Use of a Random Intercept in Latent Class Regression Models to Remove Response Level Effects in Ratings Data

Magidson, Jay
Statistical Innovations Inc.
375 Concord Ave, Suite 007
Belmont, MA 02478 USA
Jay@statisticalinnovations.com

Vermunt, Jeroen
Tilburg University
Room P-1.114
NL-5000 LE Tilburg, The Netherlands
j.k.vermunt@uvt.nl

Introduction

The existence and detection of consumer segments that differ in their taste preferences have important implications for food product development. Food manufacturers should understand these different taste preferences in order to develop successful new products tailored to each consumer segment. In this paper we compare various key driver latent class (LC) regression models to identify relevant segments in a case study involving the consumer evaluation of crackers.

In LC regression models, segments are formed on the basis of predictive relationships between a dependent variable (e.g., liking ratings of different crackers) and various sensory attributes (flavor, texture, appearance attributes of the cracker rated) as prediction variables. Segments are comprised of consumers whose coefficients are similar. By including these sensory attributes as predictors, LC regression models can identify the segments and their sensory drivers in one step and thus provide highly actionable results.

Many consumer segmentation studies show evidence of individual differences in response style (i.e. consumers differ systematically in how they use the response scale). 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. A question of considerable practical importance is whether the random intercept is effective in dealing with such response style differences.

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" (m=1) to "Like Extremely" (m=9)". 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.

Research Goals

The objectives of the research are a) to estimate various LC models to determine whether consumers can be segmented on the basis of differences in their liking ratings that can be linked to meaningful differences in the inferred preferences of various cracker attributes, b) to compare the model fit of these LC models, with and without a random intercept and c) to identify and interpret the resulting segments in terms of the sensory attributes that drive liking for that segment (in the case of the regression models).

To address these goals, three sets of regression models are considered here:

A. Traditional LC regression models. LC segments differ from each other with respect to both the intercept and regression coefficients.
B. LC regression with a random intercept. Each LC segment differs with respect to the regression coefficients only, individual-level variability with respect to the intercept being modeled separately using a continuous latent variable.
C. Same as B except that the LC segments are modeled using 2 or more discrete latent variables (DFactors).

Ordinal Regression Models with Latent Class Segments

For each of the following regression models, the ratings are treated as ordinal using an ordinal logit model known as the adjacent category logit model. In addition, the models containing a random intercept are formulated using a continuous factor F (CFactor). For further details on the use of CFactors see Skrondal and Rabe-Hesketh (2004).

The homogeneous ordinal logit model without any latent variables to handle heterogeneity can be expressed as the following K=1 class model:

1-Class Model

$$\text{logit}(Y_m)=\alpha_m+\beta_1 Z_1+\ldots+\beta_{12} Z_{12}$$

where the Z-variables refer to the 12 attributes and effects coding is used for the intercept:

$$\sum_{m=1}^{m=9}\alpha_m=0$$

On the other hand, the ***LC Ordinal Regression model with K > 1 classes and a Random Intercept is expressed as:***

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

Thus,

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

Ordinal Regression Models with Segments characterized by Dfactors

The DFactor alternative utilizes discrete factors instead of latent classes to deal with heterogeneity in the attribute effects.

Results from Various Regression Models

The results indicate the following:

- The random intercept models work well
- The DFactor models fit better than the traditional LC models

Name	Random Intercept	# Classes	# DFactors	LL	BIC(LL)	# Parameters
1cl	No	1	-	-4790.8	9677.7	19
2cl	No	2	-	-4759.3	9675.3	31
3cl	No	3	-	-4744.7	9706.8	43
1clR	Yes	1	-	-4729.9	9561.0	20
2clR	Yes	2	-	-4686.5	9534.8	32
3clR	Yes	3	-	-4744.7	9706.8	43
2dfacR	Yes	-	2	-4665.3	9553.0	44
3dfacR	Yes	-	3	-4643.4	9569.9	56
3clR restricted	Yes	3	-	-4682.2	9531.2	33
2dfacR restricted	Yes	-	2	-4671.2	9509.3	33
3dfacR restricted	Yes	-	3	-4657.3	9491.6	35

Attribute	Dfactor 1	Dfactor 2
JApp1	-0.0960	0.0000
JApp2	0.4206	0.0000
JApp3	-0.0906	0.0000
JApp4	0.1796	0.0000
JFlv1	0.0831	0.0000
JFlv2	0.1551	0.1955
JFlv3	0.2481	0.1942
JFlv4	0.0000	0.0000
JTex1	0.0000	-0.1333
JTex2	0.0000	0.0000
JTex3	0.1747	0.0000

Attribute	Dfactor 1	Dfactor 2	Dfactor 3
JApp1	-0.3097	0.0000	-0.1509
JApp2	0.4316	0.2446	0.0000
JApp3	-0.2721	0.0000	0.0000
JApp4	0.0000	0.1715	0.0000
JFlv1	0.0000	0.1497	0.0000
JFlv2	0.1419	0.0000	0.0000
JFlv3	0.2064	0.0000	0.0000
JFlv4	0.0000	0.0000	0.0000
JTex1	-0.3161	0.0000	-0.4870
JTex2	0.0000	0.0000	0.0000
JTex3	0.0000	0.1395	0.0000

REFERENCES

Skrondal and Rabe-Hesketh. *Generalized latent variable modeling: multilevel, longitudinal, and structural equation models*. Boca Raton: Chapman & Hall/CRC, 2004.

ABSTRACT

Latent class (LC) regression models classify respondents into a common segment (latent class) if they share the same values on the intercept and regression coefficients. While it is generally desirable that such segments differ on the (meaningful) coefficients, quite often they differ only on the intercept, a result that may not be meaningful from a strategic perspective. Intercept differences are especially common when the dependent variable is a rating, where different values reflect different response style tendencies -- some respondent segments providing more favorable ratings than others in all situations.

In this paper we analyze data obtained from a study involving consumer evaluation of 15 different cracker products. We investigate use of a random intercept to remove response level effects so that segments may better be differentiated on meaningful attributes (sensory drivers) of the crackers (texture, appearance, flavor) used as independent variables to predict the cracker rating. We compare results obtained from traditional unstructured latent classes in a LC regression model with a random intercept, to those obtained when the classes are structured by two or more discrete latent factors. The analyses are conducted using Latent GOLD version 4.5.

The results suggest that use of the random intercept is successful in removing the response level effects. In addition, the factor regression models fit the data better than the traditional LC regression models. 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 differential effects of the attributes across the segments suggest that different cracker products can be designed for each segment.

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