IES2025 Bressanone

Innovation & Society: Statistics and Data Science for Evaluation and Quality

SVQS

Società Italiana di Statistica

Research Group of Statistics for the Evaluation and Quality in Services of the Italian Statistical Society (SIS)

Innovation & Society

Statistics and Data Science for Evaluation and Quality

Editors: Giovanna Boccuzzo, Enrico Bovo, Marica Manisera, Luigi Salmaso

University of Padova
Casa della Gioventù
Brixen – Bressanone
June 25-27, 2025

cleup

First edition: June 2025
ISBN 978 88 5495 849 4

2025 Cleup sc
"Coop. Libraria Editrice Università di Padova"
via G. Belzoni 118/3, Padova (t. +39 049 8753496), Italy
www.cleup.it
www.facebook.com/cleup

Model-Based Clustering of Multivariate Rating Data to Evaluate Consumer Perceptions on Sustainability

Matteo Ventura[1], Julien Jacques[2], Paola Zuccolotto[1]
[1]Department of Economics and Management, University of Brescia, Brescia, Italy
[2]ERIC Laboratory, University Lumière Lyon 2, Lyon, France

Model-Based Clustering of Multivariate Rating Data to Evaluate Consumer Perceptions on Sustainability

Matteo Ventura, Julien Jacques, Paola Zuccolotto

Abstract Analyzing rating data is essential for understanding consumer perceptions but poses significant methodological challenges. This study applies the Multivariate Latent Class CUB (MLC-CUB) model to cluster rating data collected through a survey within the MICS project, which focuses on sustainability in the Italian furniture sector. The goal is to identify consumer segments with differing attitudes toward sustainability, highlighting key market priorities. The results reveal three distinct clusters: Sustainability Advocates, Profit Skeptics, and Balanced Realists. The analysis shows a relationship between sustainability orientation and willingness to pay a premium for Made in Italy or environmentally friendly products.

Keywords: sustainability, consumer perception, clustering, CUB model

1. Introduction

Ordinal variables are widely used in research to measure latent traits, such as perceptions and experiences, which are inherently unobservable. Questionnaires are a common tool in various scientific domains, including social sciences, marketing, and economics, to assess opinions through rating scales like the well-known Likert scale. However, analyzing ordinal data presents significant challenges, as it requires methods that respect both the ordinal nature of the variables and the structured format of questionnaires, which often consist of batteries of related items. Identifying hidden patterns within such data adds another layer of complexity, necessitating approaches that account for both the multivariate and ordinal aspects of the data.

Among the statistical models developed for the analysis of rating data, the Combination of discrete Uniform and shifted Binomial (CUB) model has proven to be particularly effective [1]. Within the framework set by this model, the recently proposed Multivariate Latent Class CUB (MLC-CUB) model provides an interesting method for uncovering latent patterns in multivariate ordinal data while preserving their inherent structure [2].

This study was carried out within the MICS (Made in Italy – Circular and Sustainable) Extended Partnership and received funding from Next-GenerationEU (Italian PNRR – M4 C2, Invest 1.3 – D.D. 1551.11-10-2022, PE00000004 CUP D73C22001250001), and applies MLC-CUB model to explore consumer perceptions of sustainability practices within the furniture industry, offering insights into market priorities and expectations. The Italian furniture sector, known for its excellence in craftsmanship and design, is increasingly considering sustainability as a key driver of innovation and competitiveness. Sustainability has become an essential focus, including economic, social, and environmental dimensions, and shaping the future of the industry. This study aims to provide an evaluation of consumer opinions toward sustainability in the Italian furniture sector, contributing to a deeper understanding of how sustainability influences consumer preferences.

2. Clustering data within the CUB framework

The CUB framework is based on the idea that the final rating chosen by a respondent is the result of the joint action of two latent components: the feeling, which represents the rational part of the decision process and the perceived sensation of the respondents regarding the feature they are evaluating, and the uncertainty, that is the indecision commonly present in every human choice.

Mathematically, the basic CUB model is a mixture of a shifted Binomial and a discrete Uniform distribution that respectively models the feeling and the uncertainty component. Formally, the rating r ($r = 1, \ldots, m$) is the realization of a discrete random variable R which is described by the following probability mass function:

$$P(R = r \mid \xi, \pi) = \pi \left[\binom{m-1}{r-1} \xi^{m-r} (1-\xi)^{r-1} \right] + (1-\pi) \frac{1}{m}, \tag{1}$$

where the parameter $(1 - \xi)$ represents the success probability of the shifted Binomial distribution. In the context of the CUB model, it is referred to as the feeling parameter, as it quantifies the respondent's level of conviction or agreement with the statement being evaluated. A higher value of $(1 - \xi)$ indicates stronger agreement, reflecting a more decisive and positive perception of the feature under assessment. the parameter π represents the mixture proportion, determining the relative weight of the feeling component. Since π takes values in the range $(0, 1]$, it also implicitly defines the contribution of the uncertainty component. Specifically, the quantity $(1 - \pi)$ is referred to as the uncertainty parameter in the CUB framework, as it quantifies the degree of indecision in the respondent's evaluation. A higher $(1-\pi)$ value indicates greater uncertainty, suggesting a less confident or more hesitant response.

The CUB model has been proven to be identifiable for $m > 3$. Parameter estimation is typically performed using the Expectation-Maximization (EM) algorithm, a standard approach for finite mixture models. This iterative procedure efficiently maximizes the likelihood function, providing robust estimates of the model parameters.

2.1 The MLC-CUB Model

Over the years, several researchers have proposed extensions to the CUB model, giving rise to what is now known as the CUB framework. This framework comprises all models designed for the analysis of rating data that assume a respondent's final rating results from a combination of feeling and uncertainty. Within this framework, a recent advancement is the Multivariate Latent-Class CUB (MLC-CUB) model, developed to perform model-based clustering of multivariate rating data. Clustering is a fundamental tool for researchers, commonly used to show hidden structures within data. However, while traditional clustering methods typically are used to analyze multivariate data, the original CUB model is used for the analysis of univariate data. To fill this gap, the MLC-CUB model was introduced as a mixture of multivariate CUB models, formulated under the assumption of variable independence.

A multivariate rating variable $\mathbf{R} = (R_j)_{j=1,\dots,J}$ composed of J rating variables with m_j categories is defined. Considering that ω_k is the mixture proportion of cluster k, such that $\omega_k > 0$ and $\sum_{k=1}^{K} \omega_k = 1$, by assuming the independence of the variables, the marginal distribution of $\mathbf{R}$ is defined as follows:

$$P(\mathbf{R} = \mathbf{r} \mid \boldsymbol{\xi}, \boldsymbol{\pi}, \boldsymbol{\omega}) = \sum_{k=1}^{K} \omega_k \prod_{j=1}^{J} \left\{ \pi_{jk} \left[\binom{m-1}{r-1} \xi_{jk}^{m-r} (1-\xi_{jk})^{r-1} \right] + (1-\pi_{jk}) \frac{1}{m} \right\}, \quad (2)$$

where $\boldsymbol{\xi} = (\xi_{jk})$ and $\boldsymbol{\pi} = (\pi_{jk})$ are the matrices containing the $J \times K$ feeling and uncertainty parameters. These parameters characterize both the clusters (indexed by k) and the variables (indexed by j), allowing the model to capture variations in response behavior across different groups and rating dimensions. The vector $\boldsymbol{\omega}$, instead, contains the K parameters ω_k which are the mixing proportions of the mixture.

Since the MLC-CUB model is a finite mixture model, the estimation of its parameters is carried out using the Expectation-Maximization (EM) algorithm. The process involves maximizing the complete log-likelihood, which is extended by introducing two indicator variables: one to account for the cluster membership of each multivariate observation and another to identify whether each univariate observation is drawn from the Uniform or shifted Binomial component, conditional on its cluster assignment.

To ensure reliable estimates, the algorithm is initialized randomly multiple times, and the model with the highest log-likelihood is selected. The number of clusters is determined using the Bayesian Information Criterion (BIC).

2.2 Identifiability of the model

The MLC-CUB model suffers from non-identifiability due to the fact that the parameter π_{jk}, which determines the weight of the uncertainty component, can vary across both clusters and variables. This variability makes it difficult to determine the level of uncertainty that characterizes each cluster. The challenge arises because the Uniform component does not have a specific parameter that defines its contribution, unlike the shifted Binomial component, which has a well-defined feeling parameter.

Since identifiability is a significant issue both theoretically and practically, a bootstrap procedure can be employed to detect potential identifiability problems when fitting the MLC-CUB model to real data. The procedure involves bootstrapping the original dataset and fitting the MLC-CUB model to each resampled dataset. Then, the Adjusted Rand Index (ARI) is used to compare the cluster assignments obtained from the bootstrapped data with those from the original dataset. By examining the distribution of the ARI values, it is possible to assess whether the estimated model is affected by identifiability issues. If the mode of the distribution is close to 1, it indicates that the model is not affected by identifiability problems. On the other hand, if the distribution is close to 0, it suggests that the model is likely suffering from identifiability issues, and the estimated parameter distributions will likely be multimodal rather than unimodal.

2.3 Assessing the goodness of fit

Among the various fitting measures proposed for the CUB models, the most common and widely used is the *Diss* index, defined as the difference between the observed and the expected relative frequencies of the rating r. Here, a Weighted Average Diss (WAD) index, weighted by the size of the clusters, is proposed to assess the goodness of fit of the MLC-CUB Model:

$$WAD = \frac{1}{2J} \sum_{j=1}^{J} \sum_{k=1}^{K} \hat{\omega}_k \sum_{r=1}^{m} \mid fr(r_{jk}) - P(r_{jk} \mid \hat{\xi}_{jk}, \hat{\pi}_{jk}) \mid, \tag{3}$$

where $fr(r_{jk})$ is the observed frequence of rating r for variable j in cluster k, $P(r_{jk})$ is the expected frequence of rating r for variable j in cluster k, and $\hat{\omega}_k$, $\hat{\xi}_{jk}$ and $\hat{\pi}_{jk}$ are the estimated parameters of the model. This index takes values in $[0, 1]$, where 0 means poor fit and 1 means perfect fit.

3. Case Study

As part of the MICS - Made in Italy Circular and Sustainable project, this study aims to explore consumer perceptions of sustainability within the Italian furniture sector. The project itself was conceived to analyze and communicate the diverse values embedded in the "Made in Italy" label, with a particular focus on environmental, economic, and social sustainability across industries. This specific research delves into how consumers perceive the sustainability objectives of companies operating in the furniture industry, offering insights into the priorities and expectations of the market.

To achieve this, a structured survey was conducted, asking respondents to evaluate the importance of various corporate objectives on a five-point Likert scale (1 = not important at all; 5 = very important). The key objectives assessed include: improving product quality and reliability, ensuring efficient production processes, reducing costs, increasing profits, enhancing product functionality and usability, promoting social well-being (among employees, consumers, and local communities), improving resource use efficiency, reducing waste and emissions, and ensuring effective end-of-life product management.

By analyzing consumer responses, this study seeks to identify which sustainability goals are considered most critical for "Made in Italy" furniture companies. The findings are intended to offer strategic insights for businesses looking to meet consumer expectations, thus strengthening Italy's position as a global leader in sustainable industrial practices.

A total of 405 responses were collected from the questionnaire. The data were analyzed using the MLC-CUB model. Through the Bayesian Information Criterion (BIC, value 9344.65), the model identified three distinct clusters within the data, each characterized by different levels of feeling and uncertainty which are represented in Figure 1.

Cluster 1, referred to as the "Sustainability Advocates", includes individuals with high levels of feeling and generally low uncertainty. These respondents consider all the objectives presented in the questionnaire as important. Notably, the highest levels of feeling are recorded for the items "Reducing waste and emissions" and "Ensuring effective end-of-life product management". This cluster represents consumers who strongly believe in and trust that companies are committed to environmental sustainability goals.

Cluster 2, labeled the "Profit Skeptics", is the smallest group, comprising only 8% of the respondents. It is characterized by high uncertainty regarding certain items (such as "Improving product quality and reliability" and "Reducing costs") and low levels of feeling across most objectives. The only objective with a high level of feeling is "Increasing profits", indicating that these respondents are skeptical and believe that profit maximization is the primary, if not the sole, objective for companies.

Cluster 3, known as the "Balanced Realists", is the largest group and consists of respondents with intermediate levels of feeling and very low uncertainty. The highest levels of feeling within this cluster are associated with "Improving product quality and reliability" and "Increasing profits", while the lowest feeling is observed for "Reducing costs". This group represents consumers who acknowledge the importance of both quality and profitability, reflecting a balanced perspective on corporate objectives.

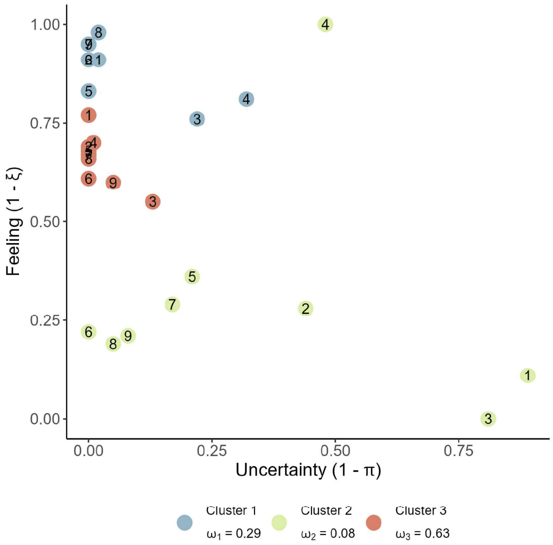

Figure 1: Representation of the model parameters which characterize each cluster. The numbers represent the associated question in the questionnaire.

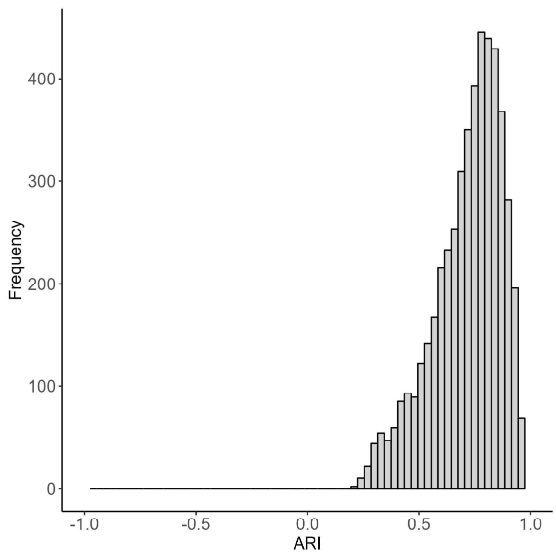

Figure 2: Distribution of the pairwise ARI indexes between the classification fitted on bootstrapped data.

To uncover valuable business insights, we analyzed the responses of different clusters to two survey questions: "How much more would you be willing to pay for a Made in Italy furniture product compared to a non-Made in Italy product?" and "How much more would you be willing to pay for a furniture product that, while maintaining the same characteristics, also ensures a lower environmental impact?". respondents could choose from three premium levels: 0%, less than 30%, or more than 30%.

To assess whether there is a relationship between cluster membership and willingness to pay a premium, we conducted a chi-square test and calculated Cramer's V to measure the strength of the association. In both cases, the chi-square test was significant at the 95% confidence level. However, the association observed through Cramer's V was weak, as the index remained below 15% of its theoretical maximum.

The findings suggest that Sustainability Advocates are slightly more willing to pay a premium for both Made in Italy products and those with a lower environmental impact. On the other hand, as expected, Profit Skeptics are less inclined to accept price increases.

The reliability of these results was assessed using a bootstrap procedure, which – as shown in Figure 2 – demonstrates stability in model identification. Additionally, the AWD index, with a value of 0.0786, indicates a good fit of the model to the data.

4. Conclusions

This study applied the MLC-CUB model to analyze consumer perceptions of sustainability in the Italian furniture sector, identifying three distinct clusters. The results indicate a weak but significant link between sustainability orientation and willingness to pay a premium. Future research could enhance clustering by incorporating copula-based models to account for variable dependence and zero-inflated models to better handle structural zeros in responses.

References

[1] Domenico Piccolo. On the moments of a mixture of uniform and shifted binomial random variables. *Quaderni di Statistica*, 5:85–104, 2003.

[2] Matteo Ventura, Julien Jacques, and Paola Zuccolotto. Clustering Multivariate Rating Data within the CUB Framework. In *Scientific Meeting of the Italian Statistical Society*, 637–642, 2024.

This book collects the papers presented at the 12th Scientific Meeting of the Group of Statistics for the Evaluation and Quality in Services (SVQS) "Innovation & Society: Statistics and Data Science for Evaluation and Quality", which took place at the Bressanone-Brixen campus of the University of Padova from June 25 to 27, 2025.
The papers, which had been selected through a refereeing process, focus on the role of statistics and data science as key tools for evaluating complex phenomena and ensuring quality in a variety of fields, including education, healthcare, environment, public administration, finance, tourism, sports, and beyond.

ISBN: 978-88-5495-849-4