

MKTG 212 - Case 3: Conjoint Analysis

Juan Manubens

University of Pennsylvania

1 Conjoint Study

Gidi, the Marketing Research Man (GMRM), is considering a launch of a new type of raisin into the packaged raisin market. Prior to launch, however, GMRM has decided that a conjoint study of potential consumers might be useful to help in his product design. Even though Gidi knows marketing research, he wants a second opinion. Pretend that you are a marketing research company hired to provide GMRM help with the following issues.

- **[1a] GMRM would like to make sure he has the right attributes and levels of those attributes to include in the conjoint study. Please provide a step-by-step process (with at most five steps) that GMRM can employ to get the right conjoint attributes and their levels.**
1. Analyze the market and competition - Understand the size of the market, target customers, and who the main competitors. Identify the attributes and their levels for the main competitors' product.
 2. Pilot studies (open ended surveys, ratings, ranking) to determine attributes to include, make sure they are unambiguous, actionable and useful for determining choice or preference (on average 6 attributes)
 3. Conduct qualitative research to decide on appropriate attribute levels.
 4. Set an empirical range in product category to determine range of attributes
 5. Determine which format of conjoint analysis you want to use - ratings-based conjoint or choice-based conjoint

- [1b] How would you validate that the attributes and levels you obtained are an “appropriate” set?

1. Confirm determined attributes with the results from pilot studies conducted before.
2. Ensure that the attributes selected are unambiguous, mutually exclusive, and actionable, and through surveys and market research make sure they are not ”must-have” attributes.

After GMRM heard your answer to (1a), he went out and collected product ratings on a 1-10 Likert-scale from consumers utilizing the following attributes and corresponding levels,

Attribute	Level 1	Level 2	Level 3	Level 4
Raisin Chewiness	Low	Medium	High	NA
Raisin Color	White	Grey	Brown	Black
Packaging Size	Small	Large	NA	NA
Free Gift	No	Yes	NA	NA
Raisin Aroma	None	Medium	Heavy	NA
Price Compared to Market Leader	Lower	Same	Higher	NA

where NA indicates not applicable (e.g. Chewiness has only three levels). The data that GMRM collected is sitting in canvas (conjoint data). Please base your answer to the following questions on this data. Note that each attribute is coded numerically. For instance, for Chewiness (Low =1, Medium =2, High =3) and similarly for the other attributes reading left to right in the table above. (Hint: GMRM remembers that one has to create dummy (0/1) variables to represent the different levels of attributes.)

For the following questions, assume a homogeneous market (ignore the segment variable for now).

- [1c] Determine the relative percentage importance of each attribute.

We computed the relative importance of each attribute by summing the ranges of their respective coefficients (including, of course, the 0-coded coefficient too), and computing the fraction each of the attribute-ranges made up of this sum. The relative percentage importance is summarized below:

- Chewiness: 21.85%
- Color: 12.11%
- Size: 22.69%

- Free Gift: 2.31%
 - Aroma: 24.51%
 - Price: 16.52%
- [1d] What Likert rating score would you predict for a raisin product that has Low Chewiness, Grey Raisins, Large Package Size, Free Gift, Medium Aroma, and the Same Price as the Market Leader? You may round your answer to the nearest integer.

Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	6.28996	0.293596	21.42	<.0001
Chew: low	-0.8659	0.243677	-3.55	0.0006
Chew: med	-0.34613	0.24384	-1.42	0.1593
Color: white	0.121095	0.287123	0.42	0.6742
Color: grey	0.214549	0.280197	0.77	0.4459
Color: brown	0.479868	0.269573	1.78	0.0785
Size: small	-0.89923	0.199967	-4.5	<.0001
Gift: no	-0.09165	0.209938	-0.44	0.6635
Aroma: none	0.546762	0.256293	2.13	0.0357
Aroma: medium	0.971456	0.232728	4.17	<.0001
Price: low	0.654813	0.215666	3.04	0.0032
Price: same	0.323676	0.289506	1.12	0.2666
Rating Score: 6.933743				

Multiplying the applicable dummy variables by 1 and adding them to the intercept, we arrive at the following predicted Likert score:

6.93 - rounded to 7

- [1e] What product has the highest predicted rating score

A raisin product with (a) High Chewiness, (b) Brown, (c) Free Gift-Yes, (d) Aroma-Medium, (e) Lower Price, (f) Large Package Size has a rating score of 8.40.

- [1f] Would you necessarily introduce this product, the one from (1e) if you were GMRM? Why or why not?

Not necessarily. There are many possible reasons why it might not make sense to introduce the product. For example, the combination of the low price and the desirable product attributes (which are likely more costly), might seriously reduce the profit margin of the product. It could also be the case that certain of the attributes do not go well together, i.e. the attributes are individually desirable but not desirable as a bundle, or there is a negative interaction effect.

- [1g] Suppose that the market currently contains the following three products:

(A) High Chewiness, Grey Raisins, Small Package Size, No Free Gift, Medium Aroma, and Same Price as Market Leader (B) Low Chewiness, Brown Raisins, Small Package Size, Free Gift, Medium Aroma, and Same Price as Market Leader (C) Medium Chewiness, Black Raisins, Large Package Size, No Free Gift, No Aroma, and Lower Price as Market Leader, and that the predicted market share of product j is proportional to R_j ; that is market share $= R_j / \sum R_j$ where R_j is the predicted Likert rating of product j . What would the predicted market share be if the product described in (1d) were introduced into the market consisting of current products A,B,C?

- Product A has an estimated Likert score of 6.81
- Product B has an estimated Likert score of 6.30
- Product C has an estimated Likert score of 7.05
- Product from 1D has an estimated Likert score 6.93

The sum of the estimated scores is 27.10. Accordingly, the predicted market shares are,

$$M_A = 25.13\%(A), M_B = 23.25\%, M_C = 26.03\%, \quad \text{and} \quad M_{1d} = 25.59\%$$

- [1h] Suppose that the market for raisin products, in total, is 60 Million Dollars and that if GMRM had not hired the market research firm he would have introduced the product in (1d) into the market with the three existing products. Assuming GMRM paid the market research firm \$1,500,000 for their services and that as a result of running the conjoint study, he would introduce the best product, i.e. the product that you found in question (1e) not the one in (1d), was the market research firm worth the money that was paid to them? Justify your answer.

With the superior product in 1e and an estimated Likert score of 8.39, GMRM's raisin product would have a market share of 29.40%.

As such, the net revenue of the two products would be:

- $Product_{1d} : \$60\text{MM} * 25.59\% = \15.354MM
- $Product_{1e} : \$60\text{MM} * 29.40\% - \$1.5\text{MM} = \$16.140\text{MM}$

Thus, the product of question 1e is more attractive by a net difference of \$786,144. In other words, the marketing research paid off.

- **[1i] Do the product attributes (as a whole) provide significant predictive power for the rating scores? Justify your answer.**

The Omnibus F -statistic for significance is less than .0001 (see Appendix 3.1) at a $\alpha = 0.05$ significance level (F -ratio = 5.9763) . Therefore, this mix of product attributes as a whole does provide significant predictive power for the rating scores.

- **[1j] Which, product attributes, if any, have no statistically significant explanatory power for rating scores? State clearly how you arrived at your answer.**

Our results show that the color attribute is not significant at any level, and the free gift variable isn't either (see Appendix 3.1). This can be inferred from the p -values which are statistically significant if below 0.05 (and highlighted in orange/red, accordingly).

The chewiness attribute is only significant on its difference between low chewiness and high chewiness, but it is not significant on its difference between medium chewiness and high chewiness. Similarly, price is not significant on same as market as compared to higher than market price; it is only statistically significant in the difference between lower and higher than market.

The product attributes that have statistically significant explanatory power for all levels are the package size and aroma.

- **[1k] For each segment, determine the percentage importance for each attribute. How do the segments differ? Please be specific.**

(See Appendix 3.2 for calculations)

Segment 1:

- Chewiness: 22.29%
- Color: 7.97%
- Size: 18.52%
- Free Gift: 3.74%
- Aroma: 30.86%
- Price: 16.63%

Segment 2:

- Chewiness: 18.47%
- Color: 15.90%
- Size: 30.75%
- Free Gift: 5.91%
- Aroma: 17.33%
- Price: 11.65%

The segments differ in that Segment 2 deems size by far the most important attribute at 30.75 percent, whereas Segment 1 has size as the third most important attribute, at a mere 18.52 percent. For its part, Segment 1 considers aroma to be the most important attribute, at 30.86 percent, while Segment 2 has aroma as the third most important, at 17.33 percent. Other differences are that Segment 1 considers chewiness to be a somewhat more important (22.29 percent vs. 18.47 percent), color to be considerably less important (7.97 percent vs. 15.90 percent), the presence of a free gift to be about the same in importance (3.74 percent vs. 5.91 percent), and price to be a bit more important (16.63 percent vs. 11.65 percent) than Segment 2 does.

- **[1l] For each segment, what Likert rating score would you predict for a raisin product that has Low Chewiness, Grey Raisins, Large Package Size, Free Gift, Medium Aroma, and the Same Price as the Market Leader?**

For segment 1, this product would have an expected Likert score of 7.094

For segment 2, this product would have an expected Likert score of 6.702

- **[1m] For each segment, which product has the highest predicted rating score?**

For segment 1, the combination which yields the highest predicted rating score is: (a) high chewiness, (b) brown color, (c) large package size, (d) no free gift, (e) medium aroma, and (f) lower price.

For segment 2, the combination which yields the highest predicted rating score is: (a) high chewiness, (b) brown color, (c) large package size, (d) free gift, (e) medium aroma, and (f) lower price.

In other words, regardless of the differences of importance as described in question 1L, only one variable changes in terms of yielding the highest predicted Likert score between the two segments: whether or not a free gift is included.

2 Survey Analysis

In addition to the conjoint study that was done, GMRM also decided that it would be useful to survey a sample of supermarket owners for their opinions on the raisin industry. To conduct this survey, GMRM constructed a list of names containing the largest 200 supermarket owners who had email accounts. He then conducted an email study in which these supermarket owners were sent a questionnaire with 10 items on it regarding the raisin category in their supermarket. Out of the 200 supermarket owners emailed, 120 responded, for a response rate of 60%. The 10 questions asked of the supermarket owners are located at the website above and are labeled “supermarket owner questionnaire”, and their responses are labeled “survey data with item non-response.txt”. Note that some of the questions were not answered by some of the supermarket owners; hence the item non-response (labeled as ‘.’). Based on this information, please answer the following questions.

- [2a] List three potential biases in the way that this study was conducted.

Possible biases:

1. Selection Bias - The sample of the 200 supermarket owners emailed does not reflect the population of supermarket owners. Because only the largest supermarket owners who had email accounts were surveyed, this leaves out small businesses or supermarket owners that don't use email. The circumstances and observations of small and large businesses are often different, which renders this survey skewed toward larger and more technological advanced supermarkets.
2. Response Bias - Perhaps the observations of the 120 respondents do not reflect the observations of the 200 supermarket owners emailed. The circumstances and observations of the businesses that respond are often very different from those that don't. (For example, if the reason that some owners didn't respond to the survey is that their business is too busy, their answers had they responded would probably have significantly differed than the recorded responses.) Indeed, not all the questions were answered by all the survey participants, so some questions' responses only represent a certain group of owners.
3. Anchoring effect from Order and Leading Question bias: The subjects' response may depend in the order in which they are asked the questions in the survey. Similarly, through statements like 'Raisin sales are highly linked to the sales of other fruits.' included in the questionnaire, some respondents may be encouraged to say they agree more than they

would if the statement didn't contain words like "highly." For instance, another way to phrase the question could be "On a scale from 1 to 10, how linked are raisin sales to the sales of other fruits?"

- [2b] GMRM decided to delete all observations with any missing data. Based on the data set with all such observations deleted, which variables (out of Q1-Q9) are significant predictors, at the 10% significance level, of Q10, the overall profitability in the raisin category? Justify your answer.

We cleaned the data using R (code and results in Appendix 3.4 and 3.5 respectively), and ran a Multi-variable OLS regression on the clean data set, using $X : Q_1, \dots, Q_9$ and $Y : Q_{10}$. As we can see, at $\alpha = 0.1$, $X^* : Q_1, Q_3, Q_4, Q_7, Q_8$ and Q_9 are significant predictors for $Y : Q_{10}$

- [2c] Suppose that GMRM decides to use a sophisticated method to fill in the missing values. The data set, filled in by stochastic imputation is located on Canvas ("Sophisticated imputation survey data.txt"). Re-run your model with this data set and compare and contrast your answers to that from questions (2b). What are the main findings? Any differences inform us about how deleting missing data can have an impact on the main findings.

The main findings are that running a least squares regression, all variables except Q5 and Q6 are significant at the 10 percent level. Before, Q2 wasn't a significant variable, and Q1 and Q4 had larger p-values (were less significant). When missing data is deleted, less information is available in the model (because there are fewer observations compared to when all data is present), which means the variables are less predictive.

Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	-0.472662	0.544518	-0.87	0.3873
Q1	-0.279853	0.078109	-3.58	0.0005*
Q2	-0.183641	0.063661	-2.88	0.0047*
Q3	0.1268482	0.065624	1.93	0.0558
Q4	0.3575565	0.071513	5.00	<.0001*
Q5	-0.056478	0.059862	-0.94	0.3475
Q6	0.0544211	0.058184	0.94	0.3517
Q7	0.1617241	0.052052	3.11	0.0024*
Q8	0.3933121	0.059171	6.65	<.0001*
Q9	0.5342735	0.054325	9.83	<.0001*

- [2d] Imagine that for each of the 120 supermarket owners, you also had information from the three questions listed at the end of the questionnaire. List three ways in which that information could be used for target marketing and/or inferential purposes.

(See Appendix 3.3 for the Questionnaire)

Answers to the question "How many different SKUs of raisins does your store carry?" could be the first question to ask when trying to find out whether stores would carry different types of raisins we could launch or would rather keep things simple and only carry the basics. This could be used to figure out which stores to target for distribution because it would tell you which stores are focused on selling raisins and which aren't. It could also be viewed alongside the attribute importances of the consumers that shop in these stores in order to determine if the store is carrying the type of raisins its customer desires or falling short and would benefit from selling more types of our raisins.

Answers to the question "What is the monthly dollar volume of your store?" could be used to figure out if there's a relationship between store size and customer preferences. For example, perhaps customers that shop in smaller stores go there for varieties of raisins not carried in larger chains. The question could also indicate where would be the best exposure if we wanted to test launch a new type of raisin, because it shows where the most people will see it.

Finally, answers to "How much marketing expenditure, per month, do you spend on the raisin category?" can be used to determine the importance stores place on marketing their raisins and whether this affects customer preferences. If, for example, one store spent a lot on raisin marketing, perhaps their customers' preferences would align with the raisin characteristics advertised by the store. In addition, if a store spends a lot of money on raisin marketing, it will likely benefit us to place our products there because they're more likely to be purchased by the consumers who see the advertisements.

3 Appendix

3.1 JMP Outputs for Q1(i,j)

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	11	58.30786	5.30071	5.9763
Error	88	78.05214	0.88696	Prob > F
C. Total	99	136.36000		<.0001*

Lack Of Fit				
Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	80	70.052136	0.87565	0.8757
Pure Error	8	8.000000	1.00000	Prob > F
Total Error	88	78.052136		0.6551
			Max RSq	0.9413

Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	6.2899602	0.293596	21.42	<.0001*
Chew: low	-0.865897	0.243677	-3.55	0.0006*
Chew: med	-0.346133	0.24384	-1.42	0.1593
Color: white	0.1210945	0.287123	0.42	0.6742
Color: grey	0.2145485	0.280197	0.77	0.4459
Color: brown	0.4798681	0.269573	1.78	0.0785
Size: small	-0.899225	0.199967	-4.50	<.0001*
Gift: no	-0.091645	0.209938	-0.44	0.6635
Aroma: none	0.5467621	0.256293	2.13	0.0357*
Aroma: medium	0.9714555	0.232728	4.17	<.0001*
Price: low	0.6548128	0.215666	3.04	0.0032*
Price: same	0.3236755	0.289506	1.12	0.2666

3.2 Calculations for Q1(k)

SEG 1:					
Term	Estimate	Std Error	t Ratio	Prob> t	
Intercept	6.0372324	0.415087	14.54	<.0001	
Chew: low	-0.900949	0.38172	-2.36	0.0235	[Abs] % Importance
Chew: med	-0.513603	0.402945	-1.27	0.2102	0.900949 22.29%
Color: white	0.2331064	0.434023	0.54	0.5943	0.321904 7.97%
Color: grey	0.1444352	0.457271	0.32	0.7538	
Color: brown	0.3219036	0.485653	0.66	0.5114	
Size: small	-0.748257	0.33366	-2.24	0.0308	0.748257 18.52%
Gift: no	0.1512235	0.373222	0.41	0.6876	0.151224 3.74%
Aroma: none	0.491878	0.418169	1.18	0.2468	1.246964 30.86%
Aroma: medium	1.2469641	0.361752	3.45	0.0014	
Price: low	0.6718513	0.352722	1.9	0.0644	0.671851 16.63%
Price: same	0.5658396	0.499033	1.13	0.264	
					Aggregate: 4.041149 100.00%
SEG 2:					
Term	Estimate	Std Error	t Ratio	Prob> t	
Intercept	6.9018905	0.474075	14.56	<.0001	[Abs] % Importance
Chew: low	-0.815026	0.356478	-2.29	0.0279	0.815026 18.47%
Chew: med	-0.206695	0.328564	-0.63	0.5331	
Color: white	-0.435942	0.514865	-0.85	0.4025	0.701372 15.90%
Color: grey	-0.139461	0.413076	-0.34	0.7375	
Color: brown	0.2654299	0.369212	0.72	0.4766	
Size: small	-1.35646	0.285732	-4.75	<.0001	1.35646 30.75%
Gift: no	-0.260542	0.288104	-0.9	0.3715	0.260542 5.91%
Aroma: none	0.560364	0.366177	1.53	0.1342	0.764679 17.33%
Aroma: medium	0.7646785	0.349111	2.19	0.0347	
Price: low	0.5138337	0.313636	1.64	0.1096	0.513834 11.65%
Price: same	-0.010472	0.40217	-0.03	0.9794	
					Aggregate: 4.411912 100.00%

3.3 Supermarket Owner Questionnaire

Please answer each of the following questions on a 1-10 scale where 1 indicates that you disagree completely with the given statement and 10 indicates perfect agreement.

1. Raisin color is a key determinant of raisin category sales.
2. Raisin aroma is a key determinant of raisin category sales.
3. Raisin consumers have differing preferences for varying raisin package sizes.
4. Giving away in-package free gifts is a strong driver of brand sales.
5. Raisin consumers are price sensitive.
6. Raisin chewiness is an important determinant of raisin preference.
7. Raisin sales are highly linked to the sales of other fruits.
8. Marketing for the raisin category can significantly affect people's purchasing habits.
9. I enjoy raisins as part of my daily diet.
10. The raisin category is extremely profitable in my store.

Additional Questions:

What is the monthly dollar volume of your store? How much marketing expenditure, per month, do you spend on the raisin category? How many different SKUs of raisins does your store carry?

3.4 Q2(b,c) R Code for Data-cleaning

```
27. ```{r}
28 q2b <- read.table("Survey Data with item non-response.txt", stringsAsFactors = FALSE, header = TRUE)
29 findNA <- function(i){ length(which(q2b[i,] != ".")) }
30 NAperrow <- unlist(lapply(c(1:120),findNA))
31 NAperrow.delete <- (NAperrow < 10)
32 index <- grep(TRUE,NAperrow.delete)
33 q2b.clean <- q2b[-c(index),]
34 q2b.clean <- as.data.frame(q2b.clean)
35 for (i in 1:10){q2b.clean[,i] <- as.numeric(q2b.clean[,i])}
36 lm.q2b <- lm(Q10 ~ Q1 + Q2 + Q3 + Q4 + Q5 + Q6 + Q7 + Q8 + Q9, data = q2b.clean)
37 lm.summary.q2b <- summary(lm.q2b); lm.summary.q2b
38 ```
```

3.5 Q2(b) R Output

```
Call:
lm(formula = Q10 ~ Q1 + Q2 + Q3 + Q4 + Q5 + Q6 + Q7 + Q8 + Q9,
    data = q2b.clean)

Residuals:
    Min       1Q   Median       3Q      Max
-1.1271 -0.2492  0.1079  0.2762  0.6726

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.89673     0.80571  -2.354  0.024870 *
Q1             -0.29690     0.13373  -2.220  0.033621 *
Q2             -0.13556     0.09311  -1.456  0.155167
Q3              0.19852     0.11079   1.792  0.082630 .
Q4              0.35441     0.10862   3.263  0.002626 **
Q5             -0.01870     0.11089  -0.169  0.867140
Q6             -0.02065     0.09291  -0.222  0.825525
Q7              0.29722     0.09194   3.233  0.002843 **
Q8              0.40254     0.09123   4.413  0.000109 ***
Q9              0.59027     0.07975   7.401  2.03e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4808 on 32 degrees of freedom
Multiple R-squared:  0.8093,    Adjusted R-squared:  0.7557
F-statistic: 15.09 on 9 and 32 DF,  p-value: 3.402e-09
```