Use of latent class regression models with a random intercept to remove the effects of the overall response rating level

Jay Magidson[1]

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

Jeroen .K. Vermunt

Department of Methodology and Statistics, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands

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, each satisfying at least one of the segments.

In latent class (LC) regression models (Wedel and Kamakura, 1998), 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. A problem with the application of LC regression analysis to this type of rating data sets is that the solutions tend be dominated by the overall liking (or the respondents' response tendency) rather than that one captures differences in the liking of the presented products. In other words, latent classes tend to differ much more with respect to the intercept of the regression model than with respect to slopes corresponding to the product attributes.

This paper illustrates an elegant way to overcome this problem. More specifically, we illustrate that the inclusion of a random intercept in a LC ordinal regression model is a

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

good way to separate an overall response level effect from differences in relative preferences for one cracker over another. As such, it provides a model-based alternative to within-case ‘centering’ of the data, which is the common practice. The use of a random intercept in regression model is very common in multilevel analysis (Goldstein, 1995; Snijders and Bosker, 1999) of which also variants for ordinal data have been proposed (Hedeker and Gibbons, 1996). Similar hybrid models combining latent classes and random effects have been proposed by Lenk and DeSarbo (2000) and Vermunt (2006).

Below we first introduce the substantive research question of interest and the data set at hand. Then, we present the extended LC ordinal regression model that we used for our analyses as well as specific the results obtained with the cracker data set. We end with a general discussion of the proposed approach.

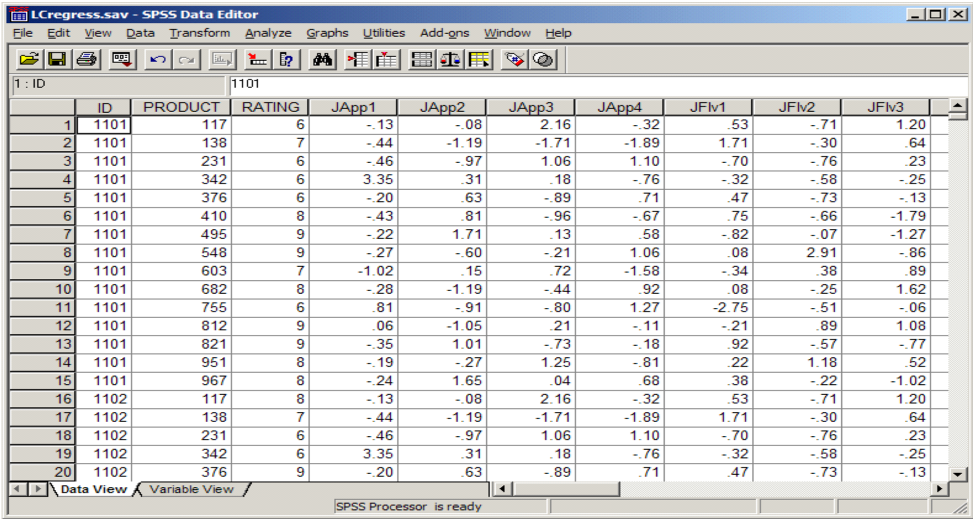
Description of the cracker case study

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.

The data layout required for these analyses is shown in Figure 1. 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.

As described in more details below, LC regression models were estimated with and without a random intercept to account for individual differences in average liking across all products. In Latent GOLD 4.0 (Vermunt and Magidson, 2005a) -- the program that was used to estimate the models -- a random intercept model is specified as continuous latent factor (C-Factor). Inclusion of such a random intercept is expected to result in segments that represent *relative* as opposed to absolute differences in cracker liking.



LCregress.sav - SPSS Data Editor

1 : ID 1101

	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	1.71	-.30	.64
3	1101	231	6	-.46	-.97	1.06	1.10	-.70	-.76	.23
4	1101	342	6	3.35	.31	.18	-.76	-.32	-.58	-.25
5	1101	376	6	-.20	.63	-.89	.71	.47	-.73	-.13
6	1101	410	8	-.43	.81	-.96	-.67	.75	-.66	-1.79
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	-1.58	-.34	.38	.89
10	1101	682	8	-.28	-1.19	-.44	.92	.08	-.25	1.62
11	1101	755	6	.81	-.91	-.80	1.27	-2.75	-.51	-.06
12	1101	812	9	.06	-1.05	.21	-.11	-.21	.89	1.08
13	1101	821	9	-.35	1.01	-.73	-.18	.92	-.57	-.77
14	1101	951	8	-.19	-.27	1.25	-.81	.22	1.18	.52
15	1101	967	8	-.24	1.65	.04	.68	.38	-.22	-1.02
16	1102	117	8	-.13	-.08	2.16	-.32	.53	-.71	1.20
17	1102	138	7	-.44	-1.19	-1.71	-1.89	1.71	-.30	.64
18	1102	231	6	-.46	-.97	1.06	1.10	-.70	-.76	.23
19	1102	342	6	3.35	.31	.18	-.76	-.32	-.58	-.25
20	1102	376	9	-.20	.63	-.89	.71	.47	-.73	-.13

Data View / Variable View

SPSS Processor is ready

Figure 1. Data Layout for the Regression Models

The LC ordinal regression model with a random intercept

Let Y_{it} denote the rating of respondent i of product t, with i = 1, 2,…, 157, t = 1, 2,…, 15. The rating Y_{it} takes on discrete values which are denoted by m; m=1, 2,…, 9. Since the rating is *not* a continuous but a discrete response variable, we work with an ordinal logit model or, more specifically, with an adjacent-category logit model (Agresti, 2002). As far as the explanatory variables in the model is concerned, we use two different specifications: in one model the ratings are assumed to depend on the products (modeled by 14 independent product effects) and in the other model the ratings are assumed to depend on 12 product characteristics. More specifically, we have:

- Model 1: a LC ordinal regression model with a random intercept and *product* effects that vary across latent classes
- Model 2: a LC ordinal regression model with a random intercept and *product-attribute* effects that vary across latent classes

Model 1 used the nominal variable PRODUCT as the sole predictor. It included a random intercept 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. The model of interest has the following form:

$$\log\left[\frac{P(Y_{it}=m\,|\,x)}{P(Y_{it}=m-1\,|\,x)}\right]=\alpha_{im}+\beta_{xt}=\alpha_m+\lambda F_i+\beta_{xt}.$$

As can be seen, this is a regression model for the logit associated with giving rating m instead of m-1 for cracker t conditional on membership of latent class x, for x = 1,2,…, K. In this model, α_{im} is the intercept, which, as can been seen from its indices, is allowed to vary across individuals. The specific parameterization we used is $\alpha_{im}=\alpha_m+\lambda F_i$, where F_i is a normally distributed continuous factor (the C-Factor score for the ith respondent), which has a mean equal to 0 and a variance equal to 1, and where λ is a factor loading. The implication of this parameterization is that expected value of the intercept $E(\alpha_{im}) = \alpha_m$ and its variance $Var(\alpha_{im}) = \lambda^2$. So, both the expectation and the square root of the variance are model parameters. More details on the factor-analytic parameterization of random effects models can, for example, be found in Skrondal and Rabe-Hesketh (2004) and Vermunt and Magidson (2005b).

The β_{xt} parameter appearing in the above equation is the effect of the t^{th} product for latent class or segment *x*. Effect coding is used for parameter identification, which implies that the α_m parameters sum to zero over the 9 possible ratings and that the β_{xt} sum to zero over the 15 product. Because of the effect coding, a positive value for β_{xt} means that segment *x* likes that product more than average, and a negative value that it likes the product concerned less than average.

Model 2 is the same as Model 1, except that it used the 12 sensory attributes as predictors. This yields the following LC regression model:

$$\log\left[\frac{P(Y_{it} = m \mid x)}{P(Y_{it} = m-1 \mid x)}\right] = \alpha_{im} + \sum_{q=1}^{12} \beta_{xq} z_{tq} = \alpha_m + \lambda F_i + \sum_{q=1}^{12} \beta_{xq} z_{tq}.$$

Here, z_{tq} denotes the value for attributes *q* for product *t*, and β_{xq} is the effect of the q^{th} attribute for latent class *x*. The remaining part of the model specification identical to Model 1.

Latent GOLD 4.0 estimates the LC ordinal logit model with random effects using maximum likelihood (ML), where the integral associated with the continuous factor F_i is solved by the Gauss-Hermite numerical integration procedure (Stroud and Secrest, 1966). In the current application, we used the Latent GOLD default setting of 10 quadrature nodes. To find the ML solution, Latent GOLD makes use of a combination of the EM and Newton-Raphson algorithms. More specifically, the program starts with a number of EM iteration and when close enough to the maximum it switches to the faster Newton-Raphson method. Lenk and DeSarbo (2000) proposed using Bayesian estimation methods for finite mixture models of generalized linear models with random effects of which the models described above are special cases

Results obtained with the cracker data set

Model 1

The correlation of the random intercept F_i with respondents' average liking across all cracker products was almost perfect (> 0.99). This shows that including the random intercept in the model is similar to 'centering', where the average liking rating is subtracted from the individual's ratings. The advantage of adjusting for the average rating through the use of a random intercept is that the original ordinal rating metric is preserved, so that the distributional assumptions made by the restricted multinomial likelihood function remain appropriate. A two-class solution provided the best fit to the data, with a model R^2 of 0.39.

Figure 2 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).

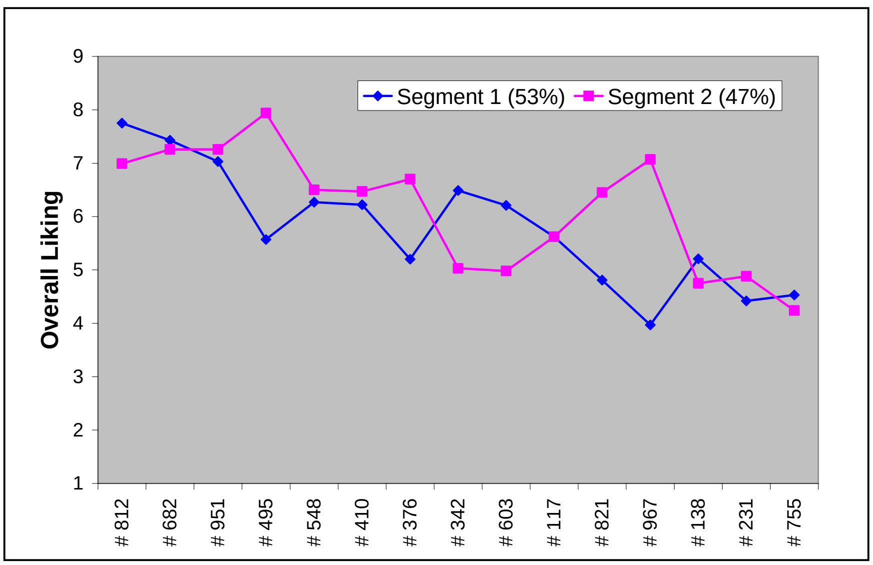


Fig. 2. Model 1 (2 classes) Results

Model 2

The correlation of the random intercept with average liking across products was again almost perfect (>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 lower 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 3 shows the average product liking scores for the three-class restricted model. The plot of regression coefficients in Figure 4 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.

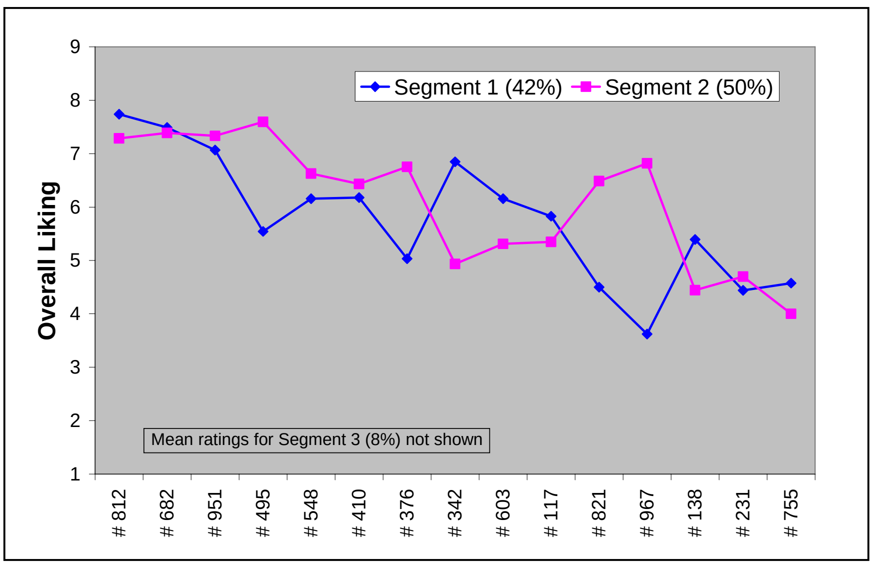


Fig. 3. Model 2 (3 classes) Results

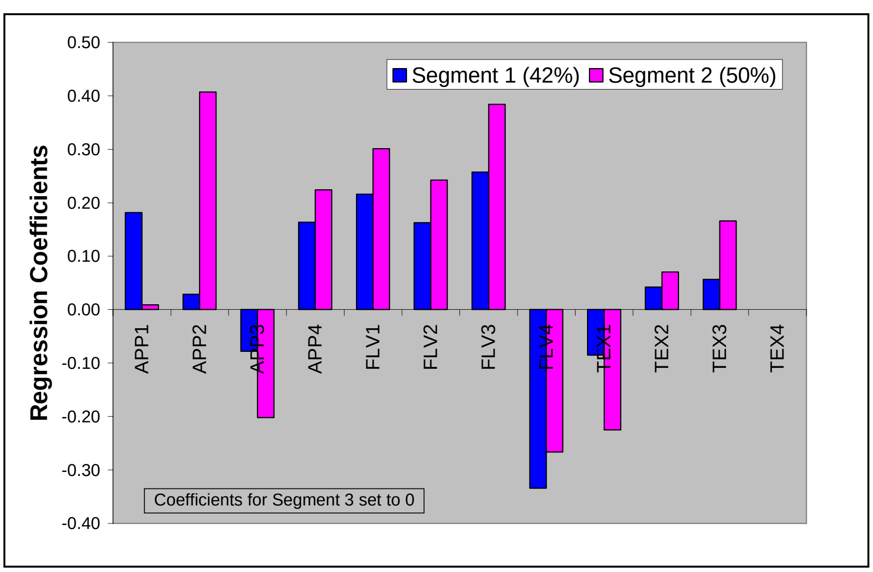


Fig. 4. Regression Coefficients for Model 2 (3 classes)

General discussion

Both Models 1 and 2 provided clear 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 correlation of the random intercept was in excess of 0.99 for both LC Regression models, which shows that 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 three reasons to prefer the regression approach in general. As mentioned above, 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, also in the case that of missing data or unbalanced designs. Third, the regression approach allows inclusion of multiple predictors, as was illustrated here in Model 2.

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). This suggests that the attributes can explain the segment level differences in product ratings.

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, Model 1 provides an upper bound on the R^2 for Model 2, when the same number of latent classes are specified. A comparison of the R^2 of Models 2 and 1 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.

With the data structure used in this study (while the attributes take on different values for each of the 15 products, they tale on the same values for each individual rater), there is a maximum number of predictors (here 14) that can be included in the regression model. Effect estimates of predictors beyond this number are not identifiable.

The use of restrictions in LC Regression Model 2 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 of "random responders" could be of substantive interest and warrant follow-up. If nothing else, the members of such a group can be excluded as outliers.

Among the 2 models tested Regression Model 2 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.

Regression models consisting of 1 CFactor to account for a random intercept, and additional CFactors instead of latent classes could be specified as a way of specifying continuously varying product or product-attribute effects. Such as specification is similar to what is done in Hierarchical Bayes (HB) (Andrews, Ainslie, and Currim, 2002) and multilevel model (Goldstein, 1995; Skrondal and Rabe-Hesketh, 2004; Snijders and Bosker, 1999). 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). In addition, the prior distribution used in HB may lead to somewhat different results than the ML framework. Such HB models were investigated with these data by Popper, Kroll, and Magidson (2004) who found that the BIC did not support the use of more than two C-Factors. The advantage of the LC regression models is that one obtains distinct segments that in the current application were found to be meaningful from a product development perspective.

References

Agresti, A. (2002). Categorical Data Analysis, Second Edition, New York: Wiley.

Andrews, R.L., Ainslie, A., and Currim, I.S. (2002), An empiracle comparison of logit choice models with discrete vs. continuous representations of heterogeneity, Journal of Marketing Research, 39, 479-487.

Goldstein, H. (1995). Multilevel Statistical Models. New York: Halsted Press

Hedeker, D. and Gibbons. R.D. (1996). MIXOR: A computer program for mixed effects ordinal regression analysis. Computer Methods and Programs in Biomedicine, 49, 157-176.

Lenk, P.J., and DeSarbo, W.S. (2000). Bayesian inference for finite mixture models of generalized linear models with random effects, Psychometrika}, 65, 93-119.

Snijders, T.A.B., and Bosker, R.J. (1999). Multilevel Analysis. London: Sage Publications.

Popper, R., Kroll, J., and Magidson, J. (2004). Applications of latent class models to food product development: a case study, Sawtooth Software Proceedings.

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

Stroud, A.H., and Secrest. D. (1966). Gaussian Quadrature Formulas. Englewood Cliffs, NJ: Prentice Hall.

Vermunt, J.K. (2006). Growth models for categorical response variables: standard, latent-class, and hybrid approaches. K. van Montfort, H. Oud, and A. Satorra (eds.). Longitudinal Models in the Behavioral and Related Sciences. Erlbaum. (in press).

Vermunt, J.K. and Magidson, J. (2005a) Latent GOLD 4.0 User's Guide. Belmont MA. Statistical Innovations Inc.

Vermunt, J.K. and Magidson, J. (2005b) Technical Guide to Latent GOLD 4.0: Basic and Advanced. Belmont MA. Statistical Innovations Inc.

Wedel, M., and Kamakura, W.A (1998). Market Segmentation: Conceptual and Methodological Foundations, Boston: Kluwer Academic Publishers.