SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2024
ORGANIZATION AND MANAGEMENT SERIES NO. 207

SEGMENTATION OF FOOD SERVICE CONSUMERS WITH SIMILAR CHOICE PREFERENCES

Tomasz BARTŁOMOWICZ

Wroclaw University of Economics and Business, Faculty of Economics and Finance, Department of Econometrics and Computer Science; tomasz.bartlomowicz@ue.wroc.pl, ORCID: 0000-0001-9986-0841

Purpose: The paper presents the application of latent class analysis and conjoint analysis methods in the segmentation of food service consumers with similar choice preferences. The main aim of the research was to identify latent classes of consumers and then measure the preferences within the segments. An additional main of the research was to demonstrate the combination of the `poLCA` and `conjoint` R packages as the one common tool.
Design/methodology/approach: Latent class analysis (including latent class models and latent regression class models) was used in the segmentation of consumers, while traditional conjoint analysis was applied in the measurement and analysis of stated consumers' preferences. All calculations were carried out using the R program and appreciate R packages.
Findings: The research confirmed that consumers' preferences of food service are so diverse that they should not be analyzed at an aggregate level but rather within specific segments. Understanding detailed preferences in segments allows business to tailor their offerings and profiling food services for specific consumers' groups.
Research limitations/implications: It may be interesting to combine other clustering and preference measurement methods, as well as additional R packages in future research of stated preferences.
Practical implications: The combination of `poLCA` and `conjoint` R packages enables the measurement of heterogeneous preferences in the homogeneous food consumers' segments and can also be applied in other contexts. The presented research results can be utilized by practitioners of food service market.
Social implications: The second R package used in the research – authoring `conjoint` R package implements the traditional conjoint analysis method similarly to the module Conjoint IBM SPSS program. The statistics (over half a million of downloads by RStudio users) indicate that the non-commercial `conjoint` package is popular among R users.
Originality/value: The research results as well as the combined use of different R packages can interest of students and researchers in the field of microeconomics and marketing research in the practical application of the latent class analysis and conjoint analysis methods.
Keywords: stated consumers' preferences, latent class analysis, conjoint analysis, R program.
Category of the paper: research paper.
JEL Classification: C6, C8, D1.

1. Introduction

One of the most important elements of marketing research is the measurement and analysis of consumers' preferences. Measurement of preferences can be conducted at the individual level for each consumer separately, at the aggregate level for the entire consumer community, or at the segment level for previously identified groups of consumers.

The individual approach assumes that modeling the preferences of each consumer individually is the most natural and intuitive, because each consumer perceives the products and services offered on the market in an individual (heterogeneous) way. In this case, the estimation of the parameters of the preference model is carried out at the individual level by assigning a separate utility function to each respondent. Traditionally, data on preferences are collected using the full profile method, and partial utilities are estimated based on the multiple regression model. Although individual models, in terms of fit to data and forecasting accuracy, are characterized by good statistical properties, the limitation of the individual approach is the lack of theoretical foundations that would allow for the transformation of individual preferences into group preferences, which enable the estimation and forecasting of market shares (Moore, 1980; Bąk, 2013).

In the aggregated (homogeneous) approach, the estimation of the preference model parameters is carried out at the aggregate level, which means that one utility function is used for the entire group of respondents. Partial utilities are most often estimated based on probabilistic models (e.g. multinomial logit model, conditional logit model, probit model). The advantage of the aggregated approach is the possibility of obtaining an estimate of market shares, but homogeneous models (due to the heterogeneity of preferences) are not characterized by statistical properties as good as individual models in terms of fit to data and forecasting accuracy (Moore, 1980; Bąk, 2013).

Both approaches can be considered as extremes, which are characterized by mentioned advantages and disadvantages. A compromise approach is the measurement of preferences at the segmental level, in which the parameters of the preference model are estimated for homogeneous groups and the heterogeneous nature of the preference measurement is preserved. A feature of the segmental approach is taking into account the advantages of extreme approaches while eliminating their main disadvantages. In the segmental approach, with a reduced number of estimated utility functions[1] it is possible to estimate market shares, and one can also have the appropriate predictive accuracy of the model (Moore, 1980; Bąk, 2013).

The paper presents the results of measurement and analysis of preferences of food service consumers at the segmental level by combining proven methods that enable consumer segmentation and preference analysis. For segmentation, latent class analysis (latent class

[1] In comparison to the number of utility functions necessary to estimate at the individual approach.

models and latent class regression models) was used, allowing for the clustering of consumers into homogeneous groups and the preliminary measurement of preferences. Then, for each group separately, a detailed measurement of preferences was conducted at the individual (heterogeneous) level using the conjoint analysis method (traditional conjoint model).

The main originality of the paper is demonstrated by the obtained results of the research, which confirmed some of the expected results, including the importance of some variables, and helped to discover unexpected differences between the segments of respondents at the detailed level of preference measurement.

All calculations and visualizations of the obtained results were carried out using R program and appropriate R packages.

2. Methods

In the segmentation and measurement of consumers' preferences, various multidimensional data analysis methods can be used. For segmentation, subjective *a priori* or formal *post hoc* methods (such as cluster analysis, latent class analysis models, models with random parameters) are most commonly applied (Bąk, 2013). For measurement of preferences, conjoint analysis and discrete choice models are typically used. Given the tools employed, latent class analysis and traditional conjoint analysis are discussed in more detail in the next sections of the paper.

2.1. Latent class analysis

The latent class analysis (LCA) method has over fifty years of history and a well-established position among other clustering methods. The first publications discussing the concept of a latent variable were introduced by Lazarsfeld (1950), but due to the lack of formal methods for estimating model parameters, the application of latent class analysis was limited at that time. The breakthrough period was 1968, when Lazarsfeld and Henry (1968) presented the mathematical workshop and the concept of the method, and then 1974, when Goodman (1974) presented statistical solutions for estimating model parameters, developed the maximum likelihood method, extended the application of latent class analysis to polytomous variables and multidimensional latent class models (Goodman, 1974). To date, many publications have been published on the issue of latent class analysis, presenting various models of data analysis. Also many computer programs and packages supporting empirical research have been developed.

Latent class analysis is a cluster analysis tool representing an approach based on a probability model, in which the research material is marketing data on the declared preferences of respondents, obtained mainly as a result of survey research. Unlike classic cluster analysis methods, which use distance measures (similarity, dissimilarity) for classification, latent class analysis uses a model approach in which the probabilities of objects belonging to classes are calculated and then objects are classified based on these values. In this way, objects (individuals, respondents, entities) are grouped (divided) into separate and homogeneous classes (segments). The advantage of the model approach is the ability to take various variables (continuous and discrete) measured on different scales (metric and non-metric) in the research (Everitt et al., 2011; Bąk, 2013; Brzezińska 2021).

Latent class analysis assumes the existence of some abstract characteristic that cannot be observed directly by the researcher. This means that there are some hidden, unobserved variables that constitute the basis of interest in latent class analysis. Hidden dependencies between variables describing the examined objects determine whether the objects belong to particular classes.

The following types of variables can be distinguished in latent class models:

- manifest variables or dependent variables that can be measured on different measurement scales,
- latent variables that can be measured on nominal or ordinal scales,
- predictor variables and covariates (concomitant variables) that can be measured on different scales.

A model in latent class analysis must contain at least one manifest (or dependent) variable and at least one latent variable. In addition, the model may include concomitant variables. Variables in the model can be continuous or discrete.

The basic types of models used in the approach based on probability models include:

- mixture models,
- latent class models,
- latent class regression models.

The basic types of models used in preference research and segmentation using latent class analysis method include latent class models and latent class regression models.

Latent class models can be written using the formula (Wedel and DeSarbo, 1994; Vriens, 2001), (Bąk, 2013):

$$f(y|\Phi) = \sum_{c=1}^{C} \pi_c \, f_c(y|\theta_c) \tag{1}$$

where:

f – distribution function of observations (empirical preferences),

y – empirical preferences,

$\Phi = (\pi, \theta)$ – unknown model parameters,

π_c – unknown size of c-th segment (interpreted as a mixing parameter representing the affiliation of observations to particular hidden classes),

f_c – distribution function of observations in c-th class,
θ_C – parameters estimated for c-th segment,
$c = 1, \ldots, C$ – number of segment.

Latent class regression models can be divided into two cases – a latent class regression model with explanatory variables (Vriens, 2001; Bąk, 2013):

$$f(y|x) = \sum_{c=1}^{C} \pi_c \, f_c(y|\pi_C, x) \tag{2}$$

where:
y – observation vector,
x – predictor variables that affect y,
π_C – probability of belonging to the c-th class or the the y-segment size; and a latent class regression model with predictor variables and covariates (Vriens, 2001; Bąk, 2013):

$$f(y|x) = \sum_{c=1}^{C} (\pi_C|z) \, f_c(y|\pi_C, x) \tag{3}$$

where: z – covariates that affect the latent variable (membership in latent classes).

Predictor variables should be understood as attributes of products or services. Including these variables in the model enables consumer segmentation using consumer characteristics (geographical, demographic, cultural, socio-economic and others) (Bąk, 2013).

One of the most important stages of using latent class analysis is estimating the model parameters. Among the most commonly used statistical methods for estimating the parameters of latent class models, the maximum likelihood method should be distinguished. The values of the maximum likelihood estimators are usually found using optimization algorithms, including the Expected Maximization (EM) algorithm (Dempster, Laird, Rubin, 1977) or the Newton-Raphson algorithm (Raphson, 1960). Both approaches rely on iterative estimation of the maximum likelihood value and both algorithms start with a certain initial value. The algorithms continue to operate until the specified criteria are met (Everitt, 1987; McLachlan, Krishman, 1977; Wedel, Kamakura, 2000).

The main advantage of the EM algorithm is the increase in the value of the likelihood function in each subsequent iteration, which makes it a more frequently used method compared to the Newton-Raphson method. In addition, the EM algorithm can also be used to supplement missing data in the sample. The disadvantages of the EM algorithm include a large number of necessary iterations, slower operation, and difficult estimation of standard errors (Brzezińska, 2021). The EM algorithm and its applications for LCA are presented e.g., in the following studies (Aitkin, Anderson, Hinde, 1981; McLachlan, Krishnan, 1997; Dempster, Laird, Rubin, 1977; Chen, 1981; Dempster, Rubin, Tsutakawa, 1981; Kamakura, Russell, 1989; Hamilton, 1991; DeSarbo, Wedel, Vriens, Ramaswamy, 1992).

In assessing the fit of latent class models to data and selecting the best-fitting model, relative fit criteria are most often used, of which the basic inferential rate is the likelihood ratio test. Additionally, various forms of information-heuristic rates are used, such as the Akaike Information Criterion (AIC) (Akaike, 1973; Akaike, 1987) and the Bayesian Information Criterion (BIC) (Schwarz, 1978; Konishi, Kitagawa, 2008).

LCA belongs to a larger family of latent variable techniques called finite mixture models (FMM) (Bouveyron et al., 2019), which comprises a wide range of cross-sectional and longitudinal models that all involve one or more latent class variables. More information about mixture models and latent class analysis can be found in the subject literature (Goodman, 1974; Wedel, DeSarbo, 1994; McLachlan, Peel, 2000; Oberski, 2016; Linzer, Lewis, 2022; Vriens 2001; Nylund et al., 2007; Vermunt, Magidson, 2002; Vermunt, 2010; Colins, Lanza, 2010; Nylund-Gibson, Choi, 2018; Lubke, Muthén, 2005; Masyn, 2013).

2.2. Traditional conjoint analysis

The traditional conjoint analysis method also has a long history of over forty years and holds a well-established position among methods of measurement and analysis of stated preferences. The first publication presenting conjoint measurement in psychometrics appeared in 1964 (Luce, Tukey, 1964), followed by additional works in the 1970s (Green, Rao, 1971; Green, Wind, 1973; Green, Srinivasan, 1978). Since then, numerous studies have discussed the methodological challenges and applications of conjoint analysis in marketing research. Nowadays, conjoint analysis is a widely used method for studying consumer preferences for products and services, as well as political opinions and religious attitudes. A comprehensive review of the existing achievements and future development perspectives in conjoint analysis is provided in (Green, Krieger, Wind, 2004; Gustafsson, Herrmann, Huber, 2007; Rao, 2014).

In empirical research, conjoint analysis methods are often used in the analysis of stated preferences measured on metric scales. In such cases, a multiple regression model with dummy variables is usually used, the parameters of which are estimated by the classical method of least squares (OLS).

The research material used in the conjoint analysis method is marketing data on respondents' declared preferences, obtained mainly as a result of survey research. Respondents evaluate product or service profiles[2] (real or hypothetical) described by a set of features (attributes), thus expressing their (empirical) preferences. Based on the collected data, the total preferences are decomposed using statistical methods by calculating the share of each attribute in the estimated total utility value of the profile. Part-worth utilities are estimated for each respondent individually and as average values for the studied sample (Green, Wind, 1975).

[2] Attributes and their levels generate different variants (profiles) of goods or services. The number of all possible profiles to be generated depends on the number of attributes and the number of levels (it is the product of level numbers of all attributes). In practice, only a subset of variants meeting the relevant conditions (e.g. of the system orthogonality) is ranked by respondents in the form of the so-called fractional factorial design.

One of the more important stages of the conjoint analysis procedure is the estimation of the parameters of conjoint analysis model. In the traditional conjoint analysis, the linear multiple regression model is developed, the parameters of which (part-worth utilities of the attribute levels) are estimated using the classical Ordinary Least Squares (OLS) method. The model for the selected respondent can be presented in the following form (Hair, Anderson, Tatham, Black, 1995):

$$\widehat{Y_s} = b_{0s} + b_{1s}X_{1s} + \cdots + b_{ms}X_{ms} = b_{0s} + \sum_{j=1}^{m} b_{js}X_j \tag{4}$$

where:
$s = 1, \ldots, S$ – respondent's number;
S – number of respondents.

The knowledge of part-worths utilities allows conducting the analysis covering:

- the theoretical total utilities of the profiles in the cross-section of respondents,
- the analyzed sample and the identified groups (segments) of respondents,
- the relative "importance" ranking of individual attributes in the cross-section of respondents in the analyzed sample,
- the simulation market shares of the selected profiles,
- the segmentation of respondents.

The total utility of i-th profile for s-th respondent (U_i^s) is calculated based on the following formula (Hair, Anderson, Tatham, Black, 1995; Walesiak, 1996):

$$U_i^s = \sum_{j=1}^{m} b_{0s} + U_{l_j^i}^s \tag{5}$$

where:
b_{0s} – the intercept for s-th respondent;
$U_{l_j^i}^s$ – part-worths utility of l-th level of j-th attribute of i-th profile for s-th respondent;
l_j^i – level number of j-th attribute in i-th profile.

The average theoretical total utility (at an aggregated level, i.e., for the whole sample covering S respondents) of i-th profile (U_i) is calculated based on the following formula (Hair, Anderson, Tatham, Black, 1995; Walesiak, 1996):

$$U_i = \frac{1}{S}\sum_{s=1}^{S}\left(\sum_{j=1}^{m} b_{0s} + U_{l_j^i}^s\right) \tag{6}$$

The knowledge of part-worths utilities also allows estimating the "importance" for every attribute in the assessment of profiles, which are the subject of research. The relative importance of j-th attribute for s-th respondent (W_j^s) is calculated using the formula (6) (Hair, Anderson, Tatham, Black, 1995):

$$W_j^s = \frac{max\{U_{l_j^i}^s\} - min\{U_{l_j^i}^s\}}{\sum_{j=1}^{m}\left(max\{U_{l_j^i}^s\} - min\{U_{l_j^i}^s\}\right)} \times 100\% \tag{7}$$

The average "importance" of particular attributes in the cross-section of the whole sample covering S respondents (W_j) is calculated based on the formula:

$$W_j = \frac{1}{S}\sum_{s=1}^{S} W_j^s \tag{8}$$

where: W_j^s – defined by a formula (7).

The results in the form of estimated partial utilities obtained in the conjoint analysis procedure can be used in simulation models of market events, the so-called choice simulators, which enable the analysis of what-if scenarios. Making some simulation analysis of market shares it is also possible to estimate the total utility of additional profiles, which were not ranked by the respondents in the survey. The anticipated market share of the selected profiles is estimated based on the maximum utility model, probabilistic BTL (Bradley-Terry-Luce) model and logit model (Hair et al., 1995; Walesiak, 1996; Walesiak, Bąk, 2000; Bąk, 2013).

The parameter values of the estimated conjoint analysis model (estimated part-worth and total utilities) can additionally constitute the basis for consumers' segmentation, as they reflect the respondents' preferences presented in the research regarding the specific profiles of products and services.

More information about conjoint analysis methods and its applications in practice can be found in subject literature (Hair et al., 1995; Coombs, Dawes, Tversky, 1977; Green, Rao, 1971; Green, Srinivasan, 1978, 1990; Green, Wind, 1975; Wilkinson, 1998; Vriens, Wittink, 1994; Zwerina, 1997; Poortinga et al., 2003; Gustafsson, Herrmann, Huber, 2007; Rao, 2014; Lu, Zhang, 2020; Walesiak, 1996; Walesiak, Bąk, 2000; Bąk, 2004; Bąk, Bartłomowicz, 2012, 2018b; Bartłomowicz, Bąk, 2021).

3. Segmentation and measurement of preferences

In the segmentation and measurement of food service consumers' preferences data from a survey conducted in 2022 were used. The survey questionnaire included questions for conjoint analysis method, questions for latent class analysis and questions about basic respondents' characteristics. A total of 154 survey questionnaires were distributed electronically using the Microsoft Teams, of which 122 questionnaires were correctly completed and used as a source of statistical data. For the purposes of segmentation and the preliminary measurement of the consumers' preferences the appropriate latent class model and latent class regression models were used, while for the purposes of detailed measurement of stated preferences the traditional conjoint analysis method was used.

All calculations were carried out using the R program with `poLCA` package (Linzer, Lewis, 2024) for latent class analysis, the `conjoint` package (Bąk, Bartłomowicz, 2018a) for conjoint analysis method, the `ltm` package (Rizopoulos, 2022) for basic descriptive statistics and the `ggplot2` package (Wickham et al., 2024) for visualization of the obtained results.

3.1. Segmentation

For the purpose of respondents' segmentation, a question from the second part of the survey was used, in which 8 observed variables were selected: quality of products, originality of the dish, quality of service (by the salesperson or waiter), additional amenities, free parking, description of food allergens, use of organic ingredients and food service location. Respondents indicated the significance of each variable on a 5-point polytomous scale. Sample answers for first respondent are presented in Table 1.

Table 1.
Sample answers for first respondent

Variable	Significance to respondent				
	very small (1)	small (2)	average (3)	high (4)	very high (5)
Quality of products					X
Dish originality					X
Quality of service				X	
Additional amenities				X	
Free parking		X			
Food allergens	X				
Ecological ingredients	X				
Location			X		

Source: survey questionnaires.

The options for this question were included in the latent class models as manifest variables and served as the basis for respondents’ segmentation. In the research, respondents were also asked to provide some of their own characteristics[3] which allowed for the inclusion of some selected covariates in the research using latent class regression models.

[3] The asked characteristics included the following variables (with corresponding levels): sex (male, female), age (open question), education level (basic education, secondary education, some university education, higher education), net income (up to 2000 PLN, between 2000 and 3500 PLN, between 3500 and 5000 PLN, above 5000 PLN), frequency of using food service (less than once a month, once a month, once a week, 1-3 times a week, more than 3 times a week).

First, the collected data were summarized using the `descript` function from `ltm` R package, which allows obtaining basic descriptive statistics for polytomous data (cf. Table 2):

```
> data=read.csv2("food_lca.csv", header=TRUE)
> library(ltm)
> des=descript(data)
> print(des)
```

Table 2.
Frequency of selecting a given option by respondents

Variable	Respondents' answers				
	very small (1)	small (2)	average (3)	high (4)	very high (5)
Quality of products	0,0082	0,0082	0,0246	0,1803	0,7787
Dish originality	0,0492	0,0984	0,2951	0,3197	0,2377
Quality of service	0,0082	0,0328	0,2213	0,4426	0,2951
Additional amenities	0,0984	0,1967	0,2213	0,3033	0,1803
Free parking	0,1885	0,1311	0,1475	0,2377	0,2951
Food allergens	0,2705	0,1311	0,1639	0,1803	0,2541
Ecological ingredients	0,1803	0,1721	0,2787	0,2541	0,1148
Location	0,0410	0,0246	0,2377	0,3525	0,3443

Source: author's compilation using `ltm` R package.

The initial data analysis indicates that not all variables are equally strongly preferred. The quality of products is very important to almost all respondents (77.87% of "very high" and 18.03% of "high" responses). In the case of the food allergens and ecological ingredients, respondents most often selected the option "very small" and "very high", which allows us to conclude that these features are important only to some respondents. In the case of the rest variables (dish originality, quality of service, additional amenities, free parking and location) respondents' responses were highly dispersed, which indicates the division of respondents into separate classes and need for further analysis using the latent class analysis method.

The first analyzed was the latent class model with all manifest variables. In order to select the optimal number of classes, models for 2, 3 and 4 classes were estimated (each model was estimated 3 times with different starting values of the optimization algorithm, each time from 100 models). In this way, using the `poLCA` function latent class models were estimated, the fit of which was the best based on the AIC and BIC criteria:

```
> model=cbind(quality, originality, service, amenities, parking, allergens,
ecology, location)~1
> for (k in 2: 4) {
+      min_ll=0
+      min_aic=10000
+      min_bic=min_aic
+      for (m in 1:100) {
+             lca=poLCA(model, data, nclass=k, nrep=3, tol=1e-10, verbose=FALSE)
+             if (lca$ll<min_ll) {min_ll=lca$ll}
+             if (lca$aic<min_aic) {min_aic=lca$aic}
+             if (lca$bic<min_bic) {min_bic=lca$bic}
+      }
+      cat("Model nr", k, "\n")
+      cat("LL: ", min_ll, "\n")
```

```
+      cat("AIC: ", min_aic, "\n")
+      cat("BIC: ", min_bic, "\n")
+ }
Model nr 2
LL:  -1279.827
AIC:  2639.755
BIC:  2822.016
Model nr 3
LL:  -1232.605
AIC:  2620.049
BIC:  2894.843
Model nr 4
LL:  -1201.035
AIC:  2621.632
BIC:  2988.958
```

The indications of the AIC and BIC criteria are not unambiguous (the AIC criterion indicates 3 segments, while the BIC criterion indicates 2 segments). The comparison of both models (cf. Figures 1-2) in terms of the assessment of the probability of choosing a food service due to the selected variables indicates that the division of the studied sample according to the AIC criterion (3 segments) has a greater interpretative value. Among the selected variables, the most important in all segments is "very high" products' quality, while the perception of the remaining variables by the respondents depends on the given segment. Therefore, the basis for further analysis was the division of the respondent population into 3 segments, with percentage shares of 38.9%, 32.7% and 28.4%:

```
> min_aic=10000
> min_bic=min_aic
> for (k in 2: 3) {
+      for (m in 1:100) {
+             lca=poLCA(model, data, nclass=k, nrep=3, tol=1e-10, verbose=FALSE)
+             if (lca$aic<min_aic) {min_aic=lca$aic; mod3=lca}
+             if (lca$bic<min_bic) {min_bic=lca$bic; mod2=lca}
+      }
+ }
> print(mod2)
> windows(width=5, height=4, pointsize=9)
> plot(mod2)
Conditional item response (column) probabilities,
 by outcome variable, for each class (row)

$quality
          Pr(1) Pr(2)  Pr(3)  Pr(4)  Pr(5)
class 1:  0.019 0.019 0.0190 0.1883 0.7548
class 2:  0.000 0.000 0.0289 0.1743 0.7969

$originality
           Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)
class 1:  0.0945 0.2085 0.2323 0.2736 0.1910
class 2:  0.0146 0.0145 0.3429 0.3547 0.2733

$service
          Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)
class 1:  0.019 0.0569 0.3030 0.4104 0.2107
class 2:  0.000 0.0144 0.1591 0.4671 0.3594

$amenities
           Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)
class 1:  0.2275 0.3380 0.1573 0.1477 0.1295
class 2:  0.0000 0.0891 0.2700 0.4218 0.2190
```

```
$parking
           Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)
class 1:  0.3408 0.2392 0.1153 0.1163 0.1885
class 2:  0.0725 0.0488 0.1721 0.3302 0.3763

$allergens
           Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)
class 1:  0.5965 0.3033 0.0000 0.1002 0.0000
class 2:  0.0221 0.0000 0.2888 0.2414 0.4477

$ecology
           Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)
class 1:  0.3600 0.3083 0.1374 0.1942 0.0000
class 2:  0.0435 0.0684 0.3863 0.2997 0.2022

$location
           Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)
class 1:  0.0744 0.0569 0.2836 0.3201 0.2650
class 2:  0.0155 0.0000 0.2028 0.3771 0.4046

Estimated class population shares
 0.4324 0.5676

Predicted class memberships (by modal posterior prob.)
 0.4344 0.5656

=========================================================
Fit for 2 latent classes:
=========================================================
number of observations: 122
number of estimated parameters: 65
residual degrees of freedom: 57
maximum log-likelihood: -1254.878
AIC(2): 2639.755
BIC(2): 2822.016
G^2(2): 1345.892 (Likelihood ratio/deviance statistic)
X^2(2): 1079715 (Chi-square goodness of fit)
```

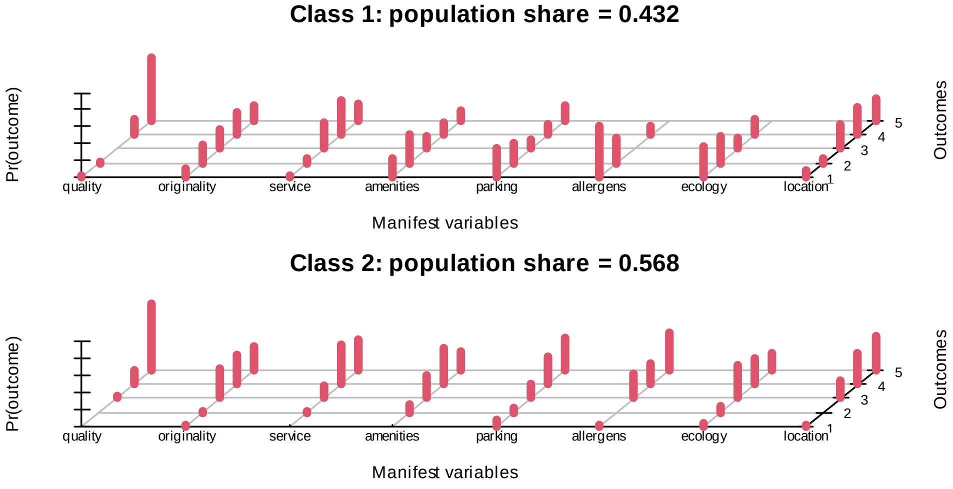


Figure 1. Estimated probabilities of selecting options for 2 classes.

Source: author's compilation using `poLCA` R package.

```
> print(mod3)
> windows(width=5, height=4, pointsize=9)
> plot(mod3)
Conditional item response (column) probabilities,
```

```
 by outcome variable, for each class (row)

$quality
           Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)
class 1:  0.0211 0.0211 0.0211 0.1871 0.7497
class 2:  0.0000 0.0000 0.0253 0.2035 0.7712
class 3:  0.0000 0.0000 0.0287 0.1443 0.8270

$originality
           Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)
class 1:  0.1053 0.2315 0.2161 0.2922 0.1550
class 2:  0.0251 0.0252 0.4637 0.4042 0.0819
class 3:  0.0000 0.0000 0.2092 0.2599 0.5309

$service
           Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)
class 1:  0.0211 0.0632 0.3335 0.4102 0.1721
class 2:  0.0000 0.0000 0.1961 0.6459 0.1580
class 3:  0.0000 0.0289 0.0964 0.2527 0.6220

$amenities
           Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)
class 1:  0.2526 0.3549 0.1706 0.1489 0.0729
class 2:  0.0000 0.1790 0.3289 0.4582 0.0340
class 3:  0.0000 0.0000 0.1669 0.3366 0.4966

$parking
           Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)
class 1:  0.3826 0.2741 0.0992 0.1121 0.1320
class 2:  0.0000 0.0462 0.2114 0.5934 0.1490
class 3:  0.1396 0.0328 0.1403 0.0000 0.6873

$allergens
           Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)
class 1:  0.5478 0.3369 0.0000 0.1154 0.0000
class 2:  0.0325 0.0000 0.2984 0.2707 0.3984
class 3:  0.1643 0.0000 0.2339 0.1652 0.4365

$ecology
           Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)
class 1:  0.3377 0.3194 0.1384 0.2045 0.0000
class 2:  0.0590 0.0000 0.5086 0.3400 0.0925
class 3:  0.1043 0.1685 0.2062 0.2232 0.2980

$location
           Pr(1)  Pr(2)  Pr(3)  Pr(4)  Pr(5)
class 1:  0.0686 0.0632 0.2954 0.3468 0.2260
class 2:  0.0000 0.0000 0.1482 0.6343 0.2174
class 3:  0.0503 0.0000 0.2616 0.0353 0.6528

Estimated class population shares
 0.3893 0.327 0.2836

Predicted class memberships (by modal posterior prob.)
 0.3852 0.3197 0.2951

=========================================================
Fit for 3 latent classes:
=========================================================
number of observations: 122
number of estimated parameters: 98
residual degrees of freedom: 24
maximum log-likelihood: -1212.025
AIC(3): 2620.049
BIC(3): 2894.843
G^2(3): 1260.186 (Likelihood ratio/deviance statistic)
X^2(3): 825988.5 (Chi-square goodness of fit)
```

The obtained results present the probability of respondents choosing a given option for each analyzed variable. The `poLCA` function optionally creates a graph presenting the probability of choosing an option. Additionally, information about the size of individual segments is visible.

Additionally, using latent class regression models, it is possible to determine the influence of variables characterizing respondents on the segmentation. For this purpose, the parameters of the latent class regression models with the covariate net income, with the covariate education and with the covariate frequency of service use were estimated (sample for first covariate):

```
> model=cbind(quality, originality, service, amenities, parking, allergens,
ecology, location)~income
> min_aic=10000
> for (m in 1:100) {
+      lca=poLCA(model, data, nclass=3, nrep=3, tol=1e-10, verbose=FALSE)
+      if (lca$aic<min_aic) {min_aic=lca$aic; mod3i=lca}
+ }
> ps_aic=poLCA.reorder(mod3i$probs.start, order(mod3i$P, decreasing=TRUE))
> lca_aic=poLCA(model, data, nclass=3, probs.start=ps_aic)
> windows(width=3.0, height=3.5, pointsize=10)
> par(cex.main=0.9, cex.lab=0.85, cex.axis=0.85)
> pd=cbind(1, c(1:4))
> exb=exp(pd%*%lca_aic$coeff)
> par(cex.main=0.9, cex.lab=0.85, cex.axis=0.85)
> pic=cbind(1, exb)/(1+rowSums(exb))
> matplot(c(1:4), pic, ylim=c(min(pic), max(pic)), xaxt="n",
+ col=c("#C79a9a", "#9fc79a", "#ff0010"), type="l", lwd=3,
+ main="Model with covariate (3 classes)",
+ xlab="Net income low (1) - high (4)", ylab="Probability of class membership")
> axis(1, at=c(1, 2, 3, 4))
> text(1.1, 0.34, "1", col="#C79a9a", cex=0.9)
> text(3.9, 0.22, "2", col="#9fc79a", cex=0.9)
> text(1.1, 0.22, "3", col="#ff0010", cex=0.9)
> model=cbind(quality, originality, service, amenities, parking, allergens,
ecology, location)~education
> min_aic=10000
> for (m in 1:100) {
+      lca=poLCA(model, data, nclass=3, nrep=3, tol=1e-10, verbose=FALSE)
+      if (lca$aic<min_aic) {min_aic=lca$aic; mod3e=lca}
+ }
> ps_aic=poLCA.reorder(mod3e$probs.start, order(mod3e$P, decreasing=TRUE))
> lca_aic=poLCA(model, data, nclass=3, probs.start=ps_aic)
> windows(width=3.0, height=3.5, pointsize=10)
> par(cex.main=0.9, cex.lab=0.85, cex.axis=0.85)
> pd=cbind(1, c(1:4))
> exb=exp(pd%*%lca_aic$coeff)
> par(cex.main=0.9, cex.lab=0.85, cex.axis=0.85)
> pic=cbind(1, exb)/(1+rowSums(exb))
> matplot(c(1:4), pic, ylim=c(min(pic), max(pic)), xaxt="n",
+ col=c("#C79a9a", "#9fc79a", "#ff0010"), type="l", lwd=3,
+ main="Model with covariate (3 classes)",
+ xlab="Education level basic (1) - high (4)",
+ ylab="Probability of class membership")
> axis(1, at=c(1, 2, 3, 4))
> text(1.1, 0.34, "1", col="#C79a9a", cex=0.9)
> text(1.1, 0.15, "2", col="#9fc79a", cex=0.9)
> text(3.9, 0.22, "3", col="#ff0010", cex=0.9)
```

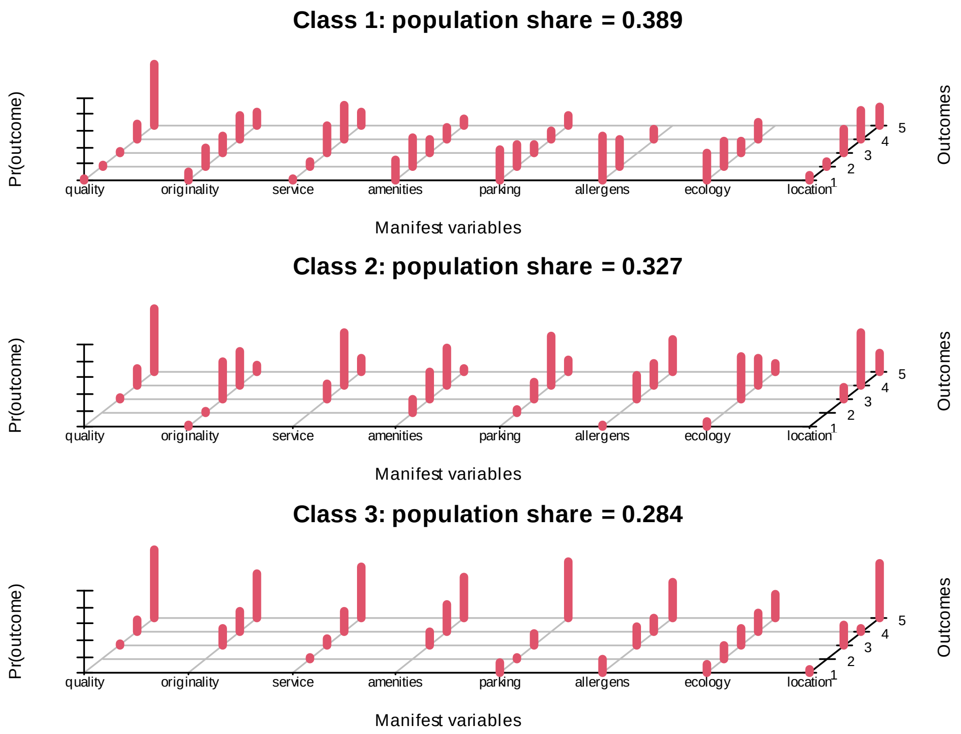
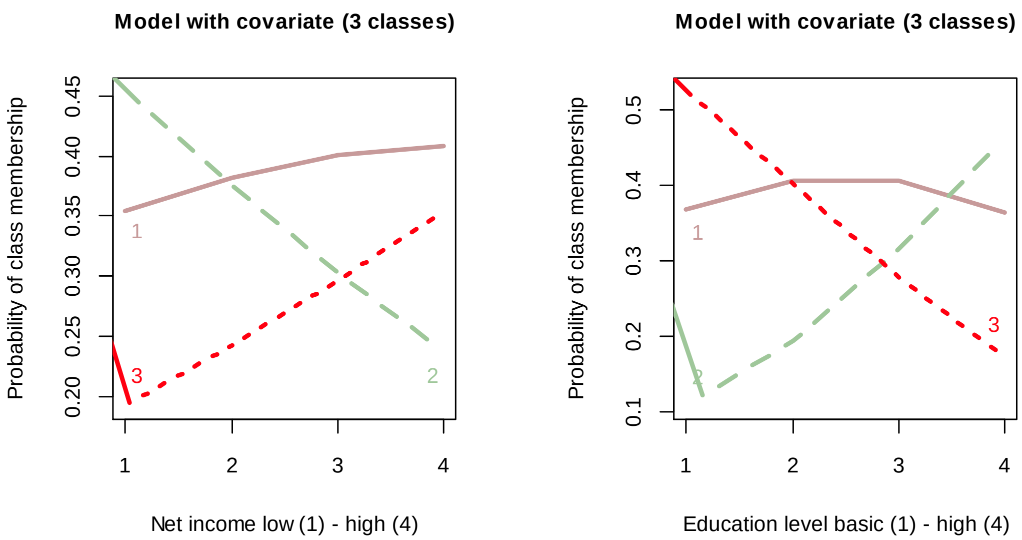


Figure 2. Estimated probabilities of selecting options for 3 classes.

Source: author's compilation using `poLCA` R package.



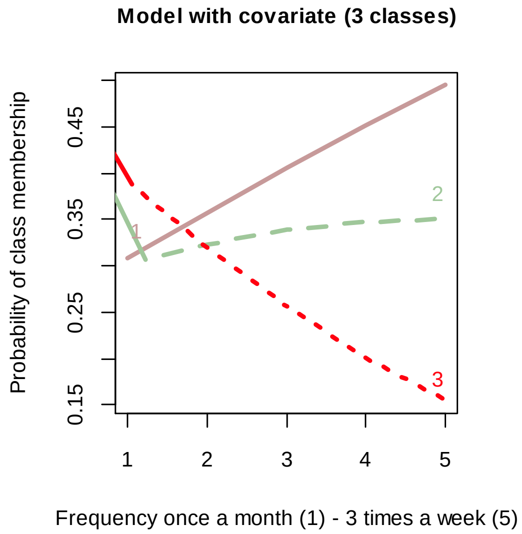


Figure 3. The influence of covariates on segments membership.

Source: author's compilation using `poLCA` R package.

The analysis of the graphs indicates that the probability of belonging to segment 1 does not depend significantly on the level of income or education. This segment is the largest and includes people who, due to lower income, pay attention mainly to such features as product quality, service and location. The increase in the frequency of using food service, which translates into higher expenses, only confirms the increase in the probability of belonging to this class.

In the case of segment 2, the probability of belonging to this segment decreases with increasing income, and any more frequent use of food services does not affect the probability of changing the segment. Segment 2 includes people with average incomes, for whom basically all variables are important, but they allow for some deviations from the ideal state (they much more prefer "average" or "high" level then "very high"). Obtaining a higher education increases the probability of belonging to this segment, which may mean stabilized preferences. In turn, an increase in income decreases the probability of belonging to the segment, which may mean migration to the class with the highest income.

Segment 3 (the smallest) are people declaring the maximum possible importance ("very high" level) of all features, including the description of food allergens and the use of organic ingredients. These are people with the highest incomes, but not necessarily with the highest education. The premise of this state is the rapidly decreasing probability of belonging to this class as a result of the increase in the level of education.

Assuming the division into 3 segments as a starting point, it is possible to learn about the belonging of each of the respondents to a given segment. In further analysis, this will allow for a detailed analysis of food service preferences in each segment separately:

```
> print(mod3$predclass)
  (1) 1 3 1 1 2 3 1 3 2 1 1 3 2 1 2 3 1 1 3 3 1 2 3 3 2 3 3 1 2 1 3 3 3 3 3
 (36) 2 2 3 1 1 2 3 1 3 2 3 1 1 3 3 3 2 3 3 2 2 1 3 2 1 1 1 2 3 2 1 1 2 3 1
 (71) 2 2 2 1 1 1 2 2 3 1 1 1 1 3 1 2 3 2 2 1 1 1 1 1 1 2 2 2 2 1 3 2 1 2 3
(106) 2 2 3 1 1 1 1 1 3 3 1 2 2 2 2 2 2 1
```

3.2. Measurement of preferences

For the purpose of detailed measurement of preferences using the conjoint method, the research identified 5 variables (attributes) of food service (with their corresponding levels): form of consumption (for here, takeaway), meal price (low, medium, high), place of consumption (bar, canteen, restaurant, food outlet), type of cuisine (polish, italian, asian, american) and menu (short, long):

```
> library(conjoint)
> full<-expand.grid(
+ form=c("for here","takeaway"),
+ price=c("low","medium","high"),
+ place=c("bar","canteen","restaurant","food outlet"),
+ cuisine=c("polish","italian","asian","american"),
+ menu=c("short","long"))
```

In the conjoint analysis method, where survey questionnaires are employed, respondents evaluate hypothetical profiles of products or services. In the research, the number of variables combined with the number of their levels enables the construction of a total of 192 different profiles of food services. Due to the respondents' limited capacity to evaluate a large number of profiles, a final set of 16 profiles was selected in a form of fractional factorial design that met the criterion of experimental orthogonality.

The building fractional factorial design and its coding in the `conjoint` package is possible using the `caFactorialDesign` and `caEncodedDesign` functions. The orthogonality of the design is validated by the identity matrix of variable correlations, as well as the appropriate determinant value of this matrix:

```
> factorial<-caFactorialDesign(full,"orthogonal")
> prof<-caEncodedDesign(factorial)
> print(prof)
    form price place cuisine menu
2      2     1     1       1    1
23     1     3     4       1    1
31     1     1     2       2    1
39     1     2     3       2    1
58     2     2     2       3    1
66     2     3     3       3    1
75     1     2     1       4    1
94     2     2     4       4    1
106    2     2     2       1    2
111    1     2     3       1    2
126    2     3     1       2    2
142    2     2     4       2    2
147    1     2     1       3    2
163    1     1     4       3    2
179    1     3     2       4    2
182    2     1     3       4    2
> print(round(cor(prof),5))
        form price place cuisine menu
```

```
form        1     0     0       0    0
price       0     1     0       0    0
place       0     0     1       0    0
cuisine     0     0     0       1    0
menu        0     0     0       0    1
> print(det(cor(prof)))
(1)  1
```

The respondents evaluated each of the profiles (cf. Table 3) on an interval scale [1-10] considering the relative attractiveness of the profiles and assigning a higher value to the profile that was more attractive to the respondents than the others. This means that the data was collected as a form of rating. With the respondent population divided into 3 segments, a detailed measurement of preferences was carried out separately for each segment.

Table 3.
Sample answers for first respondent

Number of profile	Attributes of food service					Rating [1-10]
	Form of consumption	Meal price	Place of consumption	Type of cuisine	Menu	
1	takeaway	low	bar	polish	short	10
2	for here	high	food outlet	polish	short	5
3	for here	low	canteen	italian	short	10
4	for here	medium	restaurant	italian	short	10
5	takeaway	medium	canteen	asian	short	5
6	takeaway	high	restaurant	asian	short	5
7	for here	medium	bar	american	short	3
8	takeaway	medium	food outlet	american	short	3
9	takeaway	medium	canteen	polish	long	10
10	for here	medium	restaurant	polish	long	10
11	takeaway	high	bar	italian	long	10
12	takeaway	medium	food outlet	italian	long	10
13	for here	medium	bar	asian	long	5
14	for here	low	food outlet	asian	long	5
15	for here	high	canteen	american	long	3
16	takeaway	low	restaurant	american	long	3

Source: author's compilation.

In the `conjoint` package, the experimental design, specifically the profile information (`prof`) is supplemented by a matrix containing the empirical preferences of respondents for each segment (`pref1`, `pref2`, `pref3`) and a vector with the names of the levels for all attributes (`levn`):

```
> pref1=read.csv2("preferences1.csv",header=TRUE)
> pref2=read.csv2("preferences2.csv",header=TRUE)
> pref3=read.csv2("preferences3.csv",header=TRUE)
> print(head(pref1))
  p1 p2 p3 p4 p5 p6 p7 p8 p9 p10 p11 p12 p13 p14 p15 p16
1 10  5 10 10  5  5  3  3 10  10  10  10   5   5   3   3
2 10 10 10 10  4  4  3  3 10  10   9   9   4   4   3   3
3  8  3  6  5  2  3  4  5  9   9   5   6   4   4   3   4
4  6  3 10 10  7  7  1  1  6   6  10  10   6   6   1   1
5  9  5 10  9  9  7  8  8  4   4   4   7   7   8   5   7
6  7  5  7  5  6  8  6  7  6   8   8   6   7   6   7   5

> print(levn)
        levels
```

```
1     for here
2     takeaway
3          low
4       medium
5         high
6          bar
7      canteen
8   restaurant
9  food outlet
10      polish
11     italian
12       asian
13    american
14       short
15        long
```

Having the above data sets allows obtaining the results of conjoint analysis. The estimation of the preference model is carried out using the least squares method with the help of the `caPartUtilities` function (results for the 6 respondents from segment 1):

```
> part1=caPartUtilities(pref1,prof,levn)
> print(head(part1))
     intercept for here takeaway    low medium   high    bar canteen
(1,)     6.583   -0.312    0.312  0.417  0.417 -0.833  0.313   0.312
(2,)     6.625    0.125   -0.125  0.125  0.000 -0.125 -0.125   0.125
(3,)     4.833   -0.250    0.250  0.667  0.667 -1.333  0.250   0.000
(4,)     5.625   -0.312    0.312  0.125  0.250 -0.375  0.062   0.313
(5,)     6.917    0.062   -0.062  1.583  0.083 -1.667  0.062   0.063
(6,)     6.542   -0.125    0.125 -0.292 -0.167  0.458  0.500   0.000
     restaurant food outlet polish italian  asian american  short   long
(1,)      0.312      -0.938  2.062   3.313 -1.687   -3.688 -0.313  0.313
(2,)      0.125      -0.125  3.375   2.875 -2.625   -3.625  0.125 -0.125
(3,)      0.250      -0.500  2.250   0.500 -1.750   -1.000 -0.500  0.500
(4,)      0.313      -0.688 -0.438   4.313  0.813   -4.688 -0.062  0.062
(5,)     -0.187       0.063 -1.438   0.562  0.812    0.063  1.187 -1.187
(6,)      0.000      -0.500  0.000   0.000  0.250   -0.250 -0.125  0.125
```

The summary of the results regarding part-worth utilities and attributes' importance at the segment level is possible with the `Conjoint` function (results for segment 1):

```
> Conjoint(pref1,prof,levn)

Call:
lm(formula = frml)

Residuals:
    Min      1Q  Median      3Q     Max
-5,7396 -1,6997 -0,1007  1,6840  4,9340

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         6,04861    0,09957  60,745   <2e-16 ***
factor(x$form)1     0,03472    0,09446   0,368   0,7133
factor(x$price)1    0,22222    0,14769   1,505   0,1330
factor(x$price)2    0,12500    0,12595   0,992   0,3214
factor(x$place)1    0,16319    0,16362   0,997   0,3190
factor(x$place)2   -0,10069    0,16362  -0,615   0,5385
factor(x$place)3    0,03125    0,16362   0,191   0,8486
factor(x$cuisine)1  0,38542    0,16362   2,356   0,0188 *
factor(x$cuisine)2  0,38542    0,16362   2,356   0,0188 *
factor(x$cuisine)3 -0,31597    0,16362  -1,931   0,0540 .
factor(x$menu)1     0,11458    0,09446   1,213   0,2256
```

```
---
Signif. codes:  0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Residual standard error: 2,267 on 565 degrees of freedom
Multiple R-squared:  0,04295,   Adjusted R-squared:  0,02601
F-statistic: 2,535 on 10 and 565 DF,  p-value: 0,005389

(1) "Part worths (utilities) of levels (model parameters for whole sample):"
        levnms    utls
1    intercept  6,0486
2     for here  0,0347
3     takeaway -0,0347
4          low  0,2222
5       medium   0,125
6         high -0,3472
7          bar  0,1632
8      canteen -0,1007
9   restaurant  0,0313
10 food outlet -0,0938
11      polish  0,3854
12     italian  0,3854
13       asian  -0,316
14    american -0,4549
15       short  0,1146
16        long -0,1146
(1) "Average importance of factors (attributes):"
(1)  6,67 18,71 18,89 39,95 15,77
(1) Sum of average importance:  99,99
(1) "Chart of average factors importance"
```

The application of the `caUtilities` function using the appropriate data sets (`pref1`, `pref2`, `pref3`) was repeated 3 times. The visualization of the obtained results was realized using the `ggplot` function of the `ggplot2` package:

```
> library(ggplot2)
> util1=caUtilities(pref1,prof,levn)
> h=util1(2:16)
> df=data.frame(names=levn$levels(1:15),h)
> df$type=ifelse(df$h>=0,"above","below")
> df$names<-factor(df$names,levels=rev(df$names))
> ggplot(df,aes(x=names,y=h))+
+ xlab("Atrributes' levels")+ylab("Part-worth utilities (segment 1)")+
+ geom_bar(position='stack',stat='identity',width=.9,aes(fill=type))+
+ scale_fill_manual(values=c("above"="#9fc79a","below"="#C79a9a"))+
+ theme(legend.position='none',axis.title=element_text(size=12),
+ axis.text=element_text(size=12))+
+ geom_text(aes(label=round(h,3),y=h+.00),size=4)+
+ coord_flip()+geom_hline(yintercept=0)
```

The analysis of part-worth utilities indicates that respondents' preferences regarding the levels of food service attributes are different in each of the segments. Only the perception of the menu variable is similar – in each segment, respondents at a similar level prefer a "short" menu to a "long" one (cf. Figure 4).

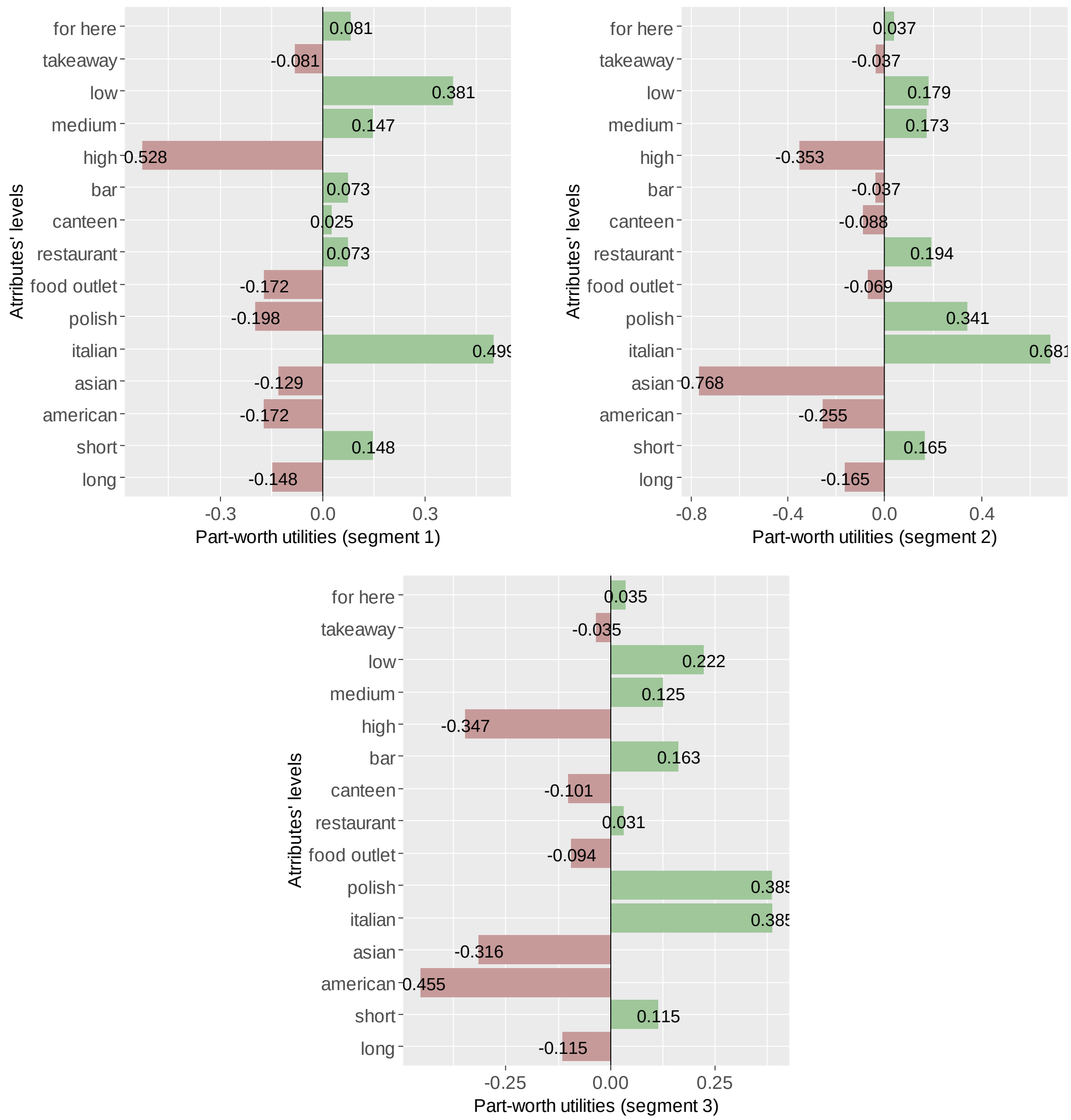


Figure 4. Part-worth utilities of attributes' levels.

Source: author's compilation using `conjoint` R package.

In segment 1, respondents exhibit a strong preference for consuming their meals "for here" rather than opting for "takeaway" options. Additionally, respondents demonstrate a marked sensitivity to pricing, with a pronounced preference for "low" prices. This is coupled with a strong rejection of "high" prices, which further reinforces the classification of this group as belonging to a lower income bracket. The aversion to "high" prices suggests that their economic situation significantly influences their dining choices, leading them to prioritize affordability above other factors. When evaluating the various levels of the variable concerning the place of consumption, it is noted that individuals in segment 1 are generally accepting of all types of dining venues except for "food outlets." This particular rejection indicates a preference for more traditional or formal dining settings rather than casual, fast-food environments. In terms of cuisine, respondents in segment 1 show clear and exclusive preference for "italian" cuisine. Their rejection of food outlets in favor of more traditional dining settings, coupled with a strong preference for italian cuisine, provides valuable insights for businesses seeking to attract this group. By focusing on affordable pricing and offering italian dishes in a comfortable and engaging dining atmosphere, establishments can effectively cater to the needs and preferences of consumers from 1 segment.

In segment 2, respondents exhibit another approach to pricing, displaying less aversion to "high" prices compared to other segments. While they show a clear preference for "low" and "average" prices, their acceptance of higher prices indicates a greater flexibility. This suggests that members of segment 2 are able to pay a more for quality ingredients, provided they feel that the expense is justified. The only acceptable places of consumption are "restaurants" offering "polish" and "italian" cuisine. The detailed results obtained for this segment, taking into account information on income, confirm the previous assumptions that this segment includes people with a solid economic situation. Understanding these characteristics allows business to tailor their offerings, ensuring they meet the expectations of this economically stable and discerning consumer group.

Segment 3 consists of respondents who demonstrate the least aversion to high prices among all segments but this does not imply they are indiscriminate in their spending. In combination with information about respondents, this is group of people with the highest income is confirmed, although these people clearly feel better in a "bar", or possibly in an informal "restaurant", than in other places of consumption. Their inclination toward bars suggests that they might prefer establishments that offer a balance of quality and comfort. People from segment 3 (similarly to segment 2) prefer "polish" and "italian" cuisine. It means that for businesses, attracting persons from segment 3 may involve creating upscale-casual dining environments.

In order to determine the importance of attributes, the `caImportance` function was used in a similar way (also for 3 segments):

```
> impo=caImportance(pref1,prof)
> df=data.frame(names=colnames(prof),impo)
> df$names<-factor(df$names,levels=c("form","price","place","cuisine","menu"))
> ggplot(df,aes(x=names,y=impo))+
+ xlab("Attributies")+ylab("Importance (%) (segment 1)")+
+ geom_bar(stat='identity',width=.9,fill="#9fc79a")+
+ theme(legend.position="none",axis.title=element_text(size=12),
+ axis.text=element_text(size=12))+
+ geom_text(aes(label=round(impo,3),y=impo+.00),size=4)+
+ geom_hline(yintercept=0)
```

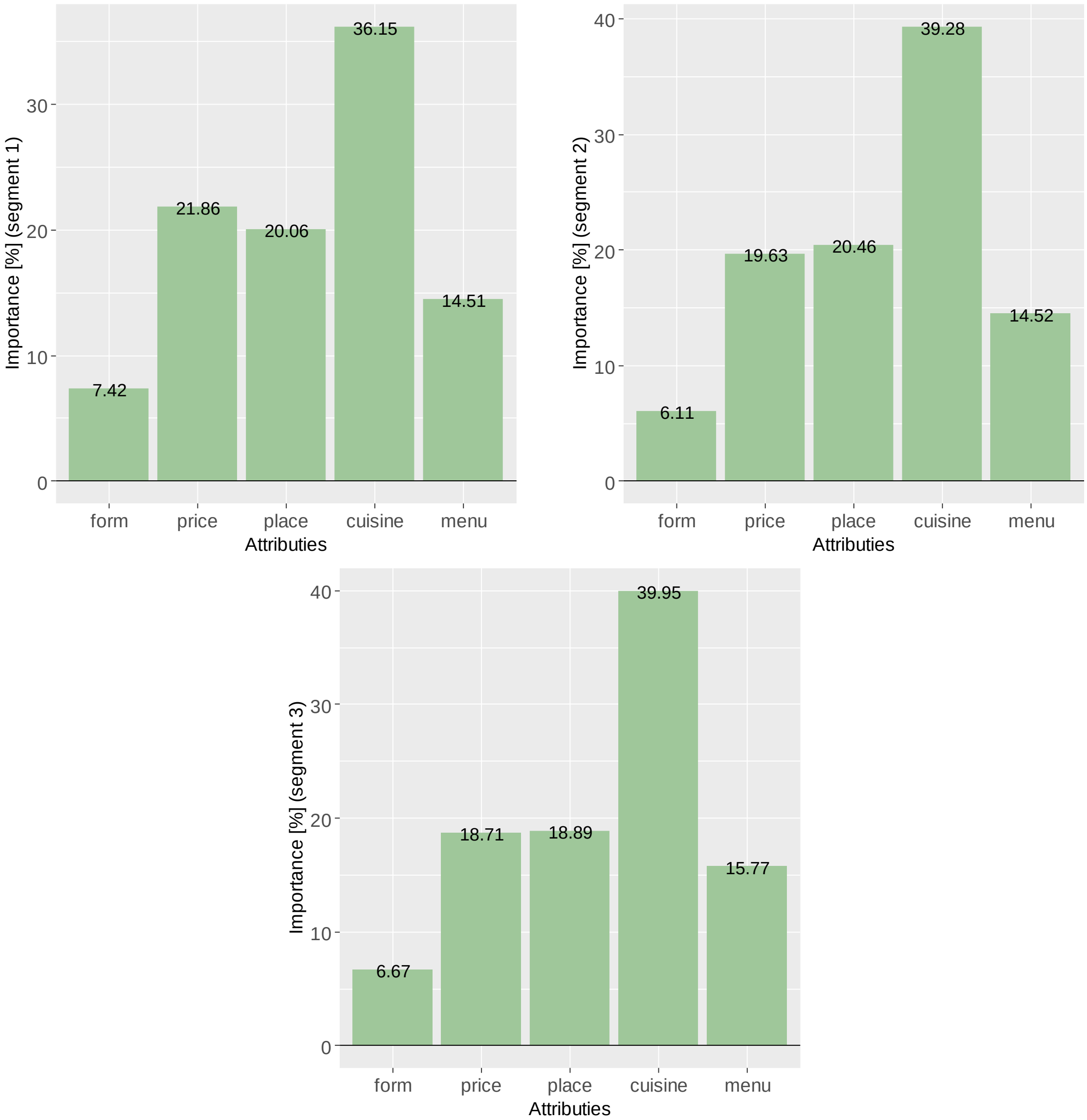


Figure 5. Importance of attributes.

Source: author's compilation using `conjoint` R package.

In examining importance of analyzed attributes across segments (cf. Figure 5), it becomes evident that there is a high degree of consistency among respondents. The type of cuisine emerges as the most significant factor. The importance of this variable is nearly identical across all segments, with respondents from segment 1 placing a 36.15% weight on cuisine type, and those from segments 2 and 3 weighing it even higher at 39.28% and 39.95%, respectively. Conversely, the form of the consumption consistently ranks as the least important attribute for all segments, with percentages ranging from 6.11% to 7.42%. This indicates that whether a consumption is "for here" or "takeaway" does not strongly influence respondents' dining decisions. The relatively low importance assigned to this variable suggests that while respondents care about what they eat (cuisine type), they are less concerned with form the consumption is served. Small differences in the perception of the importance of attributes occur between segment 1 and the others.

For respondents belonging to segment 1, the price of the meal is in second place, while in segments 2 and 3, this place is occupied by the place of consumption. These two attributes switch places of importance between the above-mentioned segments (place of consumption is in third place in segment 1, while third place in segments 2 and 3 is occupied by the price of the meal). The importance of the menu variable, like the form of consumption, is one of the least important features for all respondents.

4. Discussion

The paper explores the use of latent class analysis and conjoint analysis methods in segmentation of food service consumers based on their choice preferences. The obtained research results confirmed the value of examining food consumers' preferences within specific segments rather than at the aggregate level.

The analysis of food service preferences across all segments reveals distinct patterns shaped by income, dining priorities and the significance of certain attributes. Segment 1 is the largest group, comprised mainly of lower-income individuals who prioritize affordability, product quality, food service, and location. They prefer dining "for here" and have a strong inclination toward "italian" cuisine, avoiding "high" prices and casual "food outlets". Segment 2 represents middle-income individuals who are slightly more flexible with prices, often willing to pay more for quality. They favor "polish" and "italian" cuisine, primarily in "restaurant" settings, and maintain a balanced approach to pricing. Higher education aligns with segment 2 membership, suggests more stabilized food service preferences. Segment 3 includes high-income consumers who value all attributes, from quality of cuisine and service to the most specific attributes. This group shows the least sensitivity to price, opting bars and, to a small extent, restaurants for consumption. Across all segments, cuisine type is the most valued attribute, while the form of the meal and menu variety hold lower importance.

The results of the study confirmed earlier assumptions that perceptions of the most important factors would be similar across segments and that price is not the most important factor in choosing a food service. Respondents are willing to accept an average or even high price in exchange for high-quality products. It is also not surprising that polish and italian cuisine is the most preferred. What stands out, however, is the significant divergence between segments at a detailed level of respondents' preferences. The results confirm the need to divide the respondents into 3 separate segments, who differ fundamentally in terms of income, the possibility of going to bars (especially to food outlets), as well as aversion to asian and american cuisine. Additionally, it is noteworthy that a relatively large share – nearly 30% of respondents pays attention to allergens and organic ingredients.

Presented segmentation approach should allow food service providers to better tailor their offerings to distinct consumer groups. Businesses targeting these segments should tailor their offerings to meet the unique food service priorities of each group, from affordable italian options for segment 1 to quality-focused, upscale-casual experiences for segment 3.

5. Conclusions

The paper identifies latent consumer segments, measures their preferences and presents the integration of the poLCA and conjoint packages as complementary analytical tools. The paper shows that combining R packages effectively captures different preferences, making them useful for foodservice professionals and adaptable to other market contexts.

It is possible thanks to the possibilities (advantages) offered by the used tools - research methods and the R packages and the R environment. However, it should be noted that these solutions also have their limitations (disadvantages). The basic limitation of the conjoint analysis method, and consequently the conjoint package, is the number of possible variables to use (5-6) and their levels (3-4). In addition, the method assumes the rationality of consumer choices, which has been successfully challenged, as well as reliance on declared preferences, which may be different from the real market choices. In the case of latent class analysis, a sufficiently large research sample is needed, there are doubts to the number of classes, and the method does not offer the possibility of modeling causality. What more, computer calculations for LCA are laborious and time-consuming.

All these remarks encourage searching for other solutions, e.g. discrete choice methods and exploring some other R packages. It should be also remembered that the obtained conclusions are not timeless, which means the need to repeat the research in the future.

References

1. Aitkin, M., Anderson, D., Hinde, J. (1981). Statistical modelling of data on teaching styles. *Journal of the Royal Statistical Society, Series A (General), 144(4),* pp. 419-461.
2. Aizaki, H., Nakatani, T., Sato, K. (2015). *Stated preference methods using R*. Springer.
3. Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In: B.N. Petrov, F. Csáki (Eds.), Proceedings of the 2nd International Symposium on Information Theory (pp. 267-281).
4. Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52(3),* pp. 317-332.

5. Bąk, A. (2004). *Dekompozycyjne metody pomiaru preferencji w badaniach marketingowych.* Wrocław: Wydawnictwo Akademii Ekonomicznej we Wrocławiu.
6. Bąk, A. (2013). *Mikroekonometryczne metody badania preferencji konsumentów z wykorzystaniem programu R*. Warszawa: C.H. Beck.
7. Bąk, A., Bartłomowicz, T. (2012). Conjoint analysis method and its implementation in conjoint R package. In: J. Pociecha, R. Decker (Eds.), *Data analysis methods and its applications* (pp. 239-248). C.H. Beck.
8. Bąk, A., Bartłomowicz, T. (2018a). *Conjoint: An Implementation of Conjoint Analysis Method*. https://cran.r-project.org/web/packages/conjoint.
9. Bąk, A., Bartłomowicz, T. (2018b). *Conjoint R package*. https://pl.wikipedia.org/wiki/Conjoint_R.
10. Bartłomowicz, T., Bąk, A. (2021). *Badania preferencji i zachowań konsumentów z wykorzystaniem metod mikroekonometrii i programu R*. Wrocław: Wydawnictwo UE.
11. Bouveyron, C., Celeux, G., Murphy, T.B., Raftery, A.E. (2019). *Model-based clustering and classification for data science: With applications in R*. Cambridge University Press.
12. Brzezińska, J. (2021). *Zastosowanie analizy klas ukrytych w badaniach ekonomicznych.* Katowice: UE w Katowicach.
13. Chen, J. (1981). The EM algorithm for maximum likelihood estimates in latent class models. *Biometrika, 68(1),* pp. 209-213.
14. Colins, L.M., Lanza, S.T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. NJ: Wiley.
15. Coombs, C.H., Dawes, R.M., Tversky, A. (1970). *Mathematical psychology: an elementary introduction*. Englewood Cliffs: Prentice-Hall.
16. Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological), 39(1),* pp. 1-22.
17. Dempster, A.P., Rubin, D.B., Tsutakawa, R.K. (1981). Estimation in latent structure analysis. *Psychometrika, 46(1),* pp. 1-24.
18. DeSarbo, W.S., Wedel, M., Vriens, M., Ramaswamy, V. (1992). Latent class metric conjoint analysis. *Marketing Letters, 3(3),* pp. 273-288.
19. Everitt, B.S., Landau, S., Leese, M., Stahl, D. (2011). *Cluster Analysis*. Chichester: Wiley.
20. Everitt, B.S. (1987). *Introduction to latent variable models*. Springer.
21. Goodman, L.A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika, 61(2),* pp. 215-231.
22. Green, P.E., Rao, V.R. (1971). Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research, 8(September),* pp. 355-363.
23. Green, P.E., Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research, 5(September),* pp. 103-123.

24. Green, P.E., Srinivasan, V. (1990). Conjoint analysis in marketing: New developments with implications for research and practice. *Journal of Marketing, 54(October),* pp. 3-19.
25. Green, P.E., Wind, Y. (1973). *Multiattribute decisions in marketing. A measurement approach*. Hinsdale, Illinois: Dryden Press.
26. Green, P.E., Wind, Y. (1975). New way to measure consumers' judgments. *Harvard Business Review, 53(July/August),* pp. 107-117.
27. Gustafsson, A., Herrmann, A., Huber, F. (Eds.) (2007). *Conjoint measurement. Methods and applications*. Berlin/Heidelberg/New York: Springer-Verlag.
28. Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C. (1995). *Multivariate data analysis with readings*. Englewood Cliffs: Prentice-Hall.
29. Hamilton, J.D. (1991). A quasi-Bayesian approach to estimating parameters for mixtures of normal distributions. *Journal of Business & Economic Statistics, 9(1),* pp. 27-39.
30. IBM, SPSS (2023). *IBM SPSS Conjoint*. Chicago. https://www.ibm.com/us-en/marketplace/spss-conjoint
31. Kamakura, W.A., Russell, G.J. (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research, 26(4),* pp. 379-390.
32. Konishi, S., Kitagawa, G. (2008). *Information criteria and statistical modeling*. Springer Science & Business Media.
33. Lazarsfeld, P.E. (1950). The obligations of 1950 pollster to the 1984 historian. *Public Opinion Quarterly, 14(4),* pp. 617-638.
34. Lazarsfeld, P.E., Henry, N.W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.
35. Linzer, D.A., Lewis, J.B. (2011). poLCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software, 42(10),* pp. 1-29.
36. Linzer, D.A., Lewis, J.B. (2022). *poLCA: Polytomous Variable Latent Class Analysis*. https://cran.r-project.org/web/packages/poLCA
37. Louviere, J.J. (1988). *Analyzing decision making. Metric conjoint analysis*. Newbury Park/Beverly Hills/London/New Delhi: SAGE Publications.
38. Lu, B., Zhang, S. (2020). A conjoint approach to understanding online buyers' decisions towards online marketplaces. *Journal of Theoretical and Applied Electronic Commerce Research, 15(3),* pp. 69-83. Retrieved from: https://www.mdpi.com/0718-1876/15/3/22, doi:10.4067/S0718-18762020000300106.
39. Lubke, G., Muthén, B.O. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods, 10(1),* pp. 21-39.
40. Luce, R.D., Tukey, J.W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology, 1(1),* pp. 1-27.
41. Masyn, K.E. (2013). Latent class analysis and finite mixture modeling. In: T.D. Little (Ed.), *The Oxford Handbook of Quantitative Methods: Statistical Analysis (Vol. 2).* Oxford University Press.

42. McLachlan, G., Krishnan, T. (1997). *The EM algorithm and extensions*. NJ: Wiley.
43. McLachlan, G., Peel, D. (2000). *Finite mixture models*. NJ: Wiley.
44. Moore, W.L. (1980). Levels of aggregation in conjoint analysis: An empirical comparison. *Journal of Marketing Research, 17(4),* pp. 516-523.
45. Nylund, K.L., Asparouhov, T., Muthén, B.O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14(4),* pp. 535-569.
46. Nylund-Gibson, K., Choi, A.Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science, 4(4),* pp. 440-461.
47. Oberski, D.L. (2016). Mixture models: Latent profile and latent class analysis. In: J. Robertson, M. Kaptein (Eds.), *Modern statistical methods for HCI* (pp. 275-287). Springer.
48. Poortinga, W., Steg, L., Vlek, C., Wiersma, G. (2003). Household preferences for energy-saving measures: A conjoint analysis. *Journal of Economic Psychology, 24(1),* pp. 49-64.
49. R Development Core Team (2023). *R: A language and environment for statistical computing*. https://cran.r-project.org/. Vienna, Austria (ISBN 3-900051-07-0).
50. Rao, V.R. (2014). *Applied conjoint analysis*. Berlin/Heidelberg: Springer-Verlag.
51. Raphson, J. (1690). *Analysis Aequationum Universalis*. London: Apud Abelem Swalle.
52. Rizopoulos, D. (2006). ltm: An R package for Latent Variable Modelling and Item Response Theory Analyses. *Journal of Statistical Software, 17(5),* pp. 1-25.
53. Rizopoulos, D. (2022). *ltm: Latent Trait Models under IRT.* https://cran.r-project.org/web/packages/ltm.
54. RStudio (2023). *RStudio*: Integrated Development for R. http://www.rstudio.com/. Boston, MA.
55. Schwarz, G. (1978). Estimating the Dimension of a Model. *Gideon. The Annals of Statistics, Vol. 6, No. 2,* pp. 461-464.
56. Vermunt, J.K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis, 18(4),* pp. 450-469.
57. Vermunt, J.K., Magidson, J. (2002). Latent class cluster analysis. In: J.A. Hagenaars, A.L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89-106). Cambridge University Press.
58. Vriens, M. (2001). *Market segmentation: Analytical developments and application guidelines*. Millward Brown IntelliQuest.
59. Vriens, M., Wittink, D.R. (1994). *Conjoint analysis in marketing* [Typewritten copy].
60. Walesiak, M. (1996). *Metody analizy danych marketingowych*. Warszawa: PWN.
61. Walesiak, M., Bąk, A. (2000). *Conjoint analysis w badaniach marketingowych.* Wrocław: Wydawnictwo Akademii Ekonomicznej we Wrocławiu.

62. Wedel, M., DeSarbo, W.S. (1994). A review of recent developments in latent class regression models. In: R.P. Bagozzi (Ed.), *Advanced methods of marketing research* (pp. 353-388). Blackwell.
63. Wedel, M., Kamakura, W.A. (2000). *Market segmentation: Conceptual and methodological foundations*. Springer.
64. Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
65. Wickham, H., Winston, C., Henry, L., Pedersen, T.L., Takahasi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., van der Brand, T. (2024). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. https://cran.r-project.org/web/packages/ggplot2.
66. Wilkinson, L. (1998). *Conjoint analysis*. Chicago: SYSTAT.
67. Zwerina, K. (1997). *Discrete choice experiments in marketing*. Heidelberg/New York: Physica-Verlag.