SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2023
ORGANIZATION AND MANAGEMENT SERIES NO. 187

PREFERENCE MEASUREMENT AND SEGMENTATION OF TOURIST TRIPS CONSUMERS USING CONJOINT ANALYSIS METHOD AND CONJOINT R PACKAGE

Andrzej BĄK[1], Tomasz BARTŁOMOWICZ[2*]

[1] Wroclaw University of Economics and Business, Faculty of Economics and Finance, Department of Econometrics and Computer Science; andrzej.bak@ue.wroc.pl, ORCID: 0000-0002-5980-9694
[2] 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
[*]Correspondence author

Purpose: The paper presents the application of conjoint method and functionality of conjoint R package in measurement and analysis of preferences of tourists choosing tourist trips. The main aim of the study was to identify the key factors motivating people to make decisions regarding tourist offers. Additional aims of the research were segmentation of tourists with similar choice preferences, as well as market share simulation of trip offers not previously included in the study.
Design/methodology/approach: Traditional conjoint analysis method was used for the measurement and analysis of stated consumers' preferences. All calculations were carried out using R program and conjoint R package.
Findings: The obtained results made it possible to calculate the importance of the factors included in the study, as well as to determine the most and least preferred profile of a tourist trip at the individual level as well as for all respondents. Also, estimation of market share of the so-called simulation profiles not ranked by the respondents before and segmentation of tourists with similar choice preferences has been made.
Practical implications: Presented in the paper authoring conjoint R package is universal tool enabling the measurement and analysis of consumers' preferences of products and services, as well as political opinions and other attitudes.
Social implications: The conjoint R package implements the traditional conjoint analysis method similarly to the module Conjoint IBM SPSS program. The statistics (about half a million of downloads by RStudio users) indicate that the non-commercial conjoint package is popular among users.
Originality/value: The results of the research, as well as conjoint R package can interest of students and researchers in the field of microeconomics and marketing research in the practical application of the traditional conjoint method in analysis of stated preferences.

Keywords: behavioral economics, stated consumers preferences, conjoint analysis, R program.

JEL Classification: C6, C8, D1.

Category of the paper: research paper.

1. Introduction

Research on the rationality of market choices, considering psychological factors, has led to the emergence of new branches of research in the areas of economics and psychology. Based on economics, it is behavioral economics, and based on psychology, it is economic psychology. The co-founders of behavioral economics include D. Kahneman (Nobel Prize in 2002), A. Tversky, R. Thaler (Nobel Prize in 2017). In the research trend of behavioral economics D. Kahneman and A. Tversky formulated the prospect theory, within which utility (experienced utility or post-choice satisfaction, referring to heuristics) is analyzed, considering psychological factors (Kahneman, Tversky, 1979, 2000). R. Thaler indicated the importance of impulses and incentives addressed to consumers to induce them to make rational decisions and economic choices (Thaler, Sunstein, 2008). Direct measurement of utility is an arduous task, and the utility measurement problem has not yet been unequivocally resolved in economic theory. This results, among others, from because the utility of the good is subjective, not objective. The utility of the same goods varies across consumers. To measure utility for the purposes of empirical research in a subjective (individual) dimension, methods and models based on the concept of preferences (preference relations) are used to quantify utility across the surveyed group of consumers.

In microeconomics, a distinction is made between revealed and stated preferences. Conjoint analysis and choice-based methods are used in stated preferences research (declared at the time of conducting research).

The measurement and analysis of consumers' stated preferences uses e.g., the so-called decompositional approach (Green, Srinivasan, 1990; Zwerina, 1997; Bąk, 2004; Gustafsson, Herrmann, Huber, 2007; Aizaki, Nakatani, Sato, 2015). The basic assumption in the decompositional approach is to present the set of objects (products or services, real or hypothetical) to the respondents (e.g., in the form of a questionnaire), described using predictor variables (attributes), each of which takes certain values (levels). The main purpose of the research is to measure consumer preferences in relation to the ranked objects (profiles), which requires trade-offs. The measurement result takes the form of the set of response (outcome) variable values (empirical stated preferences measured on ordinal or interval scale). The decompositional approach uses two main groups of methods correlated with the methods and models which are different in many dimensions, however, remain similar in terms of research objectives and the obtained results' application. Conjoint analysis methods and its applications are presented e.g., in the following studies (Green, Rao, 1971; Green, Wind, 1975; Green, Srinivasan, 1978; Louviere, 1988, 1994; Green, Srinivasan, 1990; Walesiak, Bąk, 2000; Poortinga, Steg, Vlek, Wiersma, 2003; Wind, Green, 2004; Bąk, 2004, 2009; Gustafsson et al., 2007; Bąk, 2013; Rao, 2014; Bartłomowicz, Bąk, 2021; Lu, Zhang, 2020; Kim, Lee, 2023; Shim, Lee, Oh, 2022). Discrete choice methods are discussed in e.g. (Ben-Akiva, Lerman, 1985; Zwerina, 1997; Louviere, Hensher, Swait, 2000; Hensher, Rose, Greene, 2005; Garrow, 2010; Aizaki et al., 2015). The similarities and differences of conjoint analysis and discrete choice methods are demonstrated in the following studies: (Lawson, Glowa, 2000; Louviere, 2000; Louviere, Flynn, Carson, 2010; Rao, 2014). Among these methods can be distinguished traditional conjoint analysis methods (TCA) and choice-based conjoint analysis methods (CBC). Other approaches (e.g., compositional and hybrid) additionally use Adaptive Conjoint Analysis (ACA), Adaptive Choice-Based Conjoint (ACBC), Menu-Based Choice (MBC), hybrid conjoint models, as well as many other proprietary solutions.

The computer software used at the stage of factorial design, the estimation of multiple regression model with dummy variables (in the cross-section of respondents and at the entire sample level), simulation of market shares, segmentation of the respondents and visualization of the research results is required in the professional applications of conjoint methods.

The implementation of the traditional conjoint analysis method can be found primarily among commercial products in the form of either modules of popular statistical software (e.g., IBM SPSS Statistics – IBM SPSS Conjoint module, Sawtooth Software – Conjoint/Choice Software, SAS Software, XLSTAT MARKETING – XLSTAT-Conjoint module) or websites, specifically dedicated to the conjoint method in order to conduct online research.

In the case of non-commercial The R Project for Statistical Computing (R Development Core Team, 2023), the vast majority of available modules support methods other than the traditional conjoint analysis method – predominantly CBC, ACA/ACBC, MBC and other. Among the most popular packages for the analysis methods of consumer stated preference the following are listed, e.g.: mlogit (Croissant, 2022), bayesm (Rossi, 2022), DCchoice (Aizaki, Nakatani, Sato, 2022), support.CEs (Aizaki, 2022), survival (Therneau, 2023), poLCA (Linzer, Lewis, 2022). In the case of the traditional conjoint method, the following R packages are available for analyzing stated consumer preferences: conjoint (Bąk, Bartłomowicz, 2018a) and radiant.multivariate (Nijs, 2023).

The conjoint R package covers the implementation of the traditional conjoint analysis method based on the full profile method used at the stage of collecting data on the respondents' stated preference and represents a non-commercial alternative to the commercial IBM SPSS Conjoint module (SPSS, 1994; IBM, SPSS, 2023), as well as other commercial software which supports the conjoint analysis (SAS, 2023; Sawtooth Software, Inc., 2023; TIBCO, 2023). The package functions support all the research procedure stages carried out using the conjoint analysis method.

The purpose of the article is to discuss the basic assumptions of the traditional conjoint analysis method and to present the non-commercial conjoint R package (Bąk, Bartłomowicz, 2012, 2018a) for R project (R Development Core Team, 2023) supporting empirical research of stated preferences of consumers using this method.

The article also presents the results of an empirical study of the preferences of tourists coming for a holiday to a town located in the mountains (Karpacz[1]). The aim of the study is to identify important factors motivating to make a decision, such as the purpose of arrival, form of organization of the trip, winter or summer season and place of accommodation.

2. Conjoint analysis method

The traditional conjoint analysis method has over forty years of history and a well-established position among other methods measuring and analyzing stated preference of consumers representing a decompositional approach. The first publications discussing conjoint analysis were based on the theoretical research presenting conjoint measurement in psychometry. In the 1960s and 1970s the ground-breaking articles on conjoint measurement and conjoint analysis were published, e.g. (Luce, Tukey, 1964; Green, Rao, 1971; Green, Wind, 1973; Green, Srinivasan, 1978). To date, many publications have been published on the issue of conjoint analysis, presenting various methods and models of data analysis based on stated preferences, and many computer programs and websites supporting empirical research have been developed.

Conjoint analysis method is based on the axiomatic theory of measurement, originally developed at the background of psychometric studies by R.D. Luce and J. Tukey (Luce, Tukey, 1964)[2]. This theory, known in the subject literature as conjoint measurement, determines the conditions of variable measurement scales (response and predictor), in which predictor variables jointly generate the response variable values, in accordance with the specified rule of the measurement model composition (an additive rule in the traditional conjoint measurement model). This model examines the overall impact of many predictor variables on the values adopted by the response variable. It also takes into account the ordering of the response variable value at various combinations of predictor variable values. A simultaneous and additive effect of predictor variables is assumed on the response variable. Due to the response variable value measurement, considering the simultaneous impact of all predictor variables (their main effects), this measurement model is referred to as additive conjoint measurement (Coombs, Dawes, Tversky, 1970; Green, Srinivasan, 1978; Wilkinson, 1998). Therefore, conjoint measurement represents the measurement theory assuming the existence of a response variable measurement scale and measurement scales of such predictor variables allowing the quantification of joint predictor variables' impact on the response variable, according to the specific rules of model composition (cf. (Green, Srinivasan, 1978)).

[1] Karpacz town is situated in the Karkonosze Mountains. It is a spa town and ski resort in Jelenia Góra County, Lower Silesian Voivodeship, south-western Poland.
[2] The idea of nonmetric conjoint measurement was also introduced in the field of economics by G. Debreu in the work (Debreu, 1960) (Nobel Prize in 1983).

The theoretical background of conjoint measurement was developed by a psychologist and mathematician R.D. Luce and a statistician J. Tukey (Luce, Tukey, 1964; Wind, Green, 2004). An important contribution to the development of conjoint measurement was also made by the studies of J.B. Kruskal discussing the monotonic transformation of nonmetric data (Kruskal, 1964a, 1964b, 1965) and a computer program (MONANOVA), which allowed conducting experiments related to nonmetric models of additive conjoint measurement, which also significantly contributed to the development of other conjoint measurement models (Green, Wind, 1973; Wind, Green, 2004).

The first applications of conjoint analysis in the studies of stated preferences of consumers were presented in the publication by (Green, Rao, 1971) (cf. also (Green, Srinivasan, 1978; Fenwick, 1978; Hooley, Lynch, 1981)). Since then, many studies discussing methodological problems of conjoint analysis were published, including these methods' applications in marketing research. The synthetic review of the existing conjoint analysis achievements and the development perspectives of these research methods are presented in the publication dedicated to P.E. Green (Wind, Green,2004).

Currently conjoint analysis is a commonly used method to study consumer stated preferences of products and services, as well as political opinions and denominational (religious) attitudes. The basic information on conjoint analysis and software tools used in empirical research are also available on the Internet (e.g., Bąk, Bartłomowicz, 2023a, 2023b).

In accordance with the terminology used in the subject literature referring to the conjoint analysis method, predictor variables describing goods or services are called attributes[3] or factors, whereas their realizations are referred to as levels. Attributes and their levels generate different variants of goods or services, called profiles (stimuli, treatments, runs). 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).

The respondents rank product or service profiles, stating their preferences in this way. Profile ranking are referred to as total utilities and constitute the basis for further analysis. Such analysis consists in the profile decomposition of total utilities into part-worths utilities of attribute levels and in estimating the attributes' shares in the total utility development of each profile (cf. Green, Wind, 1975).

Among the most important features of traditional conjoint analysis the following are listed in the subject literature (e.g., Vriens, Wittink, 1994; Bąk, 2013):

- the number of attributes included in the study is usually limited to 6,
- the profiles presented to respondents for ranking are described using all attributes,
- the profiles are generated based on orthogonal factorial designs,

[3] The term attribute is used in statistics in relation to nonmetric variables, predominantly the nominal ones (cf. (Kendall, Buckland, 1986)).

- the profiles generated based on orthogonal designs are mutually maximally differentiated,
- the conjoint analysis model considers, apart from the main effects, also the effects of
- attributes' interactions,
- all respondents rank the same set of profiles,
- the conjoint analysis model represents the so-called decompositional approach, i.e., based on empirical total utilities profiles part-worths utilities of attribute levels are estimated.

The studies of consumer stated preferences using traditional conjoint analysis are carried out in line with the procedure presented in Table 1.

Table 1.
Conjoint analysis research procedure

Procedure stage	Procedure step
Research task specification	– response variable (empirical preferences) – predictor variables (attributes)
Model form identification	– model of predictor variables dependency (main effects or with interactions) – preference model (linear, square, parth-worths utilities)
Data collection	– data collection methods (full profiles, paired comparisons, two attributes at a time approach, simulation data) – profile generation methods (factorial designs, random sample)
Profile presentation	– presentation form (verbal description, drawing, model, physical product)) – research form (direct interview, traditional mail, phone, computer, Internet)
Preference measurement scale	– nonmetric scale – ranking – metric scale – rating
Model estimation	– nonmetric models (MONANOVA) – metric models (OLS)
Results analysis and interpretation	– preference analysis (the assessment of attributes' importance) – market share simulation – segmentation

Source: authors' compilation based on (Green, Srinivasan, 1978; Gustafsson et al., 2007; Bąk, 2004, 2013).

3. Data collection

The marketing data about the respondents' stated preferences, obtained predominantly as a result of surveys, constitute the research material used in the conjoint analysis methods. Collecting data is one of the main stages in the entire research procedure. The selection of data collection method determines the computational complexity of the parameter estimation task in the conjoint analysis model, and thus influences the nature of techniques possible to apply in estimating the value of part-worths utilities. Moreover, the method of data collection has a decisive impact on the credibility level of the rankings made by the respondents (cf. Vriens, Wittink, 1994; Bąk, 2004).

In the subject literature on conjoint analysis the following data collection methods are most often listed:

- full-profile method (the method refers to a traditional conjoint analysis),
- method of paired comparisons,
- two-attributes-at-a-time-approach (the method using the compromise matrix, the method of presenting two attributes simultaneously),
- full-profiles choice experiments (the method refers to a choice-based approach).

In a traditional conjoint analysis, the full-profile method or full-concept method are used to cover the set of all possible variants, being the combination of attributes and their levels.

In the full profile method the respondent ranks the presented variants, according to his/her own preferences stated based on the presented attributes and their levels, in terms of determining the rank order of profiles (on the ordinal scale – ranking) or determining the relative attractiveness of profiles (e.g., on the positional scale – rating).

The positive features of this method mainly include presenting the respondents with the profiles of products or services to be ranked, characterized by all the selected attributes at the same time. It is actually the situation encountered by the consumer in real life while making the specific choices among products (services) available on the market. If all the attributes are presented simultaneously, one can also take into account all the interactions occurring between them, which can generate certain synergistic effects, invisible in a different situation. The advantages of this method also include the possibility of choosing the scale of response variable value, because the preference measurement can be carried out on an ordinal, interval, or quotient scale (cf. Vriens, Wittink, 1994).

The most serious shortcoming of the method is the limited number of attributes and levels, which can be included in the designed experiment. The number of profiles (P) presented to the respondent equals the product of levels of individual attributes, i.e.: $P = \prod_{j=1}^{m} L_j$ (where: L_j – number of levels of j-th attribute; m – number of attributes describing profiles evaluated by consumers). This number can take on large values, i.e., exceeding the possibility of making a precise and reliable assessment. Therefore, in addition to the criteria for substantive selection of the attributes and levels, the statistical experiment planning systems are used in this method, which allow the reduction of the potential number of profiles.

In practice, the full factorial design can only be considered with a very small number of attributes and levels. The number of all profiles in the full factorial design results from the product of the number of attributes' levels and generally takes large values. For example, in case of $4^1 3^2 2^3$ design[4] the relevant number amounts 288 profiles. It is not possible for the respondents to rank reliably enough such a large number of profiles; therefore, it is necessary to reduce the complete set of profiles to a reasonable size using a fractional factorial design. Then, the so-called incomplete (partial, reduced, fractional factorial design) factorial

[4] The notation $4^1 3^2 2^3$ means 1 attribute with 4 levels, 2 attributes with 3 levels and 3 attributes with 2 levels.

experiment is designed, which considers only a representative subset of profiles. Using the fractional factorial design scheme, the size of the full design consisting of 288 profiles can be reduced, e.g., to 16 profiles.

The problems of designing factorial experiments have been discussed in e.g., the following studies: (Louviere, 1988; Vriens, Wittink, 1994; Kuhfeld, Tobias, Garratt, 1994; Rasch, Herrendörfer, 1991; Huber, Zwerina, 1996; Zwerina, 1997; Ott, 1984; Zwerina, Huber, Kuhfeld, 2000).

The review of current achievements in planning factorial experiments for conjoint analysis models and discrete choice models is presented in the study (Großmann, Holling, Schwabe, 2002).

In conjoint R package both full and fractional design (orthogonal and eff ective) can be generated using AlgDesign R package functions (Wheeler, 2022).

4. Conjoint model estimation

The purpose of estimating the parameters of conjoint analysis model is to assess the values of attribute levels interpreted here as the so-called part-worth utilities of attribute levels. Part-worths utilities are estimated for each respondent individually and as average values for the studied sample. Determining part-worths utilities allows, in turn, to carry out the analysis regarding (cf. Table 1):

- theoretical total utilities of profiles in the cross-section of respondents,
- theoretical total utilities of profiles in the analyzed sample,
- theoretical total utilities of profiles in the separated groups (segments) of respondents,
- the assessment of relative „importance" of individual attributes in the cross-section of respondents in the analyzed sample,
- the segmentation of respondents (Wedel, Kamakura, 2000).

In the traditional conjoint analysis, the linear multiple regression model is developed, the parameters of which (part-worths of levels attributes) are estimated using the classical Ordinary Least Squares (OLS) method.

In the multiple regression analysis, the response variable takes values (e.g., points or ranks) assigned by a given respondent to individual profiles submitted for assessment. The influence of each level of individual predictor variables (nonmetric attributes) on the rating assigned to profiles by a given respondent is taken into account by introducing dummy predictor variables to the regression model.

The linear additive multiple regression model of conjoint analysis model is presented, in general (taking into account the actual attributes of products or services), by the following formula:

$$Y = \beta_0 + \sum_{k=1}^{p} \beta_k Z_k + \varepsilon \qquad (1)$$

where:

Y – response variable, taking the values representing the respondents empirical preferences,

β_0 – model intercept;

$\beta_1, \ldots, \beta_p$ – model parameters;

$Z_1, \ldots, Z_p$ – predictor variables (the attributes describing profiles of products or services);

$k = 1, \ldots, p$ – predictor variable (attribute) number;

ε – model random component.

Next the nonmetric attributes $Z_1, \ldots, Z_p$ are encoded using dummy variables, which indicate the occurrence of particular attribute levels in individual profiles. For this purpose, indicator (dummy) coding, effects coding, deviations from means coding or orthogonal coding are used (cf. (Zwerina, 1997; Walesiak, Bąk, 2000; Bąk, 2004)). The coding results in substituting p attributes $(Z_1, \ldots, Z_p)$ with dummy variables $(X_1, \ldots, X_m)$[5], the number of which is $m = \sum_{\mathrm{k}=1}^{p} L_k - p$, where: L_k – number of levels of k-th attribute. Thus, it results that in order to encode all levels of a given attribute, the number of dummy variables by 1 less than the number of this attribute levels is sufficient, as shown in Tab. 2. Variable X_3 in both encoding methods is redundant (gray cells in Table 2), because every level of Z attribute is clearly indicated using X_1 and X_2 dummy variables.

Table 2.
Coding attributes using dummy variables

Attribute	**Dummy variables**					
	indicator coding			**effects coding**		
Z_1	X_1	X_2	X_3	X_1	X_2	X_3
level 1	1	0	0	1	0	0
level 2	0	1	0	0	1	0
level 3	0	0	0	–1	–1	–1

Source: authors' compilation.

Including the redundant variables in the model increases the phenomenon of collinearity, which affects the quality of the estimated regression model. Therefore 2 dummy variables are used in the conjoint analysis model in the case of 3-level attribute, whereas the third level serves as the so-called reference level. After transcoding the attributes, the conjoint analysis model with dummy variables can be presented in the following form:

$$\hat{Y} = b_0 + \sum_{j=1}^{m} b_j X_j \qquad (2)$$

[5] The values of dummy variables depend on the coding method, e.g. in the case of indicator (dummy) coding the respective values are 0 and 1, and in the case of effect coding, deviations from means coding they are 0, 1 and −1.

where:
$\hat{Y}$ – theoretical values of the response variable,
b_0 – model intercept;
$b_1, \dots, b_m$ – model parameters;
$X_1, \dots, X_m$ – dummy variables representing nonmetric attribute levels;
$j = 1, \dots, m$ – dummy variable number.

Model (2) is estimated at an aggregated level jest (in the cross-section of all respondents constituting the analyzed sample). Conjoint analysis models are also estimated at an individual level (for each respondent individually).

The linear regression model for the selected respondent can be presented in the following form:

$$\widehat{Y_s} = b_{0s} + b_{1s}X_{1s} + \cdots + b_{ms}X_{ms} = b_{0s} + \sum_{j=1}^{m} b_{js}X_j \tag{3}$$

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

As a result of model estimation (2) the values of $b_1, \dots, b_m$ parameters are obtained and interpreted as part-worths utilities of attribute levels. Part-worths utilities of reference levels (related to dummy variables skipped in the coding process) are calculated depending on the adopted coding method. Table 3 presents the method for calculating part-worths utilities for the 3-level attribute in the case of indicator (dummy) coding and eff ects coding, deviations from means coding taking into account the reference level (level 3, b_3X_3, U_3) in gray cells.

Table 3.
The method for calculating part-worths utilities for the 3-level attribute

Attribute	**Dummy variables**					
	indicator coding			**eff ects coding**		
Z_1	b_1X_1	b_2X_2	b_3X_3	b_1X_1	b_2X_2	b_3X_3
level 1	b_1	0	$U_1 = b_1$	b_1	0	$U_1 = b_1$
level 2	0	b_2	$U_2 = b_2$	0	b_2	$U_2 = b_2$
level 3	0	0	$U_3 = 0$	$-b_1$	$-b_2$	$U_3 =-(b_1 + b_2)$

U_1, U_2, U_3 – part-worths utilities of levels of attribute Z_1; X_1, X_2, X_3 –dummy variables; b_1, b_2, b_3 – part-worths utilities.
Source: authors' compilation.

Part-worths utilities are calculated at an aggregated level (one model is estimated for the whole sample) and at an individual one (the number of estimated models equals the number of respondents). The knowledge of part-worths utilities allows estimating theoretical total utilities of the profiles being the subject of research. The total utility of i-th profile for s-th respondent (U_i^s) is calculated based on the following formula (Walesiak, 1996):

$$U_i^s = \sum_{j=1}^{m} b_{0s} + U_{l_j^i}^s, \tag{4}$$

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 (cf. (Walesiak, 1996)):

$$U_i = \frac{1}{S}\sum_{s=1}^{S}\left(\sum_{j=1}^{m} b_{0s} + U_{l_j^i}^s\right). \tag{5}$$

The knowledge of part-worths utilities also allows estimating the so-called attribute „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) (cf. (Hair, Anderson, Tatham, Black, 1995)):

$$W_j^s = \frac{max\{U_{l_j^i}^s\} - min\{U_{l_j^i}^s\}}{\sum_{j=1}^{m}\left(max\{U_{l_j^i}^s\} - min\{U_{l_j^i}^s\}\right)} \times 100\%. \tag{6}$$

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:

$$W_j = \frac{1}{S}\sum_{s=1}^{S} W_j^s, \tag{7}$$

where: W_j^s – defined by a formula (6).

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. The simulation analysis of market shares allows estimating the total utility of additional profiles, which were not ranked by the respondents in the questionnaire. The anticipated market share of the selected profiles is estimated based on the following models (cf. (Hair et al., 1995; Walesiak, 1996)):

- maximum utility model, used in calculating the percentage of respondents for which a particular product received the highest total utility score, among the products covered by the simulation:

$$P_i^s = \begin{cases} 1, & \text{if } \widehat{U}_i^s = max\left(\widehat{U}_i^s\right) \\ 0, & \text{otherwise} \end{cases}, \tag{8}$$

where: P_i^s – the probability of i-th profile selection by s-th respondent,

- probabilistic BTL (Bradley-Terry-Luce Model), following which the total utility, corresponding to a given profile, is divided by the sum of total utilities of profiles covered by the simulation (the calculations are carried out separately for each respondent and next their average value is computed):

$$P_i^s = \frac{\hat{U}_i^s}{\sum_{i=1}^{n} \hat{U}_i^s}, \quad (9)$$

where: n – number of profiles;

- logit model, in which the calculations, as opposed to the probabilistic BTL model, use natural logarithms of total utilities' values rather than the utilities themselves:

$$P_i^s = \frac{e^{\hat{U}_i^s}}{\sum_{i=1}^{n} e^{\hat{U}_i^s}} = \frac{exp(\hat{U}_i^s)}{exp(\sum_{i=1}^{n} \hat{U}_i^s)}. \quad (10)$$

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

In the practice of segmentation studies, using conjoint analysis methods, the post hoc approach is most frequently used, which applies data classification methods (cluster analysis) in the division of respondent's set into classes (segments), based on individual part-worths utilities, representing the heterogeneity of preferences. Due to certain specific features (unequivocal qualification of objects into groups, eff ective processing of large data sets) the *k*-means method is frequently used, which belongs to the group of iterative optimization methods.

5. Overview of the functions in the conjoint package

The conjoint R package (Bąk, Bartłomowicz, 2018a) is an implementation of the traditional conjoint analysis method. The source code of the package was written in R language and represents an extension of the computational tools off er in microeconometrics and consumers' preference studies available under the terms of GNU license („free and open software")[6], thus free of charge and providing access to the source code. The correct functioning of the package requires installing the base GNU R program (R Development Core Team, 2023) and a dozen additional packages which, starting from 3.3.2 version of R base program, are downloaded, and installed along with the conjoint package. The package can be downloaded and installed from the CRAN R project repository website (https://cran.r-project.org/package=conjoint) or from the GitHub website https://github.com/packagesR/conjoint). The conjoint package has been available on the CRAN R repository website (R Development Core Team, 2023) since October 2011.

[6] The GNU General Public License is intended to guarantee your freedom to share and change all versions of a program – to make sure it remains free software for all its users, https://www.gnu.org.

The current version of conjoint (1.41) package offers 16 functions which allow for: model parameters estimation of conjoint analysis model and the segmentation of respondents (functions: caModel, caSegmentation), estimation of part-worths utilities and theoretical total utilities in the cross-section of respondents (functions: caPartUtilities, caTotalUtilities), measurement of attributes' importance and part-worths utilities of attributes' levels at an aggregated level (functions: caImportance, caUtilities), and also - within the framework of simulation analysis – market share estimations of simulation profiles (functions: caBTL, caLogit and caMaxUtility). The special purpose functions include the function converting the empirical preference data set (function caRankToScore) and the functions which allow obtaining the aggregate results of conjoint analysis and simulations (functions: Conjoint, ShowAllUtilities and ShowAllSimulations).

In addition, the package offers tools supporting the design of a questionnaire survey. The package includes functions creating the appropriate factorial designs, allowing the reduction of the complete set of profiles in the form of fractional designs (orthogonal and effective). For this purpose, the conjoint package uses AlgDesign package (Wheeler, 2022) provided in CRAN R (R Development Core Team, 2023) repository. The application of the selected AlgDesign package functions in conjoint package is carried out in the form of functions which allow e.g., obtaining orthogonal or effective fractional factorial designs (functions: caFactorialDesign, caEncodedDesign and caRecreatedDesign) and their coding using dummy variables (effects coding, deviations from means coding is applied). It means the possibility of designing an experiment to be implemented in the form of a questionnaire survey (using indirect and direct methods of collecting data from primary sources, e.g., direct, or online surveys). In order to generate the relevant fractional factorial design, the data on the number of attributes (factors) taken into account and their levels, including names, are sufficient. Fractional designs are presented in two versions: with the names of levels (questionnaire version) and with the numbers of levels (version for further calculations).

The Table 4 presents the concise description of the purpose of conjoint package function and Table 5 meaning of function arguments. The detailed features of all available functions, data sets and practical examples of the package application in measuring consumers' stated preferences is included in the documentation of conjoint R package.

Table 4.
The functions of conjoint *R package (version 1.41)*

Functions of the conjoint package
caFactorialDesign(data, type="null", cards=NA, seed=123) – the function generates full or fractional factorial design maintaining the names of variables and levels
caEncodedDesign(design) – the function encodes the experiment design obtained using caFactorialDesign function for the needs of conjoint package functioning
caRecreatedDesign(attr.names, lev.numbers, z, prof.numbers) – the function recreates the fractional factorial design based on the number of profiles from the full factorial design
caRankToScore(y.rank) – the function transforms the empirical preference data measured on a rank scale into a data set in the form of point grades (on a positional scale)

Cont. table 4.

caPartUtilities(y, x, z) – the function calculates the part-worths utility matrix of attribute levels in the cross-section of respondents (including an intercept)
caTotalUtilities(y, x) – the function calculates the theoretical total utilities matrix of profiles in the cross-section of respondents
caImportance(y, x) – the function calculates an average relative „importance" of all attributes (as %) at an aggregated level
caUtilities(y, x, z) – the function calculates part-worths utilities of attribute levels at an aggregated level
caBTL(sym, y, x) – the function estimates market shares of simulation profiles based on the BLT probability model (Bradley-Terry-Luce Model)
caLogit(sym, y, x) – the function estimates market shares of simulation profiles based on logit model
caMaxUtility(sym, y, x) – the function estimates market shares of simulation profiles based on the maximum utility model
caSegmentation(y, x, c=2) – the function caries out respondents' segmentation using *k*-means method based on kmeans function
caModel(y, x) – the function estimates conjoint analysis model parameters
Conjoint(y, x, z, y.type="score") – the function calculates basic results of conjoint analysis at an aggregated level
ShowAllUtilities(y, x, z) – the function calculates all utilities available in the conjoint package (part-worths and total)
ShowAllSimulations(sym, y, x) – the function estimates market shares of simulation profiles based on all simulation models available in conjoint package

Source: authors' compilation.

Table 5.

The functions arguments of conjoint *R package (version 1.41)*

Argument name	Argument meaning
data	data describing the object of an experiment (product, service) – the set of attributes (factors) and their levels in the form of expand.grid function
type	optional parameter describing the type of generated factorial design (default type="null" – fractional design is generated with no specific criteria)
cards	optional parameter describing the number of generated profiles (default cards=NA – the number of profiles results from the type of generated factorial design)
seed	optional parameter describing the seed value of the random number generator (default seed = 123)
design	factorial (fractional or full) experiment design
attr.names	vector representing names of attributes (factors)
lev.numbers	vector representing numbers of attributes' (factors) levels
prof.numbers	vector representing numbers of reconstructed profiles
z	vector representing names of attributes' (factors) levels
y.rank	matrix (or vector) of empirical preferences in the ranking form (the ranking data require transformation to rating data using caRankToScore function)
y	matrix (or vector) of empirical preferences (in the form of importance assessments on a rating or ranking scale)
x	matrix representing profiles (including names of attributes)
y.type	type of data about preferences – data in the form of profile importance assessments on a rating or ranking scale (default type is rating)
sym	matrix representing simulation profiles (including attributes' names)
c	optional parameter specifying the number of segments (default c = 2 – division into 2 segments)

Source: authors' compilation.

6. Application of the conjoint R package

The conjoint package was used in an empirical study of the stated preferences of tourists choosing a place and form of recreation. The main aim of the research was to identify the factors (attributes) that guide tourists when choosing a trip from among many offered on the market. Additional aims of the research were segmentation of tourists with similar choice preferences, as well as forecasting the market share of trip offers not previously included in the study. The following features, along with the respective levels, were listed in the set of variables describing the examined product – tourism trips: purpose (cognitive, vacation, health, business), form (organized, own), season (summer, winter), accommodation (1-2-3 star hotel, 4-5 star hotel, guesthouse, hostel). Due to too many profiles resulting from the combination of levels of all features (in this case the so-called full factorial design consists of 64 profiles[7]), the following fractional factorial design of 14 profiles was used:

```
> library(conjoint)
> data(journey)
> journey<-expand.grid(
+ purpose=c("cognitive","vacation","health","business"),
+ form=c("organized","own"),
+ season=c("summer","winter"),
+ accommodation=c("1-2-3 star hotel","4-5 star hotel","guesthouse","hostel"))
> jprof<-caFactorialDesign(data=journey,type="fractional")
> print(jprof)
 purpose form season accommodation
1 cognitive organized summer 1-2-3 star hotel
8 business own summer 1-2-3 star hotel
10 vacation organized winter 1-2-3 star hotel
15 health own winter 1-2-3 star hotel
19 health organized summer 4-5 star hotel
21 cognitive own summer 4-5 star hotel
30 vacation own winter 4-5 star hotel
34 vacation organized summer guesthouse
39 health own summer guesthouse
41 cognitive organized winter guesthouse
48 business own winter guesthouse
54 vacation own summer hostel
60 business organized winter hostel
61 cognitive own winter hostel
```

Respondents rated profiles according to their preferences. Data in the form of ratings on an interval scale[8] were collected using questionnaires sent electronically[9]. Of all the surveys, in the research 306 responses were included – 166 from women and 140 from men. Ratings of all 14 profiles by 6 first respondents are as follows:

[7] The number of profiles is the product of the number of all attribute levels ($4^2 2^2 = 64$).

[8] Rating on an interval scale means the valuation of profiles within the adopted interval (in the research the scale takes $[0 - 10]$ interval).

[9] The empirical data were collected using a questionnaire presented on the website https://www.webankieta.pl/. The study was carried out by Mateusz Gordzicz for the needs of his Master's thesis. The respondents' age: less than 20 years of age – 10%, 20-40 – 85%, more than 40 – 5%.

```
> head(jpref)
 profile01 profile02 profile03 profile04 profile05 profile06 profile07
1 0 10 0 10 10 8 4
2 10 0 10 3 7 9 2
3 8 2 6 9 7 9 0
4 8 10 1 6 3 0 3
5 3 4 8 10 10 1 10
6 5 1 8 3 10 0 9
 profile08 profile09 profile10 profile11 profile12 profile13 profile14
1 5 10 2 4 0 0 6
2 7 4 0 8 10 3 7
3 1 8 5 0 0 0 5
4 1 8 4 7 4 1 10
5 4 9 4 10 0 7 10
6 5 3 10 10 4 1 8
```

Having the set of data about empirical preferences (jpref), the study design (jprof), the names of variables and their levels (jlevn) it is possible to estimate part-worths utilities using conjoint package. The part-worths utilities determine the relative importance, which the particular levels of attributes have in total utilities. The estimation of part-worths utilities is carried out by decomposing total utilities stated by the respondents. For this purpose a linear regression model with dummy variables is estimated for each respondent using formula (3), in which the response variable is the empirical total utility allocated by s-th ($s = 1,2,3,\ldots,306$) respondent to the particular profiles. The aggregate model for the whole sample is also estimated based on formula (2).

Model (3) is estimated using the least squares method and the values of parameters b_{is} ($i = 1, 2, \ldots, 12; s = 1, 2, \ldots, 306$) are obtained, which allows calculating all part-worths utilities in the cross-section of respondents using caPartUtilities function:

```
> print(head(caPartUtilities(jpref,jprof,jlevn)))
intercept cognitive vacation health business organized own
[1,] 4.938 -0.937 -2.687 3.639 -0.014 -1.563 1.563
[2,] 5.625 0.875 1.625 -0.827 -1.673 0.250 -0.250
[3,] 4.188 2.562 -2.438 3.341 -3.466 0.063 -0.063
[4,] 4.375 1.125 -2.125 0.788 0.212 -1.625 1.625
[5,] 6.688 -2.187 -1.188 3.534 -0.159 -0.062 0.062
[6,] 5.500 0.250 1.000 0.202 -1.452 0.750 -0.750
summer winter 1-2-3 star_hotel 4-5 star_hotel guesthouse hostel
[1,] 0.692 -0.692 0.063 1.639 0.313 -2.014
[2,] 1.058 -1.058 0.125 -0.452 -0.875 1.202
[3,] 0.135 -0.135 2.062 -0.034 -0.688 -1.341
[4,] 0.346 -0.346 1.875 -2.962 0.625 0.462
[5,] -2.385 2.385 -0.437 1.034 0.062 -0.659
[6,] -1.808 1.808 -1.250 1.202 1.500 -1.452
```

Part-worths utilities determine the relative contribution of individual attribute levels to the total profile utility. This contribution is interpreted in accordance with the value preference principle, i.e., the higher the part-worths utility, the more the given attribute level is appreciated by the respondent. For example, in case of the respondent no. 1 the estimations of part-worths utilities are as follows:

- for purpose attribute: $b_1 = -0.937$ (cognitive), $b_2 = -2.687$ (vacation), $b_3 = 3.639$ (health), $b_4 = -(b_1 + b_2 + b_3) = -0.014$ (business),
- for form attribute: $b_5 = -1.562$ (organized), $b_6 = -(b_5) = 1.562$ (own),
- for season attribute: $b_7 = 0.692$ (summer), $b_8 = -(b_7) = -0.692$ (winter),
- for accommodation attribute: $b_9 = 0.063$ (1-2-3 star hotel), $b_{10} = 1.639$ (4-5 star hotel), $b_{11} = 0.312$ (guesthouse), $b_{12} = -(b_9 + b_{10} + b_{11}) = -2.014$ (hostel).

It means that first respondent prefers the most health purpose and own trip, at summer in 4-5 star hotel. It should be noted that there is a differentiation between the respondents in part-worths utilities ranking of individual attributes. It means that not all respondents value equally the individual levels of features, which manifests the heterogeneity of preferences.

The estimation of model (2) parameters in the cross-section of whole sample (all respondents) can be carried out using Conjoint function:

```
> Conjoint(jpref,jprof,jlevn)
Call:
lm(formula = frml)
Residuals:
 Min 1Q Median 3Q Max
-5,4460 -3,0144 -0,0949 2,7758 5,9051

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 4,979371 0,052578 94,704 < 2e-16 ***
factor(x$purpose)1 0,139093 0,084780 1,641 0,1009
factor(x$purpose)2 0,146446 0,084780 1,727 0,0842 .
factor(x$purpose)3 0,437924 0,097823 4,477 7,78e-06 ***
factor(x$form)1 -0,070057 0,052578 -1,332 0,1828
factor(x$season)1 -0,094834 0,052172 -1,818 0,0692 .
factor(x$accommodation)1 -0,136234 0,084780 -1,607 0,1081
factor(x$accommodation)2 -0,028171 0,097823 -0,288 0,7734
factor(x$accommodation)3 0,005923 0,084780 0,070 0,9443
---
Signif. codes: 0 ‘***’ 0,001 ‘**’ 0,01 ‘*’ 0,05 ‘.’ 0,1 ‘ ’ 1

Residual standard error: 3,291 on 4275 degrees of freedom
Multiple R-squared: 0,01474, Adjusted R-squared: 0,0129
F-statistic: 7,994 on 8 and 4275 DF, p-value: 9,444e-11

[1] "Part worths (utilities) of levels (model parameters for whole sample):"
levnms utls
1 intercept 4,9794
2 cognitive 0,1391
3 vacation 0,1464
4 health 0,4379
5 business -0,7235
6 organized -0,0701
7 own 0,0701
8 summer -0,0948
9 winter 0,0948
10 1-2-3 star_hotel -0,1362
11 4-5 star_hotel -0,0282
12 guesthouse 0,0059
13 hostel 0,1585
```

[1] "Average importance of factors (attributes):"
[1] 38,62 13,30 13,97 34,11
[1] Sum of average importance: 100
[1] "Chart of average factors importance"

The obtained results indicate that at an aggregate level (for all respondents), the most popular are health purpose trips focused on regenerating physical or mental condition of tourists, formed on their own trip and in winter, with hostel as an accommodation (Figure 1). At the same time, trip purpose and accommodation type seem to be the most important among the attributes used in the example, followed by the season of the year, whereas the form (organized, own) of the trip seems to be the least important. These results are illustrated by the chart of attributes' importance (Figure 2).

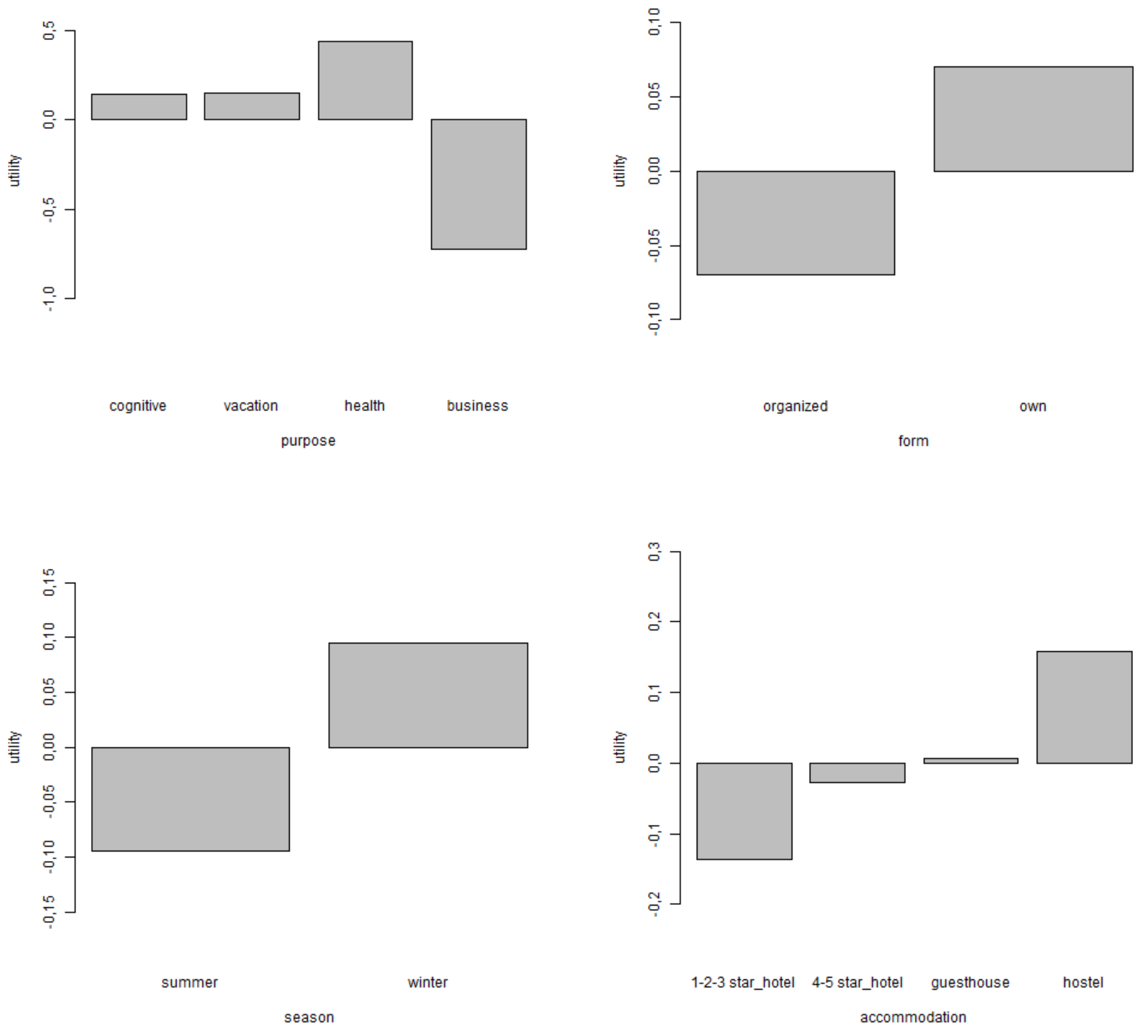

Figure 1. The chart of attributes' part-worths utilities.

Source: authors' compilation using conjoint R package.

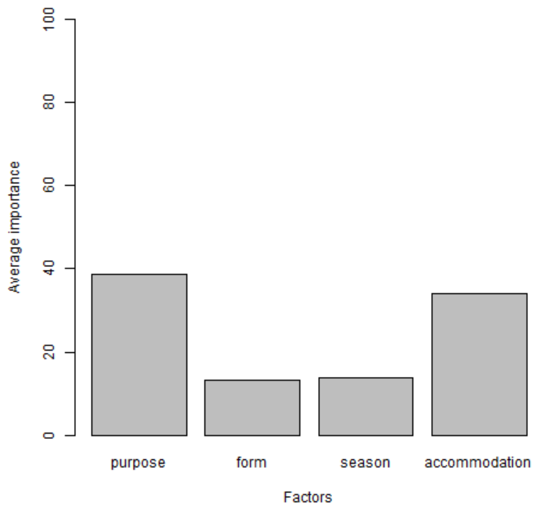

Figure 2. The chart of attributes' importance.

Source: authors' compilation using conjoint R package.

The conjoint package allows estimating market shares of the so-called simulation profiles, i.e., the profiles which were not ranked by the respondents before. Based on the analysis of the conjoint model estimation results for the whole analyzed sample (306 respondents ranking 14 tourist product profiles), 5 trip variants were chosen for simulation analysis, which were not included in the survey questionnaire. The selection of variants was carried out taking into account the average importance of features and their levels, following the trade-off principle. Profile no. 3 offers the majority of the desired features (health oriented purpose of the trip, own organization form and winter season) combined with accommodation in a 4-5-star hotel. Profile no. 2 does not offer any of the preferred features, and the other profiles one each – profile no. 1 own trip organization form, profile no. 4 – hostel as the form of accommodation, whereas profile no. 5 – trip in winter season:

```
> print(jsimp)
purpose form season accommodation
1 2 2 1 1
2 2 1 1 2
3 3 2 2 2
4 1 1 1 4
5 4 1 2 3
```

The total utility (attractiveness) of the simulation variants for all respondents was calculated using maximum utility models, the probabilistic BTL (Bradley-Terry-Luce Model) model and the logit model:

```
> ShowAllSimulations(sym=jsimp,y=jpref,x=jprof)
TotalUtility MaxUtility BTLmodel LogitModel
1 4,96 20,26 19,31 17,51
2 4,93 11,44 20,01 15,72
3 5,55 31,05 22,32 29,02
4 5,11 24,84 20,77 23,07
5 4,29 12,42 17,59 14,68
```

From among the selected trip variants subject to simulation analysis, the largest market share (according to all models – maximum utility, BTL and logit) is expected for profile no. 3. The smallest market share (according to the BTL and logit model) is expected for profile no. 5 and (according to the maximum utility model) – for the model no. 2. The comparison of relevant profiles confirms the respondents' preferences regarding the desirable features and indicates that respondents are able to accept the levels of some features (e.g., a 4-5-star hotel) in exchange for the other preferred attributes (profile no. 3).

In order to perform respondents' segmentation on the basis of estimated part-worths utilities using caPartUtilities function (individual models – one model for each respondent), the conjoint package off ers caSegmenation function using k-means method, which allows the division of respondents into the indicated number of segments (this number must take the value of 2 or higher).

Determining the relevant number of segments was carried out using a NbClust package (Charrad, Ghazzali, Boiteau, Niknafs, 2014). The NbClust package allows you to estimate 30 indexes indicating the optimal number of clusters for various methods of partitioning a data set, including the k-means method. In the procedure of selecting the optimal number of clusters, 17 indices were estimated. Clustering validity indices indicate the division of respondents into 2 (6 indices) or 11 segments (5 indices).

Figures 3 and 4 present the visualizations of segments (division of 306 respondents appropriately into 2 and 11 clusters) using the factoextra package (Kassambara, Mundt, 2020) obtained with the script:

```
library(conjoint)
library(factoextra)
data(journey)
segments<-caSegmentation(jpref,jprof,2)
print(segments$segm)
fviz_cluster(segments$segm,segments$util,
geom=c("text"),ellipse.type="convex",ellipse.alpha=0.0,
main="Cluster plot - 2 segments")
segments<-caSegmentation(jpref,jprof,11)
print(segments$segm)
fviz_cluster(segments$segm,segments$util,
geom=c("text"),ellipse.type="convex",ellipse.alpha=0.0,
main="Cluster plot - 11 segments")
```

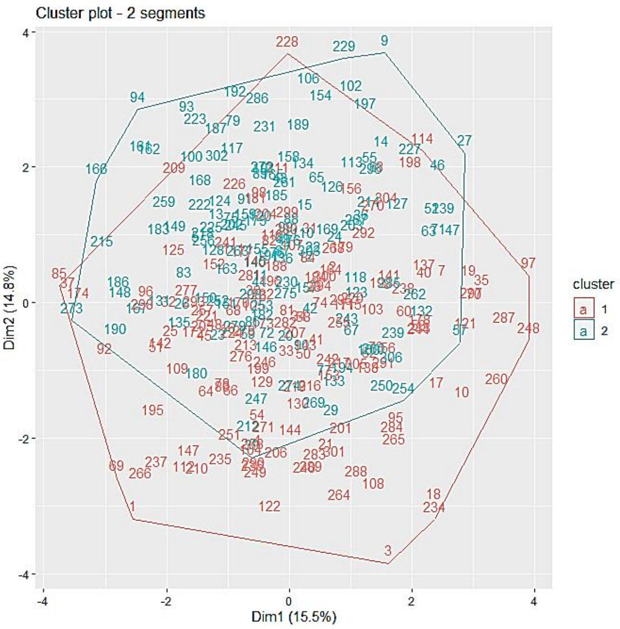

Figure 3. The visualisation of respondents' segmentation into 2 clusters.

Source: authors' compilation using conjoint and factoextra R packages.

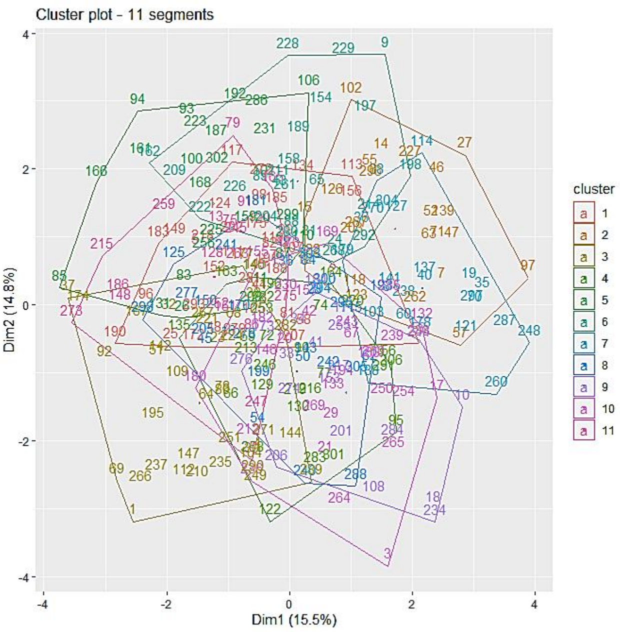

Figure 4. The visualisation of respondents' segmentation into 11 clusters.

Source: authors' compilation using conjoint and factoextra R packages.

In case of the division 306 respondents into 2 segments, the following segment sizes were obtained: 1 – 163, 2 – 143. The respondents' inclusion in the following segments is as follows:

```
K-means clustering with 2 clusters of sizes 163, 143
Cluster means:
[,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
1 5.501558 4.550515 3.716233 4.385067 4.571755 6.404399 4.618963 4.723951
2 4.037385 3.575552 6.494084 6.655217 5.967993 3.539399 5.996126 5.243678
[,9] [,10] [,11] [,12] [,13] [,14]
1 5.392896 5.070528 4.119374 5.364166 3.323951 5.710632
2 5.404783 5.238839 4.777035 5.140238 5.710329 5.135427
```

Clustering vector:
[1] 1 1 1 1 2 2 1 1 2 1 2 2 2 2 2 1 1 1 1 2 1 2 2 2 1 2 2 1 2 1 1 1 1 1 1 2
[37] 1 1 2 1 1 2 2 2 1 2 2 1 2 1 1 2 1 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 1 1 2 2
[73] 1 1 2 2 2 1 2 1 1 1 2 1 1 2 2 2 2 2 2 1 2 2 1 1 1 1 2 2 2 2 1 1 1 2 1 1
[109] 1 2 1 1 2 1 1 1 2 2 1 2 1 1 2 2 1 2 2 2 1 1 2 2 2 2 2 2 2 1 1 2 2 1 1 1 1
[145] 1 2 1 2 2 2 2 1 1 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 2 1 1 1 1 1 2 2 2 1 1 1 2
[181] 1 2 2 1 2 2 2 1 2 2 2 2 1 2 1 1 2 1 1 2 1 2 2 1 1 1 1 2 1 1 1 2 1 2 2 1
[217] 1 2 1 1 1 2 2 1 2 1 2 1 2 2 2 1 1 1 1 1 1 1 2 1 1 1 2 1 2 1 2 1 1 2 1 2
[253] 2 2 1 2 1 1 2 1 2 2 2 1 1 1 2 1 2 1 1 2 2 2 2 1 1 1 2 1 1 1 1 1 2 2 1 1
[289] 1 1 1 1 1 1 1 2 1 1 1 1 1 2 2 1 1 2

Within cluster sum of squares by cluster:
[1] 13619.73 11304.09
(between_SS / total_SS = 9.6 %)

In case of the division 306 respondents into 11 segments, the following segment sizes were obtained: 1 – 40, 2 – 25, 3 – 36, 4 – 29, 5 – 26, 6 – 25, 7 – 24, 8 – 30, 9 – 20, 10 – 29, 11 – 22. The respondents' inclusion in the following segments is as follows:

K-means clustering with 11 clusters of sizes 40, 25, 36, 29, 26, 25, 24, 30, 20, 29, 22

Cluster means:
[,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
1 4.147575 3.276725 5.921175 6.304525 4.526950 5.887350 7.660775 5.210075
2 5.911520 3.236920 7.848480 4.643080 5.290720 3.786160 5.723080 5.751520
3 4.190861 6.034694 2.003583 6.493083 5.569500 6.434083 4.246556 3.308917
4 4.827966 2.637276 3.758241 5.742034 5.713517 7.402103 6.332552 2.043483
5 3.548500 5.258846 6.701500 5.606538 5.065115 4.256231 7.409423 4.538885
6 3.021480 3.911920 6.008520 2.898080 3.770840 4.821120 7.808080 7.881480
7 7.715917 3.453458 6.554917 3.775708 1.589667 4.848167 3.687167 6.580500
8 5.225300 4.470300 2.749700 2.588033 5.823133 5.409633 2.933867 6.350300
9 6.698050 6.228800 3.701950 3.821200 5.161500 6.817400 3.821200 4.048050
10 5.763241 2.611448 5.305724 9.043724 6.644552 2.872069 2.414414 5.090828
11 2.667864 4.340364 5.820773 7.852818 8.379409 2.006455 5.159636 4.417864
[,9] [,10] [,11] [,12] [,13] [,14]
1 5.593600 4.021175 3.150150 8.082975 2.798050 6.893900
2 2.546160 7.968480 5.293840 2.116920 6.509280 4.333840
3 7.798667 3.302194 5.145778 5.996500 3.569389 5.989528
4 4.027103 6.611690 4.421172 2.762276 4.286483 7.330655
5 3.443731 3.441885 5.152423 7.850192 8.242577 6.753385
6 4.771120 5.128520 6.018880 5.701920 3.029160 2.948880
7 3.801292 6.982000 2.719542 5.885750 2.327000 6.287250
8 6.188800 5.008033 4.252867 7.066133 6.110200 5.723700
9 4.167400 4.851950 4.382600 1.278800 1.738500 2.082600
10 8.828966 6.874690 3.722759 4.085586 3.872690 5.869310
11 6.449636 3.320773 4.993545 4.238091 7.029682 3.141273

Clustering vector:
[1] 3 1 10 3 11 2 2 7 6 9 5 2 11 2 2 8 10 9 7 10 10 2 3 6
[25] 1 5 2 8 10 3 6 3 9 9 7 7 3 7 11 7 9 10 6 1 8 2 2 1
[49] 8 8 3 2 9 8 2 4 2 1 4 7 1 8 2 3 6 4 10 3 3 7 2 4
[73] 4 4 11 11 4 3 11 3 1 1 5 8 5 11 5 6 6 4 11 3 5 5 4 1
[97] 2 1 1 5 5 2 7 3 1 5 1 9 3 5 1 3 1 7 8 1 1 2 1 1
[121] 7 4 2 1 8 2 7 11 4 4 5 10 10 1 4 8 7 7 2 4 7 3 8 3
[145] 1 10 3 11 1 8 11 1 9 6 10 1 10 6 5 10 5 6 4 4 11 5 3 5
[169] 10 8 9 1 9 3 1 11 3 7 7 10 8 10 1 9 1 11 5 1 6 1 1 5
[193] 8 10 3 4 6 6 8 10 9 1 2 6 8 9 1 5 6 3 6 10 4 6 11 4
[217] 9 1 4 4 3 6 5 3 5 6 2 6 6 10 5 4 1 9 3 10 3 7 10 8
[241] 8 8 10 9 11 4 11 7 3 10 3 11 4 10 9 5 3 4 11 7 6 2 11 10

```
[265] 10 3 2 6 10 7 3 1 11 9 11 9 8 8 1 6 1 3 4 9 10 5 7 8
[289] 3 3 4 6 1 8 8 2 7 8 5 8 4 5 8 7 8 4

Within cluster sum of squares by cluster:
[1] 1907.956 1257.565 1865.649 1432.238 1223.052 1265.228 1134.321 1540.495
[9] 1214.192 1713.851 1205.163
(between_SS / total_SS = 42.9 %)
```

All results of this research and others illustrating the application of conjoint package in the analysis of stated preferences (using ranking and rating measurement scale), including simulation analysis and consumer segmentation are available on the following websites: (Bąk, Bartłomowicz, 2018b) (in Polish) and (Bąk, Bartłomowicz, 2023a) (in English).

7. Conclusions

From all conjoint methods, currently the Choice-Based Conjoint is the most popular method of stated preference analysis (Garrow, 2010; Aizaki et al., 2015). Nevertheless, the Traditional Conjoint Analysis is still highly popular with many practical applications. According to the research, the Traditional Conjoint Analysis is the third most popular and, in practice, most commonly used conjoint method, just after Choice-Based Conjoint and adaptive methods (ACA/ACBC).

One implementation of the traditional conjoint analysis method for R environment is the conjoint R package. A feature which characterizes the conjoint package is the high statistics of conjoint package downloads by RStudio users (RStudio, 2023). Until May 2023 total number of conjoint package installations exceeded 475,000. The most similar to the conjoint package in terms of the implemented conjoint method – the radiant.multivariate package has been downloaded a little more than 90,000 times (May 2023). Both results confirm the growing interest of students and researchers in the field of microeconomics and marketing research in the practical application of the traditional conjoint method in analysis of stated preferences. We can say that although the traditional conjoint analysis method has been known and used in marketing research for over forty years, it is still one of the most commonly used methods of measurement consumers' stated preferences.

Figure 5 illustrates the number of daily downloads of conjoint and radiant.multivariate R packages prepared using cran_downloads function from cranlogs package (Csárdi, 2022) and ggplot function from ggplot2 package (Wickham, 2023) – script[10]:

[10] It should be noted that the download statistics do not additionally cover the users of other R package versions, including primarily the original version of R environment, as well as the users of Microsoft R Application Network.

```
library(cranlogs)
library(ggplot2)
c<-cran_downloads("conjoint",from="2012-08-01",to=Sys.Date())
r<-cran_downloads("radiant.multivariate",from="2012-08-01",to=Sys.Date())
df<-data.frame(x=c$date,val=c(c$count,r$count),
package=c(rep("conjoint",nrow(c)),rep("radiant.multivariate",nrow(r))))
g<-ggplot(df,aes(x,val,col=package))+geom_line(linewidth=.6)
g+xlab("")+ylab("Number of downloads")+
labs(title="conjoint and radiant.multivariate R packages daily downloads")+
theme(legend.position=c(.15,.90))
sum(c$count);sum(r$count)
```

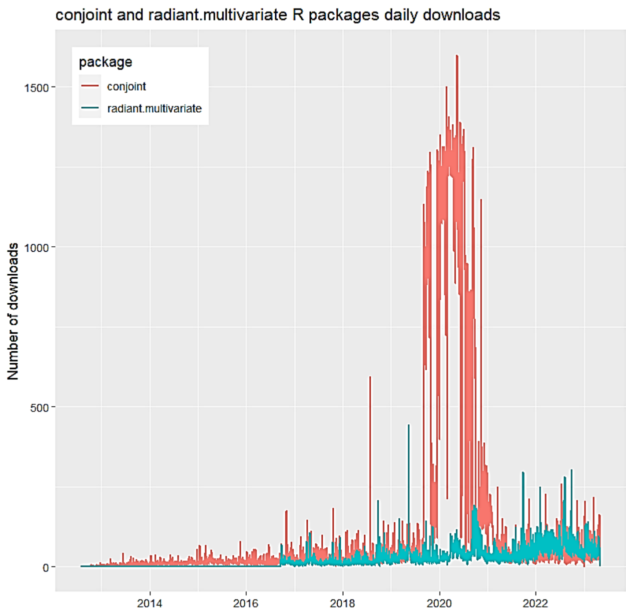

Figure 5. The number of conjoint and radiant.multivariate packages downloads by RStudio (RStudio, 2023) users.

Source: authors' compilation using cranlogs and ggplot2 R packages.

High popularity and proper functionality of the conjoint package is confirmed by the publications which recommended or at least described and cited the package: (Fiedler, Kaltenborn, Melles, 2017; Ben-Akiva, McFadden, Train, 2019; Aizaki et al., 2015; Mair, 2018; Koeser, Klein, Hasing, Northrop, 2015; Makkar, Williamson, Turner, Redman, Louviere, 2015; Le, Le, Nguyen, 2014). The users of social media channels also express their positive opinions about the conjoint package. Since 2018, websites (Bąk, Bartłomowicz, 2018b) and (Bąk, Bartłomowicz, 2023a) have also been available, which present detailed information about the conjoint package and examples of the use of the traditional conjoint analysis method and the conjoint package in empirical research on consumer preferences.

The article also presents the results of a survey of preferences of tourists choosing a place for a holiday trip. Using the data about the stated respondents' preferences and R program with conjoint package, the partial utilities of levels of attributes were estimated with OLS method. The obtained results made it possible to achieve the main aim of the study – to calculate the importance of the attributes included in the study, as well as to determine the most and least preferred profile of a tourist trip at the individual level as well as for all respondents. Additional objectives of the study were also achieved. With total and maximum utility and BTL models also implemented in the package, estimation of market share of the so-called simulation profiles not ranked by the respondents before was possible. At the end, using conjoint and some other packages the segmentation of tourists with similar choice preferences has been also made.

The obtained results of preference analysis indicate that at an aggregate level (for all respondents) the most important among all attributes used in the research trip purpose and accommodation type seem to be the most important among the attributes used in the example, followed by the season of the year, whereas the form (organized, own) of the trip seems to be the least important. At the attribute level the most popular are health trips focused on regenerating condition of tourists, organized on their own and in winter, with hostel as an accommodation. Organized business trips in the summer to a 1-2-3 star hotel turned out to be the least preferred by the respondents. At the same time, an analysis of the market share of simulation profiles revealed that respondents are able to accept the levels of some features in exchange for the other preferred attributes. Information on choice preferences made it possible to divide respondents into the segments. The results obtained in the study indicated the division of respondents into 2 or 11 segments.

References

1. Aizaki, H. (2022). *Basic Functions for Supporting an Implementation of Choice Experiments*. https://cran.r-project.org/package=support.CEs (R package support.CEs version 0.5.0).
2. Aizaki, H., Nakatani, T., Sato, K. (2015). *Stated preference methods using R*. CRC Press, Boca Raton: Chapman & Hall/CRC the R series.
3. Aizaki, H., Nakatani, T., Sato, K. (2022). *Analysing dichotomous choice contingent valuation data*. https://cran.r-project.org/package=DCchoice (R package DCchoice version 0.1.1).
4. Bąk, A. (2004). *Dekompozycyjne metody pomiaru preferencji w badaniach marketingowych*. Wrocław: Wydawnictwo Akademii Ekonomicznej we Wrocławiu.
5. Bąk, A. (2009). Analiza conjoint. In: M. Walesiak, E. Gatnar (Eds.), *Statystyczna analiza danych z wykorzystaniem programu R* (pp. 283-317). Warszawa: PWN.
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. 243-253). Warsaw: C.H. Beck.
8. Bąk, A., Bartłomowicz, T. (2018a). *Conjoint analysis*. https://cran.r-project.org/

package=conjoint. (R package conjoint version 1.41).
9. Bąk, A., Bartłomowicz, T. (2018b). *Conjoint R package*. https://pl.wikipedia.org/wiki/Conjoint_R.
10. Bąk, A., Bartłomowicz, T. (2023a). *Conjoint R package*. https://github.com/packagesR/conjoint/wiki.
11. Bąk, A., Bartłomowicz, T. (2023b). *Conjoint R package*. https://github.com/packagesR/conjoint.
12. Bartłomowicz, T., Bąk, A. (2021). *Badania preferencji i zachowań konsumentów z wykorzystaniem metod mikroekonometrii i programu R*. Wrocław: Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.
13. Ben-Akiva, M., Lerman, S.R. (1985). *Discrete choice analysis. theory and application to travel demand.* Cambridge/London: MIT Press.
14. Ben-Akiva, M., McFadden, D., Train, K. (2019). Foundations of stated preference elicitation: Consumer behavior and choice-based conjoint analysis. *Foundations and Trends in Econometrics*, *10(1-2),* 1-144.
15. Coombs, C.H., Dawes, R.M., Tversky, A. (1970). *Mathematical psychology: an elementary introduction*. Englewood Cliff s: Prentice-Hall.
16. Croissant, Y. (2022). *Multinomial logit models*. http://cran.r-project.org/package=mlogit (R package mlogit version 1.1-1).
17. Csárdi, G. (2022). *Download Logs from the RStudio CRAN Mirror*. https://cran.r-project.org/package= cranlogs (R package cranlogs version 2.1.1).
18. Debreu, G. (1960). Topological methods in cardinal utility theory. In: K. Arrow, S. Karlin, P. Suppes (Eds.), *Mathematical methods in the social sciences* (pp. 16-26). Stanford: Stanford University Press.
19. Fenwick, I. (1978). A user's guide to conjoint measurement in marketing. *European Journal of Marketing*, *12(2),* 203-211.
20. Fiedler, H., Kaltenborn, T., Melles, R.L.T. (2017). *Conjoint-analyse*. München: Rainer Hampp Verlag.
21. Garrow, L.A. (2010). *Discrete choice modelling and air travel demand: Theory and applications*. Aldershot: Ashgate Publishing.
22. Green, P.E., Rao, V.R. (1971). Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research*, *8(September),* 355-363.
23. Green, P.E., Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research*, 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),* 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), 107-117.

27. Großmann, H., Holling, H., Schwabe, R. (2002). Advances in optimum experimental design for conjoint analysis and discrete choice models. In: P.H. Franses, A.L. Montgomery (Eds.), *Econometric models in marketing, Vol. 16,* pp. 93-117. Oxford: Elsevier Science.
28. Gustafsson, A., Herrmann, A., Huber, F. (Eds.) (2007). *Conjoint measurement. Methods and applications*. Berlin/Heidelberg/New York: Springer-Verlag.
29. Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C. (1995). *Multivariate data analysis with readings.* Englewood Cliff s: Prentice-Hall.
30. Hensher, D.A., Rose, J.M., Greene, W.H. (2005). *Applied choice analysis. A primer*. Cambridge: Cambridge University Press.
31. Hooley, G.J., Lynch, J.E. (1981). Modelling the student university choice process through the use of conjoint measurement techniques. *European Research, 4(October),* 158-170.
32. Huber, J., Zwerina, K. (1996). The importance of utility balance in efficient choice designs. *Journal of Marketing Research, 33(August)*, 307-317.
33. Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software, 61(6),* 1-36. Retrieved from: https://www.jstatsoft.org/index.php/jss/article/ view/v061i06.
34. IBM, SPSS (2023). *IBM SPSS Conjoint*. https://www.ibm.com/us-en/marketplace/spss-conjoint. Chicago.
35. Kahneman, D., Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47(2),* 263-291.
36. Kahneman, D., Tversky, A. (Eds.) (2000). *Choices, values, and frames*. Cambridge University Press.
37. Kassambara, A., Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses* [Computer software manual]. Retrieved from: https://CRAN. R-project.org/package=factoextra (R package version 1.0.7).
38. Kendall, M.G., Buckland, W.R. (1986). *Słownik terminów statystycznych*. Warszawa: PWE.
39. Kim, J., Lee, C. (2023). The return of the king: The importance of killer content in a competitive ott market. *Journal of Theoretical and Applied Electronic Commerce Research, 18(2),* 976-994. Retrieved from: https://www.mdpi.com/0718-1876/18/2/50 doi: 10.3390/jtaer18020050.
40. Koeser, A.K., Klein, R.W., Hasing, G., Northrop, R.J. (2015). Factors driving professional and public urban tree risk perception. *Urban Forestry & Urban Greening, 14(4),* 968-974.
41. Kruskal, J.B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29(1),* 1-27.
42. Kruskal, J.B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika, 29(2),* 115-129.

43. Kruskal, J.B. (1965). Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society, 27(2),* 251-263.
44. Kuhfeld, W.F., Tobias, R.D., Garratt, M. (1994). Efficient experimental design with marketing research applications. *Journal of Marketing Research*, *31(November),* 545-557.
45. Lawson, S., Glowa, T. (2000). *Discrete choice experiments and traditional conjoint analysis*. Quirk’s Marketing Research Review.
46. Le, T. M., Le, S., Nguyen, D.H. (2014). Assessing consumer-perceived food quality using conjoint analysis. *Science and Technology Development Journal, 17(3),* 21-31.
47. Linzer, D., Lewis, J. (2022). *Polytomous variable Latent Class Analysis*. https://cran.r-project.org/package=poLCA (R package poLCA version 1.6.0.1).
48. Louviere, J.J. (1988). *Analyzing decision making. metric conjoint analysis*. Newbury Park/Beverly Hills/London/New Delhi: SAGE Publications.
49. Louviere, J.J. (1994). Conjoint analysis. In: R.P. Bagozzi (Ed.), *Advanced methods of marketing research* (pp. 223-259). Oxford: Blackwell.
50. Louviere, J.J. (2000). *Why stated preference discrete choice modelling is not conjoint analysis (and what SPDCM is),* http://www.memetrics.com/products/SPDCM_whitepaper.pdf (Memetrics White Paper).
51. Louviere, J.J., Flynn, T.N., Carson, R.T. (2010). Discrete choice experiments are not conjoint analysis. *Journal of Choice Modelling*, *3(3),* 57-72.
52. Louviere, J.J., Hensher, D.A., Swait, J.D. (2000). *Stated choice methods. Analysis and applications*. Cambridge: Cambridge University Press.
53. 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),* 69-83. Retrieved from: https://www.mdpi.com/0718-1876/15/3/22 doi:10.4067/S0718-18762020000300106
54. Luce, R.D., Tukey, J.W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, *1(1),* 1-27.
55. Mair, P. (2018). *Modern psychometrics with R*. Cham: Springer.
56. Makkar, S.R., Williamson, A., Turner, T., Redman, S., Louviere, J. (2015). Using conjoint analysis to develop a system to score research engagement actions by health decision makers. *Health Research Policy and Systems*, *13(22),* 1-16.
57. Nijs, V. (2023). *Multivariate menu for radiant: Business analytics using R and shiny*. https://cran.r-project.org/package=radiant.multivariate (R package radiant.multivariate version 1.5.0)
58. Ott, L. (1984). *An introduction to statistical methods and data analysis*. Boston: PWS Publishers.
59. Poortinga, W., Steg, L., Vlek, C., Wiersma, G. (2003). Household preferences for energy-saving measures: A conjoint analysis. *Journal of Economic Psychology*, *24(1),* 49-64.
60. 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).
61. Rao, V.R. (2014). *Applied conjoint analysis*. Berlin/Heidelberg: Springer-Verlag.
62. Rasch, D., Herrendörfer, G. (1991). *Statystyczne planowanie doświadczeń*. Warszawa: PWN.
63. Rossi, P. (2022). *Bayesian Inference for Marketing / Micro-Econometrics*. https://cran.r-project. org/package=bayesm (R package bayesm version 3.1-5).
64. RStudio (2023). *RStudio: Integrated Development for R*. http://www.rstudio.com/. Boston, MA.
65. SAS (2023). *Statistical Analysis System*. https://www.sas.com. Cary, NC.
66. Sawtooth Software, Inc. (2023). *Sawtooth Software*. https://www.sawtoothsoftware.com. Provo.
67. Shim, D., Lee, C., Oh, I. (2022). Analysis of ott users' watching behavior for identifying a profitable niche: Latent class regression approach. *Journal of Theoretical and Applied Electronic Commerce Research*, *17(4),* 1564-1580. Retrieved from: https://www.mdpi.com/0718-1876/17/4/79 doi: 10.3390/jtaer17040079.
68. SPSS (1994). *SPSS 6.1 Categories*. Chicago.
69. Thaler, R., Sunstein, C. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.
70. Therneau, T.M. (2023). *Survival analysis*. https://cran.r-project.org/package=survival (R package survival version 3.5-3).
71. TIBCO (2023). *Statistica*. http://statistica.io. Palo Alto, CA.
72. Vriens, M., Wittink, D.R. (1994). *Conjoint analysis in marketing* (Typewritten copy).
73. Walesiak, M. (1996). *Metody analizy danych marketingowych*. Warszawa: PWN.
74. Walesiak, M., Bąk, A. (2000). *Conjoint analysis w badaniach marketingowych*. Wrocław: Wydawnictwo Akademii Ekonomicznej we Wrocławiu.
75. Wedel, M., Kamakura, W.A. (2000). *Market segmentation. conceptual and methodological foundations*. Boston/Dordrecht/London: Kluwer Academic Publishers.
76. Wheeler, R.E. (2022). *Algorithmic experimental design*. https://cran.r-project.org/ package=AlgDesign (R package AlgDesign version 1.2.1).
77. Wickham, H. (2023). *Create elegant data visualisations using the grammar of graphics*. https://cran.r-project.org/package=ggplot2 (R package ggplot2 version 3.4.1).
78. Wilkinson, L. (1998). *Conjoint analysis*. Chicago: SYSTAT.
79. Wind, Y., Green, P. (2004). *Marketing research and modeling: Progress and prospects. a tribute to Paul E. Green*. New York: Springer-Science+Business Media.
80. Zwerina, K. (1997). *Discrete choice experiments in marketing*. Heidelberg/New York: Physica-Verlag.
81. Zwerina, K., Huber, J., Kuhfeld, W.F. (2000). *A General Method for Constructing Efficient Choice Design*. https://faculty.fuqua.duke.edu/~jch8/bio/Papers. Duke University, The Fuqua School of Business.