to customer utility up to a certain value (14h). Weight seems to contribute to customer utility from 500 g and less, thus indicating a threshold in customer preferences. The coefficient of sound_5s (1.73) is significantly higher than the one for sound_4.5s (0.98), implying that the customers value the increase in sound quality even when it is already high. Finally, the price coefficient (-0.05) is not very high, implying a relatively not so price sensitive sample. Lower statistical significance of standard deviation elements implies there is not a lot of unobserved heterogeneity in the model. Attributes that lead to highest increase in utility are battery_14h, weight_400g, sound_3.5s.

Comparing population-level with individual-level estimates (dark grey), it is observed that individual distribution generally follows population distribution, which allows for generalisation of the model findings.

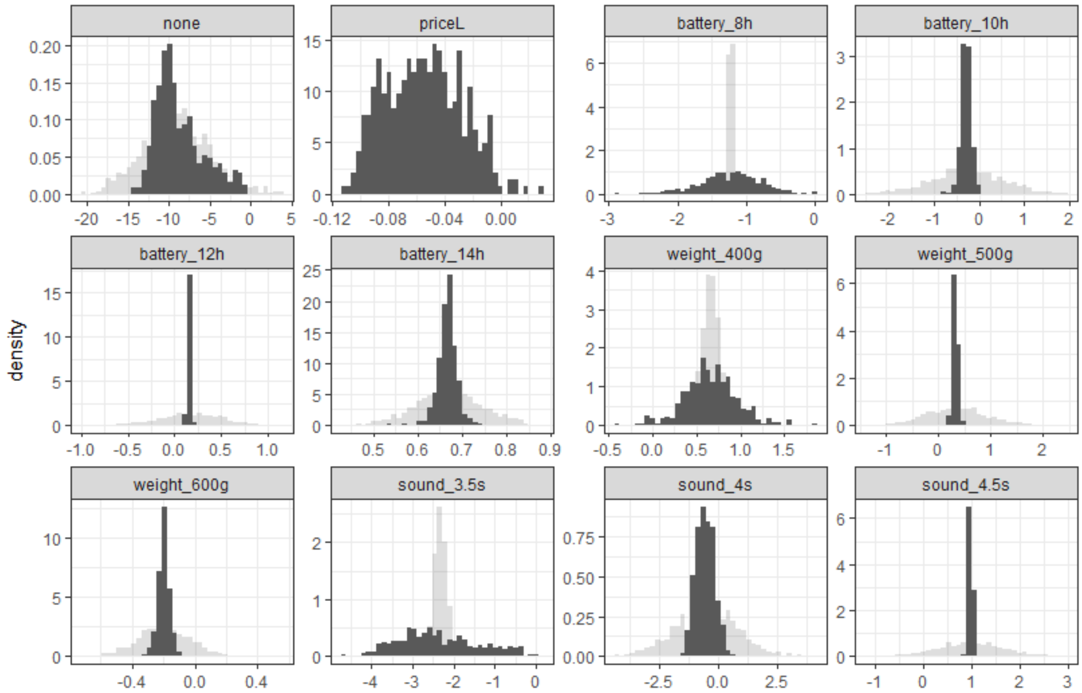

Figure 3: Individual and population-level distribution

In order to further understand which attributes drive choice more it is necessary to look into the RI measure, which is comparable across individuals. RI is defined as: $w_k = \frac{Range_k}{\sum_{k=1}^{K} Range_k}$,

where $Range_k = max(a_k) - min(a_k)$. Result of calculating RI in the data for the different product attributes is shown in Figure 4:

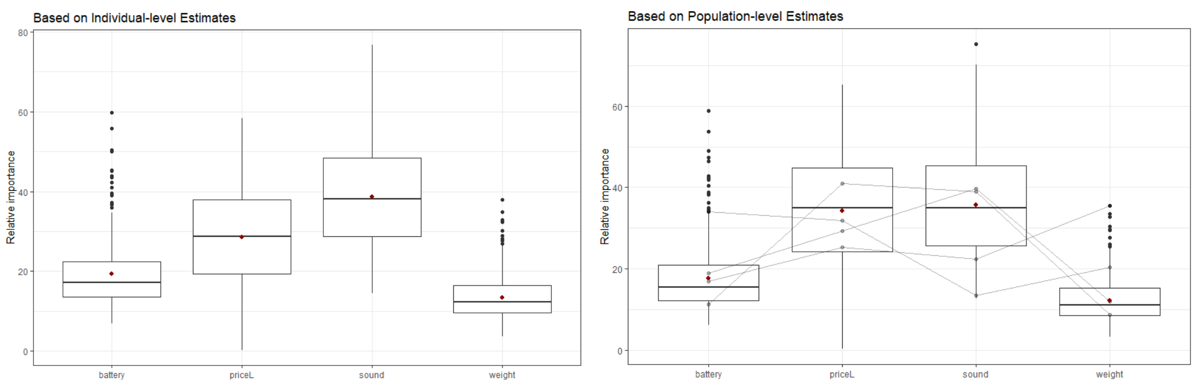

Figure 4: Relative Importance on individual and population level

Overall the sample seems to consider sound quality the most important attribute in Bluetooth speakers, a conclusion which is in line with the findings in part 1. Price matters a lot as expected, weight is valued the least relatively. On the population level price seems to be more important. The population-level estimates are useful to generalise and use on market simulation in the next part.

Next measure which will help identify choice drivers is WTP, which is defined as: Utility in Monetary Terms: $UM_{km} = \frac{a_{km}}{-\beta_{price}}$, $WTP_{km,km'} = UM_{km} - UM_{km'}$ $WTP_{km,km'} = \frac{\beta_{km}}{-\beta_{price}}$, k: index of attribute, m: index for attribute level.

Calculating WTP in R yields the following result shown in Figure 5:

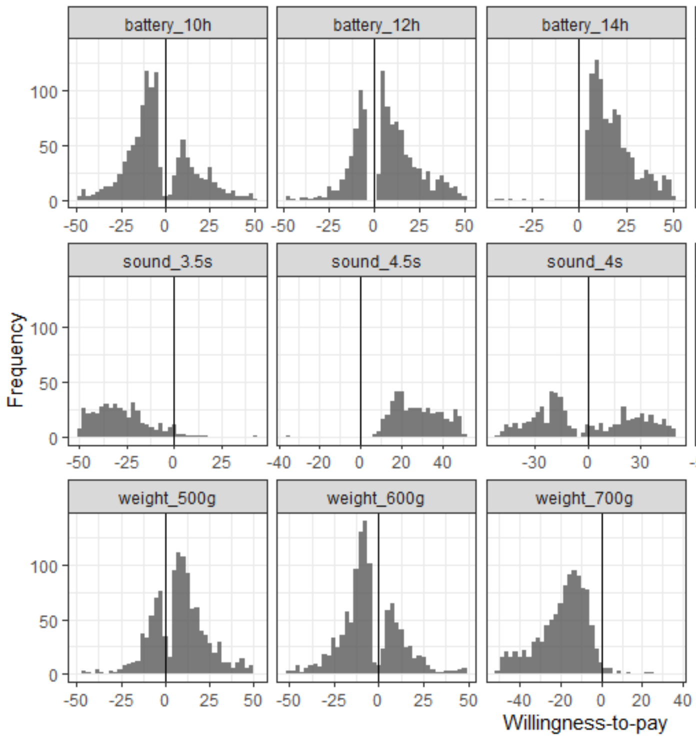

Figure 5: Willingness-To-Pay

Unsurprisingly, battery (battery_16h) seems to be a valuable attribute as well as weight (weight_400g), which implies that those attributes can be most important at driving profitability when targeting particular segments. Sound_4.5s appears important again confirming the findings from MXL and part 1 that sound quality is important for this audience.

It is also interesting to see how perceived relative importance of attributes differs from the one actually measured from the survey. To this end, stated and measured relative importance for attributes is compared and shown in table 2 below (average differences, RI difference = stated RI - measured RI): battery= -2.1975, weight=-6.23, price =-8.3475 , sound = 16.82. Individual RI differences can also be examined per respondent id on the x-axis. High positive difference (>40%) is observed frequently for battery and sound, which implies that respondents tend to overestimate how important battery life is when deciding which bluetooth speaker to purchase. Moreover, a high negative difference (>-40%) occurs frequently for price, implying that respondents tend to underestimate how much price affects decision making when buying bluetooth speakers. Those insights are important to consider when deciding which customer segments to target.

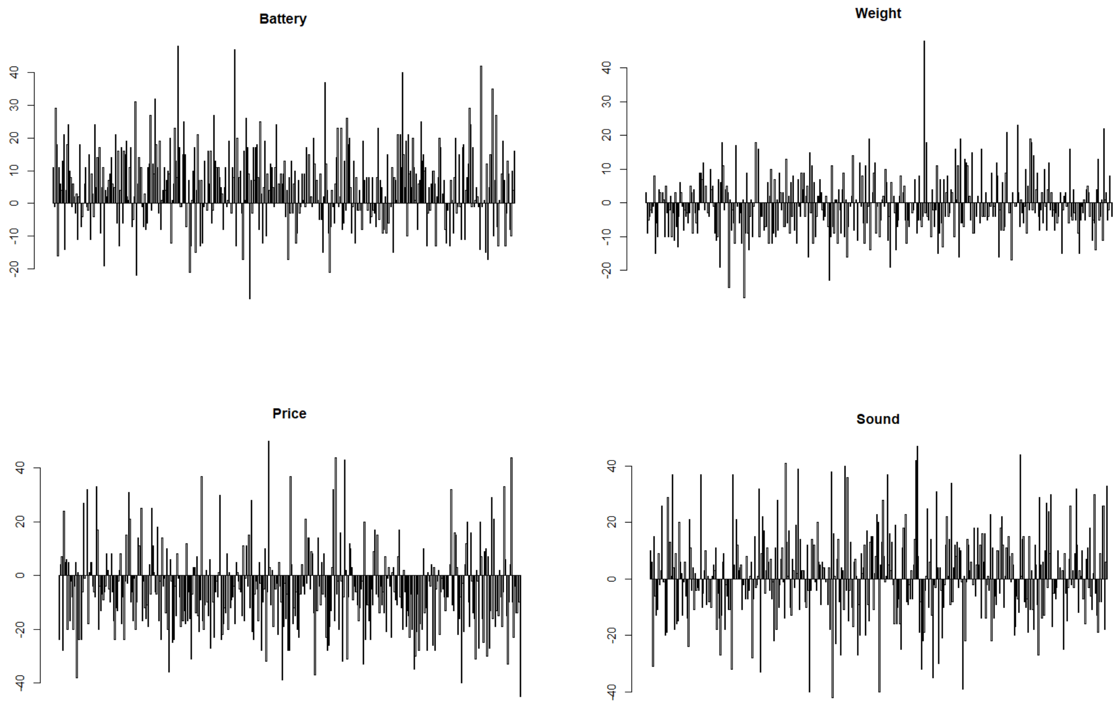

Figure 6: Stated and measured individual differences in relative importance %

Above analysis confirms and reinforces points previously uncovered in the analysis: sound quality and price are the most important attributes on average for the sample. In order to uncover specific patterns in the population and more specifically understand which attributes are most important to which part of the sample, and how those differences are distributed across the population, it is necessary to perform customer segmentation to identify different customer segments. In order to identify segments, measured preferences are input into a clustering algorithm using Ward's method as it is the only one that provides reasonably sized clusters. Scale-independent measures of preferences need to be used as input, so RI is selected for this purpose.

The necessary data is created by merging measured RI with individual data. Euclidean distance is used as a distance measure. After analysing the resulting dendrogram and "cutting" the tree we get the following distribution between the clusters: cluster 1: 114, cluster 2: 130, cluster 3: 116, cluster 4: 40. The optimal number of clusters is determined using the

VRC metric at K = 4. Next cluster means are calculated in order to help identify most important attributes for each segment. Table 4 shows the calculated cluster means:

cluster	battery	priceL	sound	weight	X	Own	IntToBuy	Income
1	16.26	35.68	37.29	10.76	355.12	0.47	0.31	3.41
2	17.24	14.61	55.94	12.21	333.4	0.67	0.43	3.73
3	14.04	53.68	22.79	9.49	337.08	0.36	0.29	3.47
4	26.82	26.72	26.26	20.2	361.37	0.42	0.43	3.83
Br.Aw.	**SubjKn_r2**	**SubjKn_r3**	**SubjKn_r4**	**SubjKn_r5**	**PII_1**	**PII_2**	**PII_3**	**PII_4**
3.19	3.98	2.82	4.27	4.06	4.23	4.33	4.44	4.26
3.59	4.22	3.46	4.9	4.83	4.87	5	4.98	5.02
3.04	3.85	2.97	4.16	3.89	4.04	3.97	4.15	4.11
2.63	3.43	2.58	4.14	3.68	4.12	4.17	4.06	4.12
RI_batt	**RI_weight**	**RI_price**	**RI_sound**	**Gender**	**Age**	**Educ.**	**WTP**	
22.8	10.26	28.66	38.28	1.61	3.05	3.14	1.89	
20.14	10.68	17.59	51.59	1.7	2.79	3.17	0.96	
20.19	9.6	42.32	27.89	1.65	3.03	3.27	1.49	
29.58	18.06	23.05	29.31	1.52	3.23	3.34	1.56	

Table 4: Cluster Means (Ward's method)

According to the data the following segments in the sample can be defined:

- Segment 1: Consider sound important (37.29) but also quite price-sensitive (35.68) , relatively high subjective knowledge and product involvement, high WTP ("Budget-conscious enthusiasts")
- Segment 2: Interested in highest sound quality possibly (55.94), have relatively high income (3.73) are knowledgeable and already own bluetooth speaker, high brand awareness, very high product involvement, relatively young ("Sound aficionados")
- Segment 3: Very price sensitive (53.68), not very interested in sound quality (22.79) or weight (9.49) ("Super-savers")
- Segment 4: Relatively indifferent across attributes (battery, priceL, sound, weight: 26.82, 26.72, 26.26, 20.2 respectively), relatively low product involvement, mostly female ("Casuals")

The segment can be evaluated for their attractiveness to decide whether they are worth targeting as follows:

- Segment 1: N= 114, due to the high WTP and interest in sound quality and product involvement , this would be considered an attractive segment to target (5/7)
- Segment 2: N=130, Biggest segment with a very engaged audience (product involvement, brand awareness, ownership, interest in sound quality) make this segment particularly attractive to target. Some caution is needed in regards to expected profitability due to relatively low WTP (6/7)
- Segment 3: N=116, even though this is the second biggest segment, the high price-sensitivity, low product involvement and not so high WTP make this segment not attractive for targeting (2/7)
- Segment 4: N=40, due to the small size and low product involvement and engagement overall this is not a particularly attractive segment to target (3/7)

3. Market Simulation

To understand how different product offerst affect decision making and affect the profitability of firms in the bluetooth speaker market, market simulation is performed. For this example, it is assumed no other competitors can enter the market. The product offers are as follows:

- Firm 1: Sound quality of 5.0 stars, weight of 600grams, battery life of 12 hours
- Firm 2: Sound quality of 4.0 stars, weight of 400grams, battery life of 16 hours

In order to profit maximising prices, market shares and firm profits a simulation algorithm is created where one firm sets an initial price and the other firm responds by adjusting the price of their product accordingly. Then, the first firm adjusts its price again and so on, until an equilibrium is reached where profitability is maximum and a price change is not beneficial. This calculation of this point is shown in table 5:

Alt.3	ms.3	price_seq.3	price.3	mc.3	profit.3
1	61.75%	100	100	75	6175
2	30.25%	100	95	70	3025
3	08.00%	100	0	0	0

Table 5: Market simulation (No of iterations=4)

After 4 rounds of adjusting prices, the firms reach an equilibrium with firm 1 price=100, market share = 61.75% and profit of 6175. Firm 2 price = 95, market share = 30.25% and profit = 3025. The total profit at this point equals 9200. In this scenario almost no consumer opts for the none option, which is possibly a consequence of the products offered being differentiated enough (attributes are completely different) to cover all possible consumer needs.

In a different market scenario, the introduction of a new product by firm 1 is examined and the effect on profitability, max prices, shares and profits is under examination. The new product has the following attributes: weighs 600 grams, battery life of 16 hours and sound quality of 5.0 stars. After running the the market simulation under the new conditions the following results are found (table 6):

Alt	ms	price_seq	price	mc	profit
1	22.75%	89	85	75	910
2	18.5%	89	85	70	1110
3	5.45%	89	89	75	3052
4	4.25%	89	0	0	0

Table 6: Market simulation with new product from firm 1 (no of iterations=100)

The introduction of the new product by firm 1 seems to lower the equilibrium prices on average (from 100 to 85 and 89 for firm 1, from 95 to 85 for firm 2). The overall profit drops for firm 1 (-2,213) as well as firm 2 (-1,915) and the overall welfare decreases (-4,128). The new product seems to dominate the market (54.5% market share). However this seems to be the result of product "cannibalization" as sales seem to move from one product of firm 1 to the other. This is not unexpected, as the new product of firm 1 is identical to the first one for all but one attribute, and this different attribute (+4 hours of battery) is considerably better, thus making it a clearly better option for most buyers. The consumers still seem to prefer one of the offered options than none, which is reasonable since the new product offered is not significantly different thus the overall options available to the consumers stay relatively the same. Conclusively, firm 1 and the market in total does not benefit from the introduction of the new product under specified market conditions, as it does not seem to be able to claim the market share of firm 2, and there is not a lot of "free" market (since very few consumer chose none) to gain. On the contrary, the introduction of the new product seems to reduce the firm's own profitability, without any measurable gains.