Application of latent class models to food product development: a case study

Richard Popper[1], Jeff Kroll[1] and Jay Magidson[2]

[1]Peryam & Kroll Research Corporation, 6323 N. Avondale Ave., Chicago, IL 60631, Richard.Popper@pk-research.com

[2]Statistical Innovations, 375 Concord Ave., Belmont MA 02478, Jay@statisticalinnovations.com

Published in *Sawtooth Software Conference Proceedings*, 2004

INTRODUCTION

Food manufacturers need to understand the taste preferences of their consumers in order to develop successful new products. The existence of consumer segments that differ in systematic ways in their taste preferences can have important implications for product development. Rather than developing a product to please all potential consumers, the manufacturer may decide to optimize the product for the most important segment (perhaps the largest or most profitable). Alternatively, the manufacturer may opt for developing a number of products with different sensory profiles, with the goal of satisfying two or more segments.

A number of analytical methods exist for conducting consumer segmentations, including such traditional methods as hierarchical clustering and k-means clustering. Recently, latent class (LC) models have gained recognition as a method of segmentation with several advantages over traditional methods (see, for example, [1], [2], [3]), but so far almost no applications of these models to food product development have been reported in the literature.

In some types of latent class models (namely, LC regression), segments are formed on the basis of predictive relationships between a dependent variable and a set of independent variables. As a result, segments are comprised of people who have similar regression coefficients. These models can be of particular utility to food developers who need to relate a segment's product preferences to the underlying sensory attributes (taste, texture, etc.) of the products. By including sensory attributes as predictors, LC regression models promise to identify the segments and their sensory drivers in one step and provide highly actionable results.

This paper presents a case study involving the consumer evaluation of crackers. The objectives of the research are a) to determine if consumers can be segmented on the basis of their liking ratings of the crackers b) to estimate and compare a number of LC models, as well as some non-LC alternatives and c) to identify and interpret segments in terms of the sensory attributes that drive liking for that segment (in the case of the regression models).

Many consumer segmentation studies show evidence of individual differences in response style (i.e. consumers differ systematically in how they use the response scale). A question of considerable practical importance is how to deal with such response style differences in the analysis. When such differences are ignored, the resulting segments often display a response level effect – one segment comprising individuals who rate all items using the upper end of the response scale, another segment comprising individuals who consistently use the lower end of the scale. Usually, the differences in response style are of little substantive interest; instead, the researcher is interested in how segments differ in their *relative* ratings of the items involved.

In addition to the objectives mentioned above, this paper also illustrates two approaches for separating an overall response level effect from differences in relative preferences for one cracker over another.

DESCRIPTION OF CASE STUDY

In this case study, consumers (N=157) rated their liking of 15 crackers on a nine-point liking scale that ranged from "Dislike Extremely" to "Like Extremely." Consumers tasted the crackers over the course of three sessions, conducted on separate days. The serving order of the crackers was balanced to account for the effects of day, serving position, and carry-over.

An independent trained sensory panel (N=8) evaluated the same crackers in terms of their sensory attributes (e.g. saltiness, crispness, thickness, etc.). The panel rated the crackers on 18 flavor, 20 texture, and 14 appearance attributes, using a 15-point intensity scale ranging from "low" to "high." These attribute ratings were subsequently reduced using principal component analysis to four appearance, four flavor, and four texture factors. The factors are referred to generically as APP1-4, FLAV1-4, and TEX1-4.

SEGMENTATION ANALYSES

Several types of models were used to obtain segments that differed with respect to their liking of crackers and to relate these differences to the sensory attributes. In all models, the liking data were treated as ordinal. Model fit was assessed using the Bayesian Information Criterion (BIC). Mathematical formulations of the models can be found in the Appendix.

LC Cluster Model. This is the traditional latent class model, which imposes no special structure to distinguish between variation due to differences in overall response level and those due to *relative* differences in the liking of crackers . That is, the latent classes simply represent unordered levels (i.e., a nominal factor). The data layout required for this model is shown in Figure 1. In this layout, each respondent occupies one row (record) and the ratings for the 15 products are arranged in successive columns. No adjustment was made to the data to account for individual differences in overall response level (i.e., the original raw ratings data were analyzed). The analysis was carried out using Latent Gold 3.0.

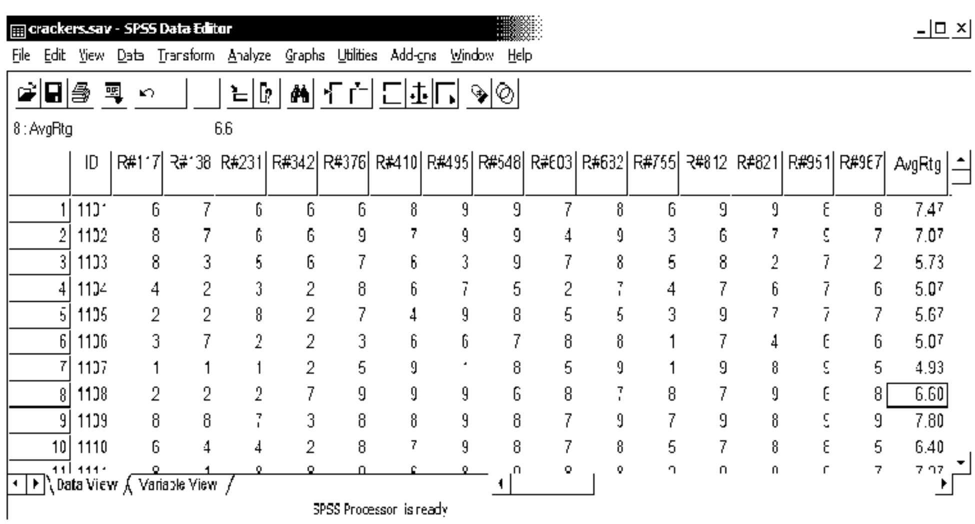


Figure 1: Data Layout for the LC Cluster and LC Factor Models

LC Factor Model. A factor-based version of the latent class model was applied in order to try to "factor out" response level effects. Ordered levels of the first discrete factor (D-Factor #1) were used to model *overall* liking, isolating a response level effect. Additional dichotomous factors (D-Factors #2, 3, etc) were considered in order to identify segments that differ in their *relative* liking of the products. The data layout required for this analysis is the same as for the LC Cluster model. Latent Gold 3.0 was used for the analysis.

Regression Models. Four types of regression models were explored. All of these models used a continuous random intercept to account for individual differences in average liking across all products. Use of a continuous factor (C-Factor) rather than a discrete factor to account for the overall level effect was expected to result in segments that better represented pure *relative* differences in cracker liking. In addition, regression models allow for the possibility of using the attributes as predictors, thus allowing the segments to be defined in terms of differences in the attribute effects. This can not be done using the cluster or factor models.

The data layout required for these analyses is shown in Figure 2. In this layout, there are 15 rows (records) per respondent. The consumer overall liking ratings of the products are contained in the column labeled "Rating", the sensory attribute information in the succeeding columns.

The regression models differ in the predictors. Two use only PRODUCT (15 nominal categories for the 15 crackers) as the sole predictor, while two others use the quantitative attributes as predictors. The models also differ in their approach to modeling the respondent heterogeneity in the effect of the products. The key differences among the four types of models are summarized in Table 1. .

Table 1: Four Types of Regression Models

		Segments Defined Using…	
		Latent Classes	Continuous Factors
Predictors	PRODUCT (15 nominal product levels)	Model 1	Model 2
	Twelve quantitative sensory attributes	Model 3	Model 4

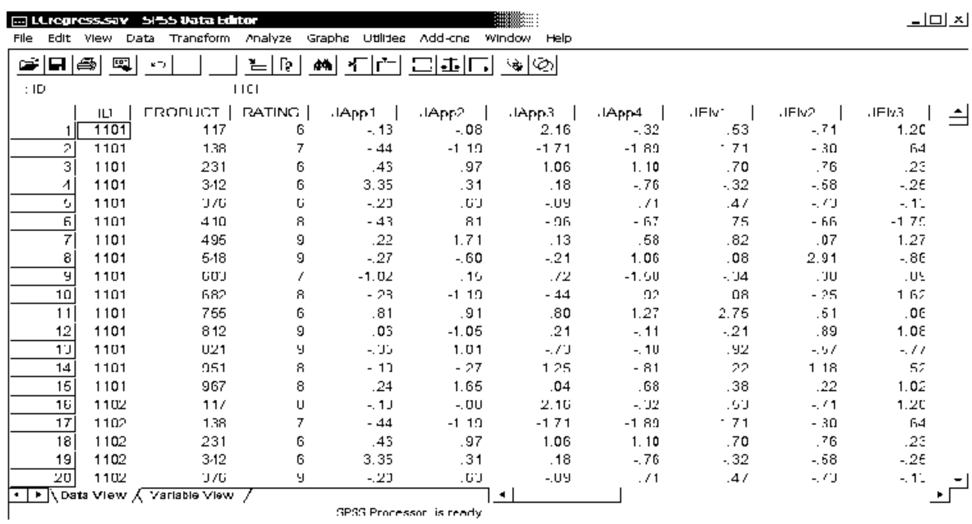
Model 1 used the nominal variable PRODUCT as the sole predictor. It included a class-independent continuous random intercept (C-Factor #1) to capture respondent differences in average liking across all products, and latent classes as a nominal factor to define the segments in terms of the heterogeneity in this PRODUCT effect.

Model 2 also used the nominal variable PRODUCT as the predictor and included a class-independent continuous random intercept (C-Factor #1) to capture response levels effect. However, in contrast to Model 1, one or more additional continuous factors (C-Factors #2, etc.) were considered in order to account for the heterogeneity in the PRODUCT effect.

Model 3 is the same as Model 1, except that it used the 12 sensory attributes as predictors.

Model 4 is the same as Model 2, except that it used the 12 sensory attributes as predictors.

The four factor- regression models (containing two or more factors) were estimated using a forthcoming version of Latent Gold (4.0).



	ID	PRODUCT	RATING	JApp1	JApp2	JApp3	JApp4	JFlv1	JFlv2	JFlv3
1	1101	117	6	-.13	-.08	2.16	-.32	.53	-.71	1.20
2	1101	138	7	-.44	-1.19	-1.71	-1.89	.71	-.30	.64
3	1101	231	6	.43	.97	1.06	1.10	.70	.76	.23
4	1101	342	6	3.35	.31	.18	-.76	-.32	-.58	-.25
5	1101	376	6	-.23	.63	-.09	.71	.47	-.73	-.11
6	1101	410	8	-.43	.81	-.06	-.67	.75	-.66	-1.75
7	1101	495	9	.22	1.71	.13	.58	.82	.07	1.27
8	1101	548	9	-.27	-.60	-.21	1.06	.08	2.91	-.86
9	1101	603	7	-1.02	.15	.72	-1.50	-.34	.30	.05
10	1101	682	8	-.23	-1.19	-.44	.02	.08	-.25	1.62
11	1101	755	6	.81	.91	.80	1.27	2.75	.51	.06
12	1101	812	9	.03	-1.05	.21	-.11	-.21	.89	1.08
13	1101	821	9	-.35	1.01	-.73	-.10	.92	-.57	-.77
14	1101	951	8	-.13	-.27	1.25	-.81	.22	1.18	.52
15	1101	967	8	.24	1.65	.04	.68	.38	.22	1.02
16	1102	117	6	-.13	-.08	2.16	-.32	.53	-.71	1.20
17	1102	138	7	-.44	-1.19	-1.71	-1.89	.71	-.30	.64
18	1102	231	6	.43	.97	1.06	1.10	.70	.76	.23
19	1102	342	6	3.35	.31	.18	-.76	-.32	-.58	-.25
20	1102	376	9	-.23	.63	-.09	.71	.47	-.73	-.11

Figure 2. Data Layout for the Regression Models.

MODEL RESULTS

LC Cluster Model

According to the BIC, a two-cluster solution was a better fit to the data than either a one-cluster or three-cluster solution. The two clusters (segments) were approximately equal in size (53% and 47%). Figure 3 shows each segment's average liking ratings for the products. This figure shows that the two segments are clearly and almost exclusively differentiated by their overall average liking of the crackers. Segment 2 rated almost all products higher than Segment 1. This result is not unexpected, since no attempt was made in the analysis to adjust the data for differences in response level.. The figure also shows *some* relative differences in liking between the two clusters. For example, Segment 2 liked Products #495, #376, #821 and #967 more relative to the other products, whereas Segment 1 liked them less.

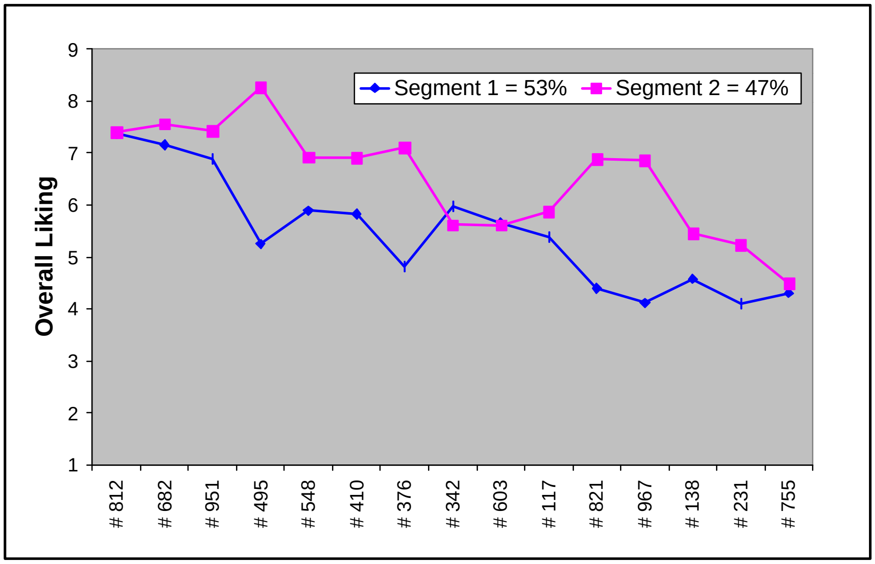


Figure 3. LC Cluster Results.

LC Factor Model

In fitting one-factor (ordered-level) models, the BIC indicated that a factor with four levels resulted in the best fit to the data. Figure 4 shows the average product liking scores for the four levels (classes) on this discrete factor (D-Factor #1). This factor corresponds largely to a level effect. Average liking scores (across products) for the four levels were 4.7 (Level 1), 4.8 (Level 2), 6.7 (Level 3) and 7.3 (Level 4). The interpretation of the factor as a level effect is further supported by the fact that the correlation between individual respondents' scores on this factor and their average liking was 0.87.

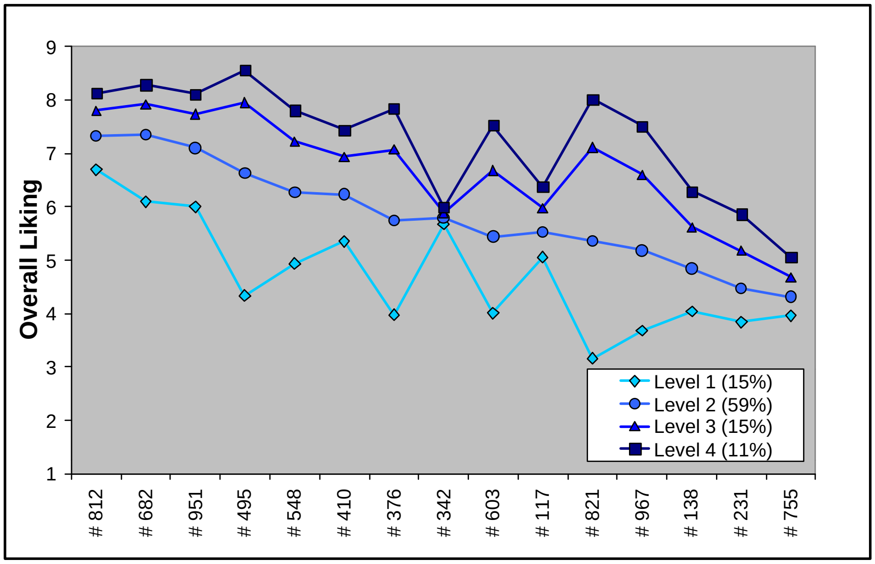


Figure 4. LC Factor Results for D-Factor #1.

According to the BIC, a second dichotomous factor further improved the fit over the one factor model, but additional dichotomous factors did not. The average product liking scores for the two levels (classes) of this second factor (D-Factor #2) are shown in Figure 5. In contrast to the first factor, the classes of D-Factor #2 are differentiated mainly in their relative liking of the products – the average liking across all products was nearly the same for both classes (5.9 and 6.0 for the two levels of D-Factor #2, respectively). Taking each level of D-Factor #2 as a segment, we see that Segment 2 liked Products #495, 376, 821, and #967 more than Segment 1, but liked Products #812, #342, and #603 less. The two factors were assumed to be uncorrelated with each other in the model.

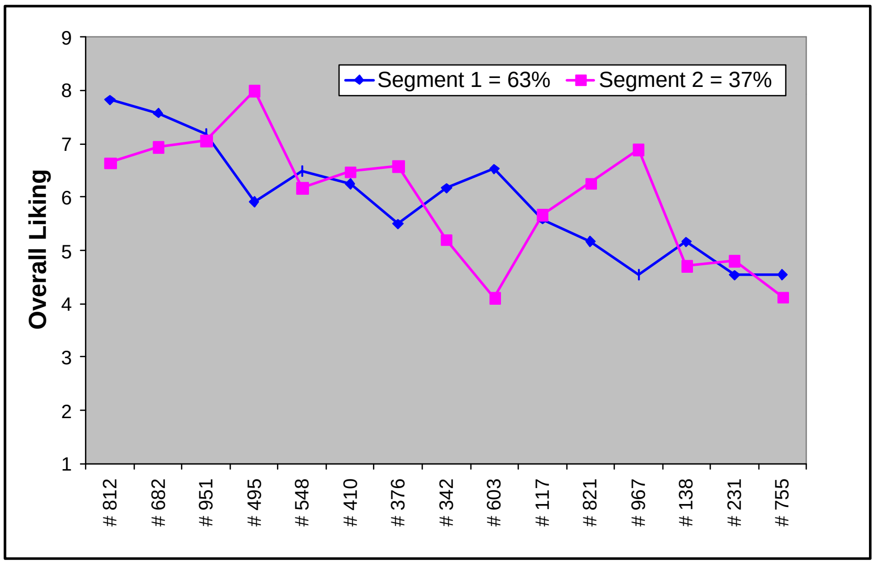


Figure 5. LC Factor Results for D-Factor #2.

Comparison of LC Cluster and LC Factor Models

The most important difference between the two models is that the LC Cluster model confounded relative differences in liking with average response level (one segment rated almost all products higher than the other). The LC Factor model, on the other hand, was able to separate out a response level effect (D-Factor # 1) from an effect that reflected the relative differences in liking (D-Factor #2). Also, the LC Factor model was preferred over the LC Cluster models according to the BIC (BIC = 9,887 for the LC Factor model, compared to 9,926 for the two-class cluster model and 9,930 for the three-class cluster model.)

The same products that differentiated the segments in the LC Cluster model differentiated the segments on D-Factor #2 (#495, 376, 821, and 967). But D-Factor #2 further differentiated the segments in their response to Products #812 and #603, whereas the LC Cluster model showed no difference between these products (see Figure 6). In the LC

Cluster solution, these differences were masked by the differences in response level between the clusters. In addition to these differences between the two models, the two models also differed in the relative sizes of the two segments.

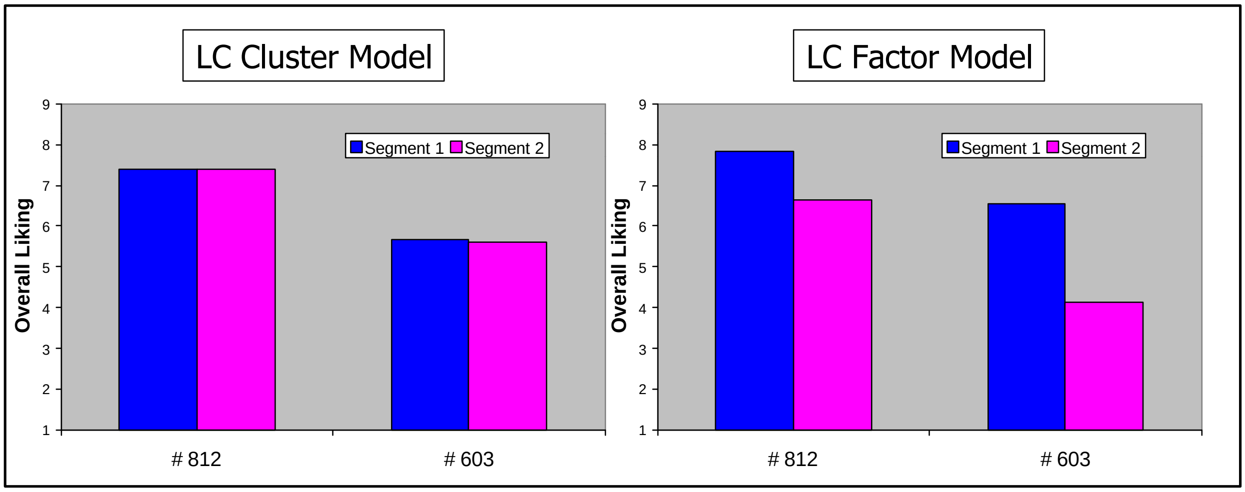


Figure 6. Comparison of LC Cluster and Factor Models for three products.

Regression Models

Model 1

The correlation of the random intercept with respondents' average liking was higher than 0.99 (compared to 0.87 for D-Factor #1), indicating that the random intercept was better able to capture individual differences in average response level than D-Factor #1 in the LC Factor model. A two-class solution provided the best fit to the data, with a model R^2 of 0.39.

Figure 7 shows the average product liking scores for the two segments. Segment 2 liked products #495, #376, #821, and #967 more than Segment 1, but liked #812, #342, #603 less. Liking when averaged across all products was nearly identical for the two segments (5.9 and 6.1 for Segments 1 and 2, respectively).

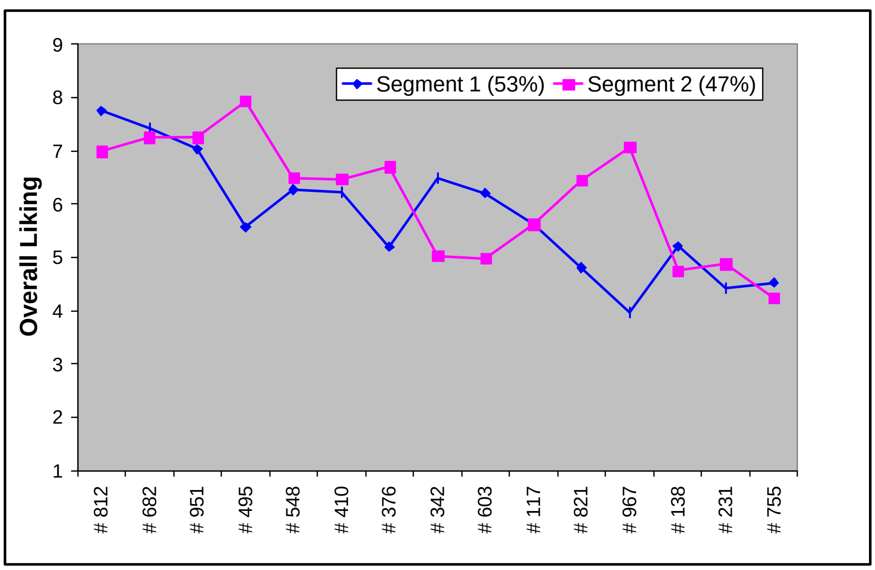


Figure 7. Regression Model 1 Results.

Model 2

As in Regression Model 1, the correlation of the random intercept in Model 2 with average liking was greater than 0.99. The addition of a second C-Factor improved the model fit according to the BIC, but a third C-Factor did not. The model with two C-Factors had an R^2 of 0.41.

Unlike Regression Model 1, Regression Model 2 is not a latent class model and thus does not provide guidance as to the number of underlying segments. For comparison to Model 1, two segments of respondents were formed on the basis of the distribution of scores on C-Factor #2 (using a cutoff point at the mean value of zero). Respondents with scores on C-Factor #2 less than zero were assigned to one segment, and those with scores greater than zero to another segment. Figure 8 shows the average product liking scores by these two groups of respondents, which were approximately equal in size and differentiated on the same products as the segments in Model 1.

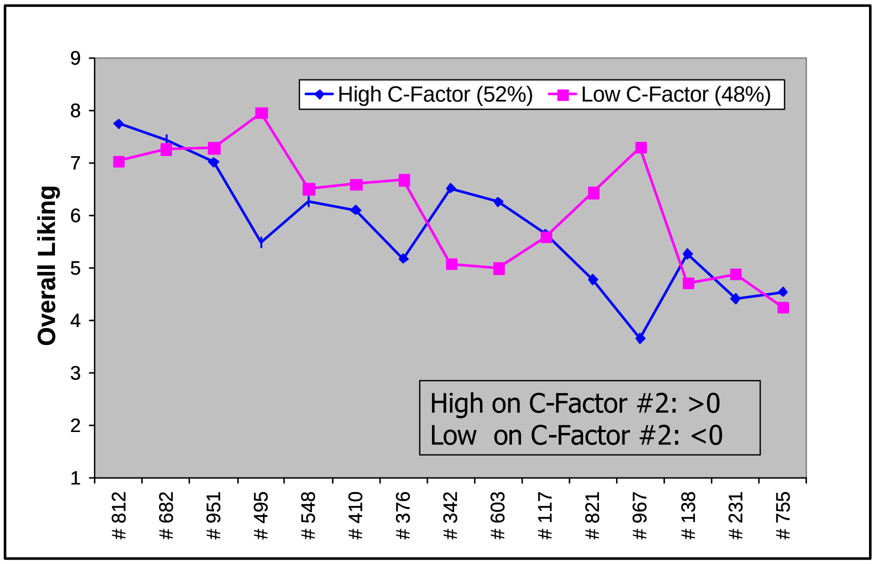


Figure 8. Regression Model 2 Results.

The respondent heterogeneity with respect to the effect of PRODUCT can be assessed more precisely without the formation of segments, by simply examining the magnitude of the interaction effects between the nominal PRODUCT effect and C-Factor #2 (see Table 2). The results represented by the interaction z-statistic are visually represented by the segment means plot in Figure 8. For example, the largest interaction effects occur for Crackers #495 and #967, which are seen in Figure 8 to yield the largest differences between the High and Low C-Factor #2 segments.

Table 2. Regression Model 2: Main Effects and Interaction Effects

Product	Main Effect	Main Effect Z-statistic	Interaction Effect	Interaction Effect Z-statistic
812	0.46	7.57	0.25	3.62
682	0.42	7.66	0.13	2.04
951	0.33	6.56	0.03	0.41
495	0.30	4.41	-0.46	-5.69
548	0.07	1.62	0.06	1.11
410	0.07	1.54	-0.03	-0.54
376	-0.03	-0.71	-0.18	-3.38
342	-0.07	-1.60	0.27	5.00
603	-0.12	-2.97	0.18	3.80
117	-0.13	-3.31	0.10	1.99
821	-0.11	-2.45	-0.23	-3.92
967	-0.13	-2.29	-0.43	-5.50

Note: Z-statistics with absolute values larger than 2 are statistically significant at the .05 level.

Comparison of Regression Models 1 and 2

The two models lead to similar conclusions regarding segment differences and are equally parsimonious (both models used 38 parameters). According to the BIC, Model 2 is preferred to Model 1 (BIC(2)=9,461 vs. BIC(1)=9,487), and the R^2 is also slightly higher. On the other hand, Model 1 provides guidance as to the number of underlying segments, Model 2 does not. Model 2 also does not offer guidance as to where to choose cut-points on the C-Factor and required the use of an arbitrary assignment rule in the formation of segments.

Model 3

The correlation of the random intercept with average liking across products was again very high (>0.99). The BIC was lower for an unrestricted two-class model (BIC=9,535) than for an unrestricted three-class model (BIC=9,560), indicating that the two-class model was preferred. However, a three-class *restricted* model that restricted the third class regression coefficients to zero for all 12 predictors had a slightly better BIC (9,531) than the two-class model. The model R^2 for the three-class restricted regression model was 0.39, the same as for Model 1 (which used the nominal PRODUCT variable as the predictor).

The interpretation of the third class is that it consists of individuals whose liking does not depend on the levels of the 12 sensory attributes. This segment was small (8%), compared to the size of the other two segments (42% and 50% for Segments 1 and 2, respectively).

Figure 9 shows the average product liking scores for the three-class restricted model. The plot of regression coefficients in Figure 10 provides a visual display of the extent of the segment differences in attribute preferences. Segment 2 prefers products high in APP2 and low in APP3. Segment 1 was not highly influenced by these two characteristics, but preferred crackers high in APP1. Both clusters agree that they prefer crackers that are high in FLAV1-3, low in FLAV4, low in TEX1 and high in TEX2-3.

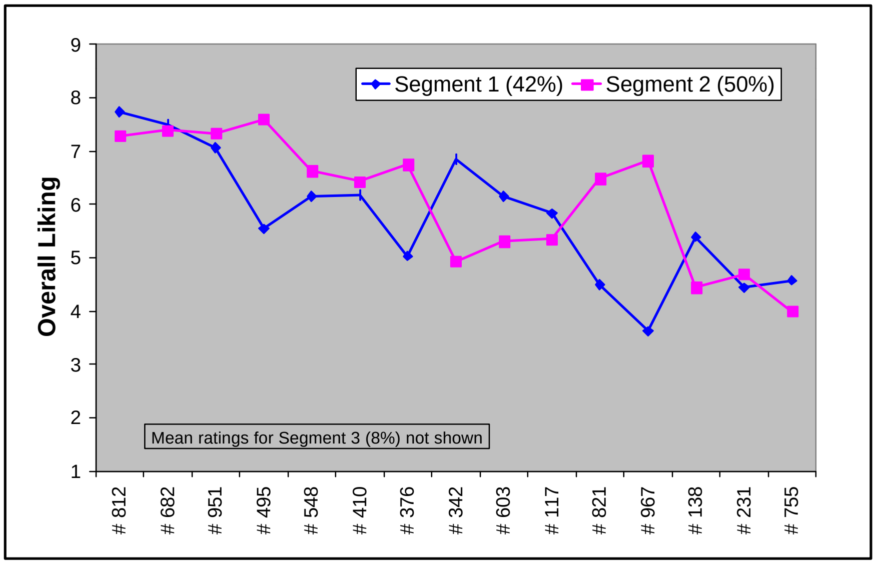


Figure 9. Regression Model 3 Results.

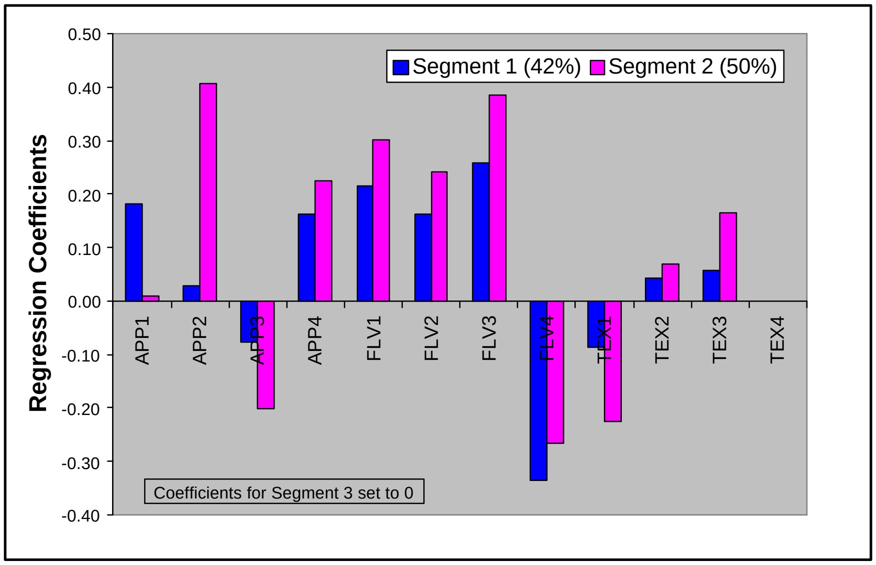


Figure 10. Regression Coefficients for Regression Model 3.

Model 4

The random intercept was again highly correlated with individual respondents' average liking (>0.99). As was the case with Model 2, the addition of a second C-Factor improved the fit (according to the BIC), but a third factor did not. The model R^2 was 0.38, slightly lower than for Model 2, which used the nominal PRODUCT variable as the predictor instead of the 12 sensory attributes.

Two segments were formed on the basis of C-Factor #2 by assigning respondents with scores less than zero to one group and respondents with scores higher than zero to the other. Figure 11 shows the average product liking scores for the two groups and indicates a pattern similar to that for the latent class model using the sensory predictors (Model 3).

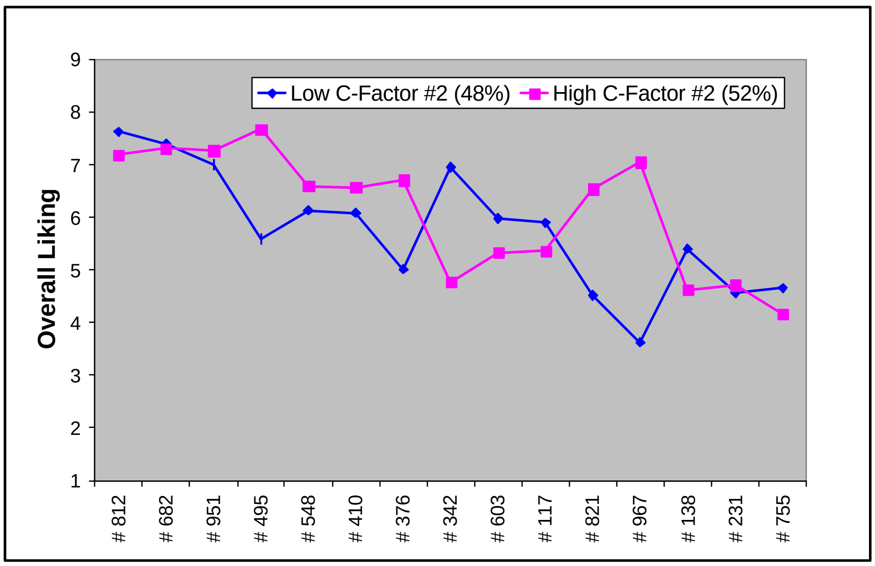


Figure 11. Regression Model 4 Results.

Table 3 provides a statistical assessment of the sensory predictors in terms of main effects and interactions with C-Factor #2. A significant *main effect* indicates that higher levels for the associated attribute significantly increase (or decrease) the rating for all respondents. All sensory predictors except for TEX4 had significant *main effects* on overall liking, but only one attribute (APP2) yielded a significant *interaction effect* ($p<0.05$). For two other predictors, APP1 and APP3, the interaction with C-Factor #2 approached statistical significance (p 0.10).

The overall effect of an attribute is given by the main effect plus the C-Factor #2 score multiplied by the interaction effect for a given respondent. For example, for a respondent scoring one standard deviation *above* than the mean on C-Factor #2 (C-Factor #2 score = +1), the overall effect of appearance attribute APP2 can be computed as 0.21 +1*0.21 = 0.4. Similarly, for a respondent scoring one standard deviation *below* than the mean on C-Factor #2 (C-Factor #2 score = -1), the overall effect of APP2 is 0.21 -1* 0.21 = 0. Thus, respondents scoring higher on C-Factor #2 are more favorably affected by APP2, whereas those scoring low on the factor (say C-Factor #2 = -1) tend to be neutral to APP2.

Note that in addition to the random intercept, only a single additional C-Factor was required to account for respondent heterogeneity. Thus, this model is substantially more parsimonious than a traditional HB (Hierarchical Bayes) type model, which would be equivalent to including 12 additional C-Factors, one for each attribute.

Table 3. Regression Model 4: Main Effects and Interaction Effects.

Product	Main Effect	p-value	Interaction Effect	p-value
APP1	0.08	0.02	-0.08	0.10
APP2	0.21	0.0001	0.21	0.0008
APP3	-0.13	0.0001	-0.07	0.06
APP4	0.19	0.0000	0.02	0.67
FLV1	0.24	0.0000	0.04	0.18
FLV2	0.19	0.0000	0.03	0.44
FLV3	0.29	0.0000	0.07	0.31
FLV4	-0.27	0.0000	0.04	0.34
TEX1	-0.14	0.0180	-0.06	0.45
TEX2	0.05	0.0026	0.01	0.63
TEX3	0.10	0.0047	0.05	0.27
TEX4	0.00	0.95	0.01	0.85

<u>Comparison of Regression Models 3 and 4</u>

The two models are similar in R^2 and are similar in parsimony (Model 3 used 33 parameters compared to 34 in Model 4). The BIC was somewhat better for Model 4 (BIC(3)=9,531 vs. BIC(4)=9,525). The models differ in their use of a discrete (Model 3) vs. continuous (Model 4) measure of respondent heterogeneity. A weakness of Model 4 is that the continuous factor (C-Factor #2) does not yield clearly differentiated segments. Figure 12 shows that the distribution of C-Factor #2 scores is normal, with many scores around zero. This makes the choice of a cut-point for formation of segments arbitrary, as was the case for Model 2. Model 3 provides clear segment differentiation. The strength of that segmentation is further indicated by the posterior membership probabilities (see Figure 13), which show that the average estimated probability of cluster membership is high (>0.8) for the two large segments (Segments 1 and 2), and above 0.5 for Segment 3.

While Model 4 does not provide guidance with respect to segment formation, the segments identified in Model 3 have a close correspondence with C-Factor #2 scores estimated in Model 4. Figure 14 shows that respondents assigned to Segment 1 in Model 3 score high on C-Factor #2, those assigned to Segment 2 score low on C-Factor #2, and those assigned to Segment 3 score around zero on C-Factor #2. Thus, it should not be surprising that Regression Models 3 and 4 provide very similar inferences about which attributes are most important (significant main effects) and which are most involved in respondent heterogeneity (significant interaction effects).

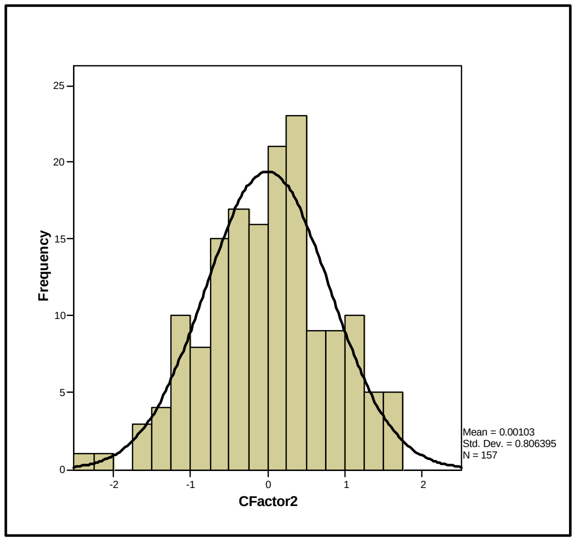


Figure 12. Distribution of C-Factor #2 Scores in Regression Model 4

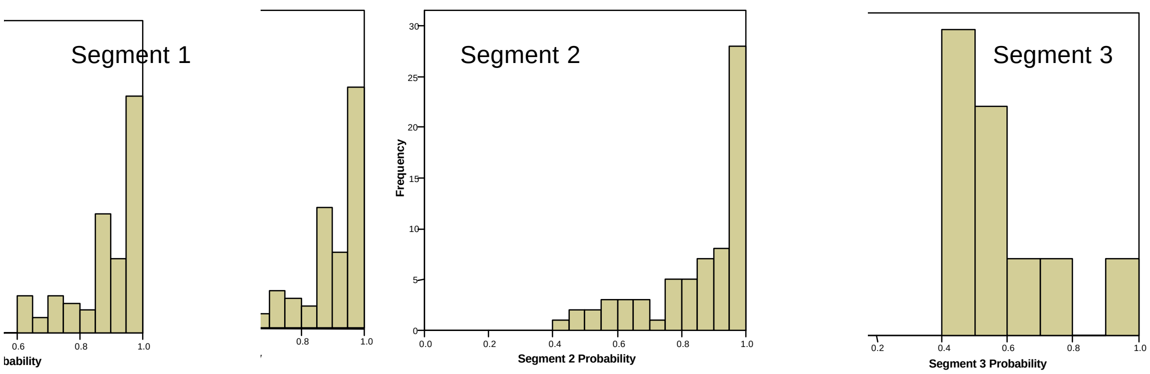


Figure 13. Posterior Membership Probabilities for Segments Determined in Regression Model 3.

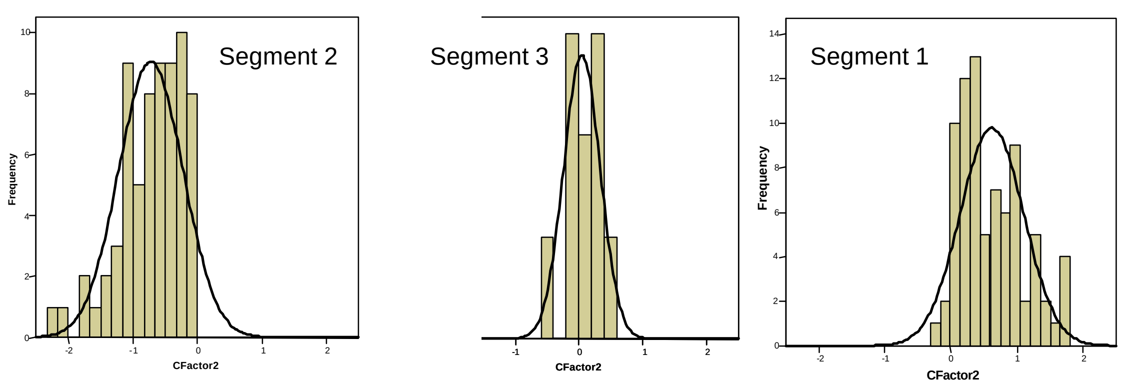


Figure 14. Distribution of C-Factor #2 Scores (Regression Model 4) for Segments Determined in Regression Model 3.

GENERAL DISCUSSION

Four alternative approaches were presented for segmenting consumers on the basis of their overall liking ratings without consideration of the products' sensory attributes: the LC Cluster model, the LC Factor model, and the two regression models (Models 1 and 2) that used the nominal product variable as the predictor. All models provided evidence of the existence of segment differences in consumers' liking ratings. While some products appealed to everybody, other products appealed much more to one segment than another.

The LC Cluster model identified two segments that differed in their average level of liking and showed some relative differences in liking as well. The LC Factor model used two factors to isolate segment differences associated with average response level from differences in relative product liking. The segments given by the second factor provided a clearer picture of relative product differences than did the LC Cluster solution.

The regression models with a random intercept yielded segments that provided an even clearer picture of the relative differences than that given by the LC Factor model, since they allowed for a cleaner separation of the response level effect. The correlation of the random intercept was in excess of 0.99 for both the LC Regression model (Model 1) and the non-LC regression model (Model 2). Both correlations exceeded the correlation of 0.87 between the first factor of the factor model and average liking.

Including a random intercept is conceptually similar to mean-centering each respondents' liking ratings. A LC Cluster analysis of the mean-centered liking data would yield similar results to those obtained with LC Regression Model 1. However, there are two reasons to prefer the regression approach in general. With the regression approach, it is possible to maintain the ordinal discrete metric of the liking data. Subtracting an individual's mean from each response distorts the original discrete distribution by transforming it into a continuous scale that has a very complicated distribution. Secondly, in studies where a respondent only evaluates a subset of products, mean-centering is not appropriate since it ignores the incomplete structure of the data. Thus, the regression approach provides an attractive model-based alternative for removing the response level effect.

The use of a continuous vs. discrete random product effect in Regression Model 2 (compared to Regression Model 1) led to a slightly improved model fit, but at a price. The non-LC regression approach does not determine the cut-points to use to identify segments. An arbitrary choice of cut-points is likely to lead to substantial misclassification, in the event that distinct segments do in fact exist.

Replacing the nominal PRODUCT predictor with the twelve quantitative appearance, flavor and texture attributes made it possible to relate liking directly to these attributes. This allowed for the identification of both positive and negative drivers of liking. Segments reacted similarly to the variations in flavor and texture, but differed with regard to how they reacted to the products' appearance. Based on such insights, product developers can proceed to optimize products for each of the identified segments.

Replacing the nominal PRODUCT variable with the sensory predictors did not lead to any substantial loss in model fit. The R^2 for Model 3 was the same as for Model 1, and the fit for Model 4 only slightly below that of Model 2 (0.39 vs. 0.41).

The non-LC regression models (Model 2 and 4) can be compared to Hierarchical Bayes (HB) models. The HB models are equivalent to regression models containing one continuous factor (C-Factor) for each (non-redundant) predictor regression coefficient plus one additional C-Factor for the intercept (15 C-Factors for Model 2 and 13 for Model 4). Since in the analysis of these models the BIC did not support the use of more than two C-Factors, Models 2 and 4 offer much more parsimonious ways to account for continuous heterogeneity than HB. HB would likely overfit these data by a substantial margin, and at least some differences in liking suggested by an HB model would therefore not validate in the general population.

Since no group of quantitative predictors is going to be able to exceed the strength of prediction of the nominal PRODUCT variable with its fourteen degrees of freedom, Models 1 and 2 provide an upper bound on the R^2 for Models 3 and 4, respectively. A comparison of the R^2 of Models 3 and 1 (and Models 4 and 2) provides an assessment how well the sensory predictors perform relative to the maximally achievable prediction. In this case study, the twelve sensory attributes captured almost all the information contained in the nominal PRODUCT variable that was relevant to the prediction of overall liking. The inclusion of additional predictors (for example, quadratic terms to model a curvilinear relationship between liking and sensory attributes) is therefore not indicated, although in other applications cross-product terms or quadratic terms could be very important in improving model fit or optimizing the attribute levels in new products.

The use of restrictions in LC Regression Model 3 improved the fit over an unrestricted model and allowed for the identification of a third segment, one whose overall liking of the products was not influenced by the sensory attributes. While this group was small, in certain applications such a group could be of substantive interest and warrant follow-up. If nothing else, the members of such a group can be excluded as outliers.

Regression Models 3 and 4 differed in their use of a discrete vs. continuous measure of respondent heterogeneity. The models yielded similar fit statistics and conclusions about the attribute effects and heterogeneity. Since Model 3 yielded clear segments more easily than Model 4 and the fit was almost the same, Model 3 was preferable for our purposes.

Among all the models tested (cluster, factor and the four regression models), Regression Model 3 yielded the most insight into the consumer liking of the products: the model provided clear segment differentiation, it isolated the response level effects from the sensory attribute effects that were of more substantive interest, and it identified the sensory drivers of liking for each segment.

ACKNOWLEDGMENT

The authors wish to thank The Kellogg Company for providing the data for this case study.

REFERENCES

[1] Magidson, J. and Vermunt, J.K. (2002). Latent class modeling as a probabilistic extension of K-means clustering. Quirk's Marketing Research Review, March 2002, 20 & 77-80.

[2] Magidson, J. and Vermunt, J.K. (2002). Latent class models for clustering: A comparison with K-means. Canadian Journal of Marketing Research, 20, 36-43.

[3] Vermunt, J.K. and Magidson, J. Latent class cluster analysis. (2002) In: J. Hagenaars and A. McCutcheon (eds.). Applied latent class models, 89-106. Cambridge University Press.

[4] Skrondal, A. & Rabe-Hesketh, S. (2004). Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models. London: Chapman & Hall/CRC.

APPENDIX

1) Latent Class Cluster Model

If the ratings were treated as *continuous*, the normal error assumption yields a linear model. In this case, the predicted rating for cracker t for latent class x, is expressed as:

$$E(Y_{i.t}) = \alpha_t + \beta_{xt}, \qquad t = 1,2,\ldots,15 \qquad (1)$$

and effect coding is used for parameter identification:

$$\sum_{x=1}^{K} \beta_{xt} = 0$$

so the intercept α_t corresponds to an overall (average) rating effect for cracker t over all cases and for K latent class segments, the β_{xt} effects x = 1,2,…,K represent the segment differences.

The prediction for individual i, is generated by weighting these class level predictions by the corresponding posterior membership probabilities obtained for that individual.

Since the ratings are *not* continuous but discrete (Y=m; m=1,2,…,9), we instead assume the following adjacent category logit model which treats the ratings as *ordinal*. For K = 2 classes, we have:

$$logit(Y_{im.t}) = \alpha_{tm} + \beta_{xt}, \qquad t = 1,2,\ldots,15; \quad m = 2, 3,\ldots, 9; \quad x=1,2$$

where

$logit(Y_{im.t})$ is the adjacent category logit associated with rating Y = m (vs. m-1) for cracker *t*,

and effect coding is used for parameter identification:

$$\sum_{m=1}^{M=9} \alpha_{tm} = \sum_{x=1}^{K} \beta_{xt} = 0$$

so, similar to the continuous case, the intercepts capture average response levels.

Thus, for segment x :

α_{tm} $m = 1,2,\ldots,9$ denote the intercepts associated with product *t*

and β_{xt}, x = 1,2 represents the effect of product *t*

2) LC Factor model

For ordinal ratings, the model is a restricted LC cluster model, where the segment differences are characterized by 2 independent discrete factors (D-Factors). The first D-Factor consists of 4 ordered levels, the second is dichotomous:

$logit(Y_{im.t}) = \alpha_{tm} + \beta_{xt}$ t = 1,2,…,15;   m = 2, 3,…, 9;

where:

$\beta_{xt} = \beta_{t1} X_1 + \beta_{t2} X_2$

X_1 = 0, 1/3, 2/3, 1   for levels x_1 = 1,2,3 and 4 respectively of D-Factor X_1

X_2 = 0, 1   for levels x_2 = 1, 2 for dichotomous D-Factor X_2

where D-Factors X_1 and X_2 are stochastically independent of each other

3) Ordinal Regression Models

For each of the following regression models, the ratings are treated as ordinal. If instead they were treated as continuous the models would be linear regression models in which case the subscript m would not appear on the intercept and logit($Y_{im.t}$) would be replaced by E($Y_{i.t}$) as in the case of the LC cluster model. In addition, each regression model contains one or two continuous factors (C-Factors). For further details on the use of C-Factors see [4].

Model 1: LC Ordinal Regression with Random Intercept and Discrete Random PRODUCT Effects

$$logit(Y_{im.t}) = \alpha_{im} + \beta_{xt}$$
$$\alpha_{im} = \alpha_m + \lambda F_i$$

Thus,

$$E(\alpha_{im}) = \alpha_m$$
$$V(\alpha_{im}) = \lambda^2$$

where m=2,3,…,9 and V denotes the variance.

logit($Y_{j.k}$) is the adjacent category logit associated with rating Y = m (vs. m-1) for product *t*

F_i is the C-Factor score for the *i*th respondent

β_{xt} is the effect of the t^{th} product for class *x*
and effect coding is used for parameter identification:

$$\sum_{m=1}^{M=9} \alpha_{im} = \sum_{t=1}^{T=15} \beta_{xt} = 0$$ for each segment x = 1,2,…,K.

Model 2: LC Ordinal Regression with Random Intercept and Continuous Random PRODUCT Effects

$$logit(Y_{im.t}) = \alpha_{im} + \beta_{it}$$
$$\alpha_{im} = \alpha_m + \lambda_{10} F_{i1} + \lambda_{20} F_{i2}$$
$$\beta_{it} = \beta_{t0} + \lambda_{2t} F_{i2}$$

Thus,

$$E(\alpha_{im}) = \alpha_m$$
$$V(\alpha_{im}) = \lambda_{10}{}^2 + \lambda_{20}{}^2$$
$$E(\beta_{it}) = \beta_{t0}$$
$$V(\beta_{it}) = \lambda_{2t}{}^2$$

where: $logit(Y_{im.t})$ is the adjacent category logit for product t
C-Factor score F_{i1} is associated with the intercept
C-Factor score F_{i2} is associated with the T product effects

and $(F_{i1}, F_{i2}) \sim BVN(0, I)$

Model 3: LC Ordinal Regression with Random Intercept and Discrete Random Product Attribute Effects

$$logit(Y_{im.t}) = \alpha_{im} + \beta_{x1} Z_1 + \beta_{x2} Z_2 + ... + \beta_{xT} Z_Q$$
$$\alpha_{im} = \alpha_m + \lambda F_i$$

Thus,

$$E(\alpha_{im}) = \alpha_m$$
$$V(\alpha_{im}) = \lambda^2$$

where:

$logit(Y_{im.t})$ is the adjacent category logit for product t with attributes $Z_1, Z_2, \ldots, Z_Q$

β_{xq} is the effect of the qth attribute for class x

and C-Factor *score* F_{i1} is associated with the intercept

Model 4: Ordinal Regression with Random Intercept and Continuous Random Product Attribute Effects

$$logit(Y_{im.t}) = \alpha_{im} + \beta_{i1}Z_1 + \beta_{i2}Z_2 + ... + \beta_{iQ}Z_Q$$
$$\alpha_{im} = \alpha_m + \lambda_{10}F_{i1} + \lambda_{20}F_{i2}$$
$$\beta_{iq} = \beta_{q0} + \lambda_{2q}F_{i2}$$

$$E(\alpha_{im}) = \alpha_m$$
$$V(\alpha_{im}) = \lambda_{10}^{\ 2} + \lambda_{20}^{\ 2}$$
$$E(\beta_{iq}) = \beta_{q0}$$
$$V(\beta_{iq}) = \lambda_{2q}^{\ 2}$$

where: $logit(Y_{im.t})$ is the adjacent category logit for product t with attributes $Z_1, Z_2, \ldots, Z_Q$

C-Factor F_{i1} is associated with the intercept

C-Factor F_{i2} is associated with the Q product attribute effects

and $(F_{i1}, F_{i2}) \sim BVN(0, I)$