INFORMATYKA EKONOMICZNA BUSINESS INFORMATICS 2(52) • 2019
ISSN 1507-3858
e-ISSN 2450-0003

Tomasz Bartłomowicz
Wroclaw University of Economics
e-mail: tomasz.bartlomowicz@ue.wroc.pl
ORCID: 0000-0001-9986-0841

Andrzej Bąk
Wroclaw University of Economics
e-mail: andrzej.bak@ue.wroc.pl
ORCID: 0000-0002-5980-9694

THE `CONJOINT` R PACKAGE AS A TOOL FOR MEASURING STATED PREFERENCES

PAKIET `CONJOINT` PROGRAMU R JAKO NARZĘDZIE POMIARU PREFERENCJI WYRAŻONYCH

DOI: 10.15611/ie.2019.2.01
JEL Classification: D12, C83, C88

Summary: Two groups of research methods are used in the decompositional approach to stated preferences – conjoint analysis methods and discrete choice methods. The most commonly applied traditional conjoint analysis method is an example of the first group. Because of its computational complexity, its practical application requires using appropriate commercial or non-commercial computer software. The purpose of the article is to present the traditional conjoint analysis method and discuss its implementation in the form of the `conjoint` package for R program, which with CRAN packages is currently one of the most important non-commercial computing environments for statistical data analysis. In addition to the detailed characteristics of the individual `conjoint` R package functions, the paper also presents the application of the `conjoint` package in marketing research, along with the interpretation of the selected results, based on the example of measuring and analysing stated preferences of beer consumers.

Keywords: stated preferences, conjoint analysis, R program.

Streszczenie: W podejściu dekompozycyjnym wykorzystuje się dwie grupy metod badawczych – metody *conjoint analysis* oraz metody wyborów dyskretnych. Przykładem pierwszej grupy jest stosowana z powodzeniem do dnia dzisiejszego tradycyjna metoda *conjoint analysis*. Ze względu na jej złożoność obliczeniową jej praktyczne zastosowanie oznacza wykorzystanie odpowiedniego komercyjnego lub niekomercyjnego oprogramowania komputerowego. W artykule omówiono tradycyjną metodę *conjoint analysis* oraz zaprezentowano implementację tej metody w postaci modułu `conjoint` programu R, który wraz z innymi pakietami oraz programem R jest obecnie jednym z najważniejszych, niekomercyjnych środowisk obliczeniowych przeznaczonych do analizy statystyczno-ekonometrycznej. Oprócz szczegółowej charakterystyki poszczególnych funkcji pakietu `conjoint`, w artykule zaprezentowane zostało także zastosowanie pakietu w badaniach marketingowych wraz z interpretacją wybranych wyników na przykładzie pomiaru i analizy preferencji wyrażonych konsumentów piwa.

Słowa kluczowe: pomiar preferencji wyrażonych, pakiet `conjoint` programu R.

1. Introduction

One of the essential marketing research components is the measurement and analysis of consumer preferences. If this measurement is based on the data collected through surveys, allowing the registration of consumers' intentions at the time of conducting the survey, the stated preferences corresponding to the hypothetical (declared) preferences of consumers become the subject of measurement. The most commonly used methods, applied in measuring and analysing stated preferences in the so-called decompositional approach, take the form of conjoint analysis methods, including the traditional conjoint analysis method as well as the discrete choice methods (Walesiak and Bąk, 2000; Bąk, 2004).

The article discusses the traditional conjoint analysis method and presents its implementation in the form of the `conjoint` R package. The R program is currently one of the most important non-commercial and dynamically developing projects for statistical data analysis. In addition to the detailed characteristics of the individual `conjoint` package functions, the paper also presents the application of this package in marketing research based on the example of measuring and analysing stated preferences of beer consumers.

2. Measuring preferences using the conjoint analysis method

The conjoint analysis method is based on the axiomatic measurement theory, originally proposed based on psychometric research. This theory, referred to in the subject literature as conjoint measurement, defines the conditions for the existence of variable measurement scales (response and predictor variables), in which the response variable values are generated jointly by the predictor variables, in accordance with the specified rule of the measurement model composition (an additive rule in the traditional conjoint measurement model). This model is used to analyse the combined effect of many predictor variables on the values adopted by the response variable. The order of the response variable values is analysed in different combinations of the predictor variable values. A simultaneous and additive influence of the predictor variables on the response variable is assumed. Due to the measurement of the response variable value, including the simultaneous impact of all predictor variables (their main effects), this measurement model is referred to as the additive conjoint measurement (Coombs, Dawes, and Tversky 1977, p. 50; Green and

Srinivasan, 1978, p. 103; Wilkinson, 1998, p. 87; Smith, 1989, p. 83). Therefore, conjoint measurement is a measurement theory assuming the existence of a measurement scale of the response variable and the measurement scales of such predictor variables which allow quantifying the joint impact of predictor variables on the response variable, in accordance with the specific model composition rules (Green and Srinivasan, 1978).

The conjoint analysis methods are often used in empirical research to analyse stated preferences measured on metric scales. In such cases, the linear multiple regression model with dummy variables is usually applied. Its parameters are estimated using the Ordinary Least Squares (OLS) method.

The marketing data about the respondents' stated preferences, predominantly obtained 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 respondents rank the product or service profiles[1], stating their preferences in this way[2]. Profile rankings 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 the attribute levels[3] and in estimating the particular attributes' shares in the total utility development of each profile (see Green and Wind, 1975).

The study of consumer stated preferences using conjoint analysis is carried out in accordance with the procedure presented in Table 1.

One of the more important stages of the conjoint analysis procedure is the estimation of the conjoint analysis model parameters, aimed at estimating the so-called part-worths utilities of the attribute levels. Part-worths utilities are estimated for each respondent separately and as the average values for the analysed sample. 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 theoretical total utilities of the profiles in the analysed sample,
- the theoretical total utilities of the profiles in the identified groups (segments) of respondents,

[1] The number of all profiles possible to generate depends on the number of attributes and the number of levels (it is the product of the number of all attributes' levels). In practice, only a subset of profiles meeting the relevant conditions (e.g. of the system orthogonality) is ranked in the form of the so-called fractional factorial design.

[2] The respondent ranks the presented profiles according to his/her own stated preferences on the basis of the presented attributes and their levels, in terms of: determining the profiles' ranking (on the ranking scale) or rating the relative attractiveness of profiles (e.g. on the rating scale).

[3] In accordance with the terminology used in the subject literature referring to the conjoint analysis method, the predictor variables describing goods or services are called attributes or factors, whereas their realizations are referred to as levels. The 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 (i.e. the product of the level numbers of all attributes).

- the relative “importance” ranking of individual attributes in the cross-section of respondents in the analysed sample,
- the simulation market shares of the selected profiles,
- the segmentation of respondents.

In the estimated model the parameters (part-worths utilities of the attribute levels) are estimated using the ordinary least squares method. In the multiple regression analysis, the response variable adopts values (e.g. points or ranks) assigned by a given respondent to the individual profiles presented for the ranking. The impact of each level of the individual predictor variables (nonmetric attributes) on the ranking assigned to profiles by a given respondent is taken into account by introducing dummy predictor variables into the regression model.

Table 1. Conjoint analysis research procedure

No.	Procedure stage	Procedure step
1	Research task specification	• response variable • predictor variables (attributes)
2	Model form identification	• model of predictor variables correlation (main effects or with interactions) • preference model (linear, square, fractional utilities)
3	Data collection	• data collection methods (full profiles, comparing profiles in pairs, presentation of attribute pairs, simulation data) • profile generation methods (factorial designs, random sample)
4	Profile presentation	• presentation form (verbal description, drawing, model, physical product) • research form (direct interview, traditional mail, phone, computer, the Internet)
5	Selection of preference measurement scale	• nonmetric scale – ranking • metric scale – rating
6	Model estimation	• nonmetric models (MONANOVA) • metric models (MNK)
7	Analysis and interpretation of the results	• preference analysis (the ranking of attributes’ importance) • market share simulation • segmentation

Source: authors’ compilation based on: (Green and Srinivasan, 1978; Bąk, 2004; 2013).

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

$$Y = \beta_0 + \sum_{k=1}^{p} \beta_k Z_k + \varepsilon, \tag{1}$$

where: Y – response variable, taking the values of 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 or orthogonal coding are used (Zwerina, 1997; Walesiak and Bąk, 2000; Bąk, 2004). After transcoding the attributes, the conjoint analysis model with dummy variables can be presented in the following form:

$$\hat{Y}_s = b_0 + \sum_{j=1}^{m} b_j X_j, \tag{2}$$

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

As a result of model estimation (2) the values of $b_0, \ldots, b_m$ parameters are obtained and interpreted as part-worths utilities of the 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.

Part-worths utilities are calculated at an aggregated level (one model is estimated for the entire 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 constituting the subject of research. The total utility of i-th profile for s-th respondent $\left(U_i^s\right)$ is calculated based on the following formula (Walesiak, 1996, p. 93):

$$U_j^s = b_{0s} + \sum_{j=1}^{m} U_{jl_j^i}^s, \tag{3}$$

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

The average theoretical total utility (at an aggregated level, i.e., for the entire sample covering S respondents) of i-th profile (U_i) is calculated based on the formula (Walesiak, 1996, p. 95):

$$U_i = \frac{1}{S} \sum_{s=1}^{S} \left(b_{0s} + \sum_{j=1}^{m} U_{jl_j^i}^s \right). \tag{4}$$

The knowledge of part-worths utilities also allows estimating the significance (the so-called attribute "importance") for every attribute in the ranking of profiles being the subject of research. The relative importance of j-th attribute for s-th respondent W_j^s is calculated based on the following formula (Hair, Anderson, Tatham, and Black, 1995, p. 608):

$$W_j^s = \frac{\max_{l_j}\left\{U_{jl_j^i}^s\right\} - \min_{l_j}\left\{U_{jl_j^i}^s\right\}}{\sum_{j=1}^{m}\left(\max_{l_j}\left\{U_{jl_j^i}^s\right\} - \min_{l_j}\left\{U_{jl_j^i}^s\right\}\right)} \times 100\%. \tag{5}$$

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

$$W_j = \frac{1}{s}\sum_{s=1}^{S} W_j^s. \tag{6}$$

The simulation analysis of market shares allows estimating 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 following models (Hair et al., 1995, p. 591; Walesiak, 1996, p. 97):

- 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:

$$P_{is} = \begin{cases} 1, & when\, U_i^s = \max\left(U_i^s\right) \\ 0, & otherwise \end{cases}, \tag{7}$$

where: P_{is} – 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 the profiles covered by the simulation (the calculations are carried out separately for each respondent and next their average value is computed):

$$P_{is} = \frac{U_i^s}{\sum_{i=1}^{n} U_i^s}, \tag{8}$$

where: n – the number of profiles;

- logit model in which, as opposed to the probabilistic BLT model, the natural logarithms of total utilities' values are used in the calculations instead of the utilities themselves:

$$P_{is} = \frac{\exp\left(U_i^s\right)}{\exp\left(\sum_{i=1}^{n} U_i^s\right)}. \tag{9}$$

The parameter values of the estimated conjoint analysis model (estimated part-worths and total utilities) can, additionally, 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 applied, which uses the data classification methods (cluster analysis) in the division of the respondent's set into classes (segments), based on individual part-worths utilities. Due to the certain specific features (unequivocal qualification of objects into groups, effective processing of large data sets) the *k*-means method is frequently used, which belongs to the group of iterative optimization methods.

More information on the procedure and the conjoint analysis method can be found in both Polish (Walesiak and Bąk, 2000; Bąk, 2004) and English-language subject literature (Hair et al., 1995; Coombs et al., 1977; Green and Srinivasan, 1978; Green and Wind, 1975; Wilkinson, 1998; Smith, 1989; Zwerina, 1997).

3. The characteristics of the `conjoint` R package

The `conjoint` statistical software package is the authors' (Bąk and Bartłomowicz 2018) extension (module) of the R program including an implementation of the traditional conjoint analysis method. The correct functioning of the package requires installing the R program and some additional packages which, starting from the 3.3.2 version of the R program are downloaded and installed along with the `conjoint` R package. The package can be downloaded from the website of the CRAN R repository (https://cran.r-project.org/package=conjoint) and the home page of the package – the website of the Department of Econometrics and Computer Science of Wroclaw University of Economics and Business (http://keii.ue.wroc.pl/conjoint). The `conjoint` package is written in R programming language and available under the GNU license (free of charge and provides access to the source code). The `conjoint` R package is also compatible with the software dedicated to R environment such as: RStudio and Microsoft R Application Network.

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

In addition, the package offers tools supporting the design of a questionnaire survey, i.e. developing the appropriate factorial designs allowing, in particular, the reduction of the complete set of profiles to the form of fractional designs (orthogonal and effective). For this purpose, the `AlgDesign` R package (Wheeler, 2015) functions

Table 2. Functions of the `conjoint` R package

Functions of the conjoint package	
caFactorialDesign(data, type=”null”, cards=NA, seed=123) – the function defines (full or fractional) factorial design using the same names of variables and levels	
caEncodedDesign(design) – the function encodes the factorial design obtained using caFactorialDesign function for the needs of the 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 ranking scale into a data set in the form of points (on a rating scale)	
caPartUtilities(*y*, *x*, *z*) – the function calculates the part-worths utility matrix of attribute levels in the cross-section of the respondents (including the intercept)	
caTotalUtilities(*y*, *x*) – the function calculates the theoretical total utilities matrix of profiles in the cross-section of the respondents	
caUtilities(*y*, *x*, *z*) – the function calculates part-worths utilities of attribute levels at an aggregated level	
caImportance(*y*, *x*) – the function calculates an average relative “importance” of all attributes (as %) at an aggregated level	
caBTL(sym, *y*, *x*) – the function estimates market shares of the simulation profiles based on the BLT (Bradley-Terry-Luce) probability model	
caLogit(sym, *y*, *x*) – the function estimates market shares of simulation profiles based on the logit model	
caMaxUtility(sym, *y*, *x*) – the function estimates market shares of simulation profiles based on the maximum utility model	
caModel(*y*, *x*) – the function estimates conjoint analysis model parameters for an individual respondent	
caSegmentation(*y*, *x*, *c*=3) – the function carries out the respondents’ segmentation using *k*-means method based on *k*-means() function	
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 (part-worths and total) utilities available in the conjoint package	
ShowAllSimulations(sym, *y*, *x*) – the function estimates market shares of the simulation profiles based on all simulation models available in the conjoint package	
Function arguments	
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 of the random number generator (default seed=123)
design	factorial experiment design (fractional or full)
attr.names	vector representing names of attributes (factors)
lev.numbers	vector representing numbers of attribute levels (factors)
prof.numbers	vector representing numbers of reconstructed profiles
y.rank	matrix (vector) of empirical preferences in the ranking form (the ranking data require transformation to rating data using caRankToScore function)
y	matrix (vector) of empirical preferences (in the form of importance assessment on the rating or ranking scale)
x	matrix representing profiles (including the names of attributes)
y.type	type of data about preferences – data in the form of profile importance assessments on the rating or ranking scale (rating is implicitly adopted)
sym	matrix representing simulation profiles (including the names of attributes)
c	optional parameter describing the number of segments (default c=3 – division into 3 segments)
z	vector representing names of attribute levels (factors)

Source: the authors’ compilation.

are used in the conjoint package. Its implementation in the conjoint package is carried out in the form of functions which allow generating orthogonal or effective fractional factorial designs and their conversion (the functions: caFactorialDesign, caEncodedDesign and caRecreatedDesign), including their coding using dummy variables. In order to generate the relevant fractional factorial design (full and fractional), the data on the number of attributes are taken into account as well as their levels and also the names of attributes and their levels. This offers the possibility of designing an experiment using the conjoint analysis method and subsequently obtaining the results of preference measurement directly in the conjoint R package.

The detailed characteristics of all available functions, data sets and the selected examples of the package application in measuring stated preferences are available in the documentation of the conjoint R package.

4. The application of the `conjoint` package in measuring stated preferences

In the conjoint R package the marketing data about the respondents' stated preferences can take the form of the ranking scale data (with the specified ranking order) and the rating scale data (with the determined relative attractiveness of the profiles).

The example illustrating the package application in empirical research presents the measurement of the stated preferences of beer consumers based on the data collected in the form of a rating[4]. In the set of variables describing the surveyed product, the following variables (features, attributes) along with the corresponding levels were listed: price (below 3 PLN, 3-5 PLN, above 5 PLN), serving form (can, bottle, mug), alcohol level (0%, 0.5-3%, 4-6%, above 6%) and taste additions (yes, no):

```
> library(conjoint)
> beer<-expand.grid(
+ price=c("below 3 PLN", "3-5 PLN", "above 5 PLN"),
+ form=c("can", "bottle", "mug"),
+ alcohol=c("0%", "0.5-3%", "4-6%", "above 6%"),
+ taste=c("yes", "no"))
> print(beer)
```

Due to the large number of profiles resulting from the combination of levels of all features (in this case the so-called full factorial design consists of 72 profiles[5]), in the presented example the following effective fractional factorial design was used:

[4] Rating means the valuation of profiles within the framework of the adopted scale (in the presented example the scale range is [0-10] but other scales, e.g. the Likert scale is possible).

[5] The number of profiles is the product of the number of levels of all attributes ($4^1 3^2 2^1$=72).

```
> profnm<-caFactorialDesign(data=beer, type="fractional")
> print(profnm)
          price   form  alcohol  taste
1    above 5 PLN bottle      0%    yes
2    below 3 PLN    mug      0%    yes
3        3-5 PLN    can  0.5-3%    yes
4    above 5 PLN    mug  0.5-3%    yes
5    below 3 PLN    can    4-6%    yes
6    above 5 PLN bottle above 6%    yes
7        3-5 PLN    mug above 6%    yes
8    above 5 PLN    can      0%     no
9        3-5 PLN    mug      0%     no
10   below 3 PLN bottle  0.5-3%     no
11       3-5 PLN bottle    4-6%     no
12   above 5 PLN    mug    4-6%     no
13   above 5 PLN    can above 6%     no
14   below 3 PLN    mug above 6%     no
> profnr=caEncodedDesign(profnm)
> print(profnr)
```

In the research 160 correctly completed questionnaires were used to analyse beer consumer preferences[6]. The examples of ratings of the first six respondents are presented below:

```
> prefer=read.csv2("beer_preferences.csv", header=TRUE)
> print(head(prefer))
  pr01 pr02 pr03 pr04 pr05 pr06 pr07 pr08 pr09 pr10 pr11 pr12 pr13 pr14
1    0    0    7    0    8    0    7    0    0    0   10    8    0    6
2    0    0    0    0    0    0    0    0    0    0   10    5    1    9
3    0    0    0    0    0    0    0    0    5    0   10   10    0    0
4    4    3    3    3    5    9   10    0    0    4    5    7    4    9
5    1    1    3    3    3    3    3    5    5    6   10   10    8    8
6    0    0    5    5    7    7    7    0    0    3    8    9    3    7
> levnms=read.csv2("beer_levels.csv", header=TRUE)
```

The availability of data on empirical preferences (prefer), the coded research design (profnr), the names of variables and their levels (levnms) allow summarizing (in the cross-section of respondents) the most important results of measuring preferences using the Conjoint() function:

```
> Conjoint(prefer, profnr, levnms)
Call:
lm(formula = frml)
Residuals:
    Min      1Q  Median      3Q     Max
-6,5858 -2,1166 -0,1166  1,9389  8,1467
```

[6] The data were collected using a questionnaire survey posted on the website: www.ankietka.pl. Daria Wolsztajn is the author of the survey which covered 160 persons (80 women and 80 men) in 2015-2016. Age of the respondents: under 20 – 3 persons, 20-30 – 114, 31-40 – 27, 41-50 – 5, 51-60 – 7 and over 60 – 4 persons.

```
Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          4,13268    0,06435  64,225  < 2e-16 ***
factor(x$price)1    -0,32891    0,09201  -3,575 0,000358 ***
factor(x$price)2     0,37422    0,09201   4,067 4,92e-05 ***
factor(x$form)1     -0,55260    0,09201  -6,006 2,21e-09 ***
factor(x$form)2      0,26146    0,09201   2,842 0,004527 **
factor(x$alcohol)1 -1,87227    0,10376 -18,045  < 2e-16 ***
factor(x$alcohol)2 -0,08992    0,11582  -0,776 0,437617
factor(x$alcohol)3  1,62664    0,11582  14,044  < 2e-16 ***
factor(x$taste)1    -0,19078    0,06305  -3,026 0,002506 **
---
Signif. codes:  0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Residual standard error: 2,912 on 2231 degrees of freedom
Multiple R-squared:  0,1791,    Adjusted R-squared:  0,1762
F-statistic: 60,85 on 8 and 2231 DF,  p-value: < 2,2e-16

[1] "Part worths (utilities) of levels (model parameters for whole sample):"
        levnms    utls
1    intercept  4,1327
2  below 3 PLN -0,3289
3      3-5 PLN  0,3742
4  above 5 PLN -0,0453
5          can -0,5526
6       bottle  0,2615
7          mug  0,2911
8           0% -1,8723
9       0.5-3% -0,0899
10        4-6%  1,6266
11    above 6%  0,3355
12         yes -0,1908
13          no  0,1908
[1] "Average importance of factors (attributes):"
[1] 17,82 18,59 46,36 17,23
[1] Sum of average importance:  100
[1] "Chart of average factors importance"
```

The obtained results of the attributes' importance indicate that among the variables included in the example, the most important at an aggregate level (for the population of respondents) is the level of alcohol (46.36%), next is the serving form (18.59%), whereas the price (17.82%) and taste additions (17.23%) are the least important. The analysis of part-worths utilities of the attribute levels indicates that canned beer with the level of alcohol ranging 4%-6%, without added tasteand at an average price (PLN 3-5) is the most preferred one. The obtained results are illustrated on the graphs of part-worths utilities of the attributes (Figures 1 to 4) and on the graph of the attributes' importance (Figure 5).

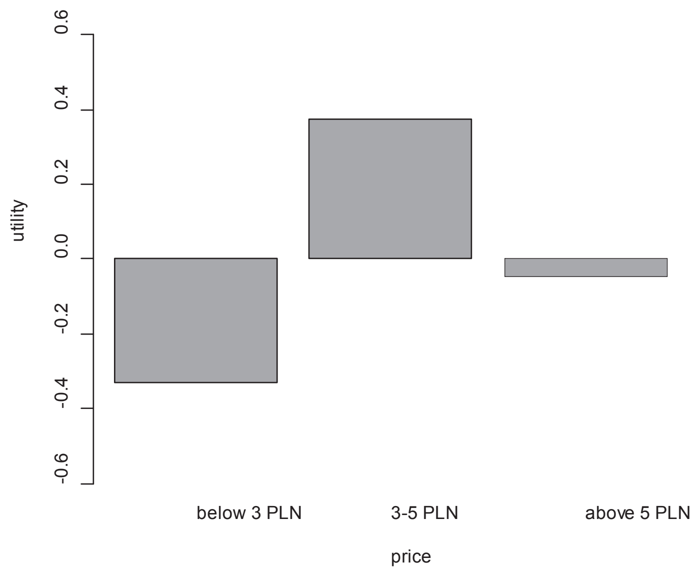

Fig. 1. The graph of part-worths utilities of the price attribute

Source: the authors' compilation using the `conjoint` R package.

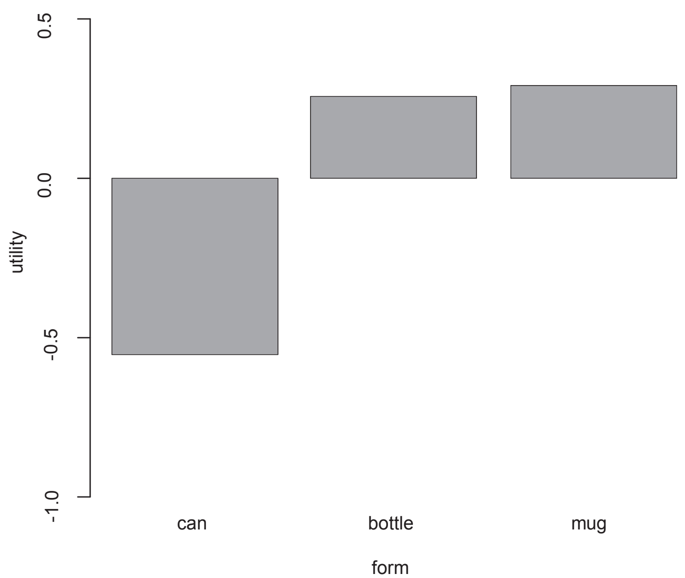

Fig. 2. The graph of part-worths utilities of serving form attribute

Source: the authors' compilation using the `conjoint` R package.

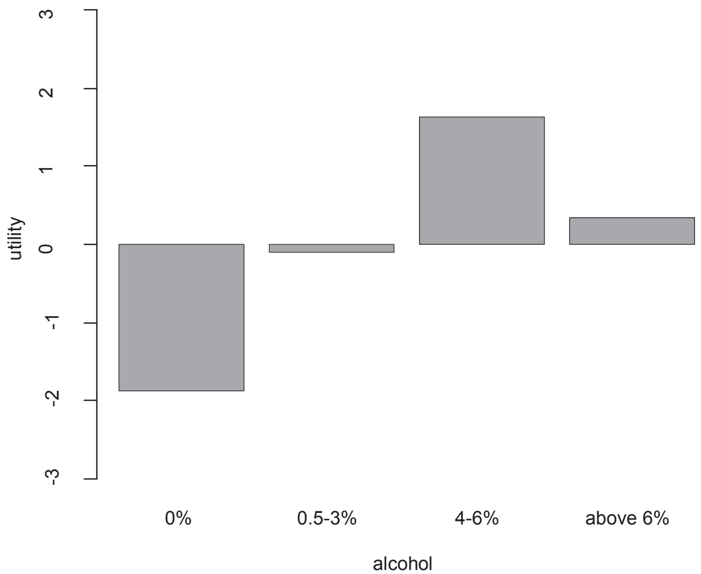

Fig. 3. The graph of part-worths utilities of the alcohol level attribute

Source: the authors' compilation using the `conjoint` R package.

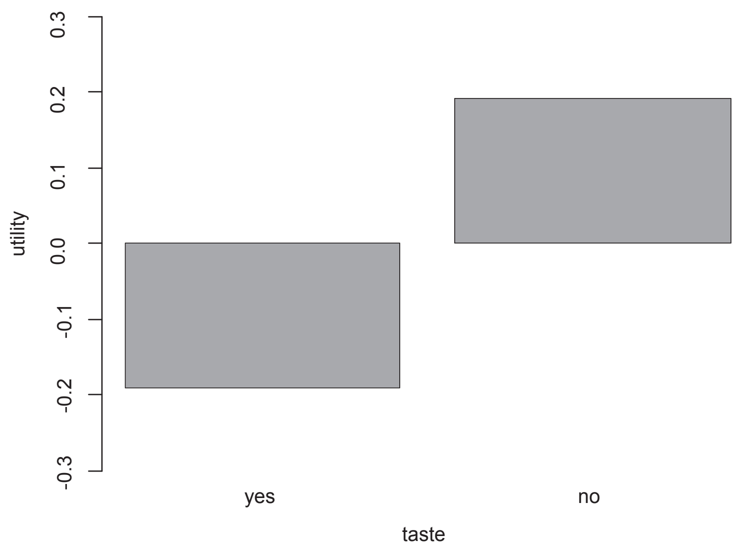

Fig. 4. The graph of part-worths utilities of the taste additions attribute

Source: the authors' compilation using the `conjoint` R package.

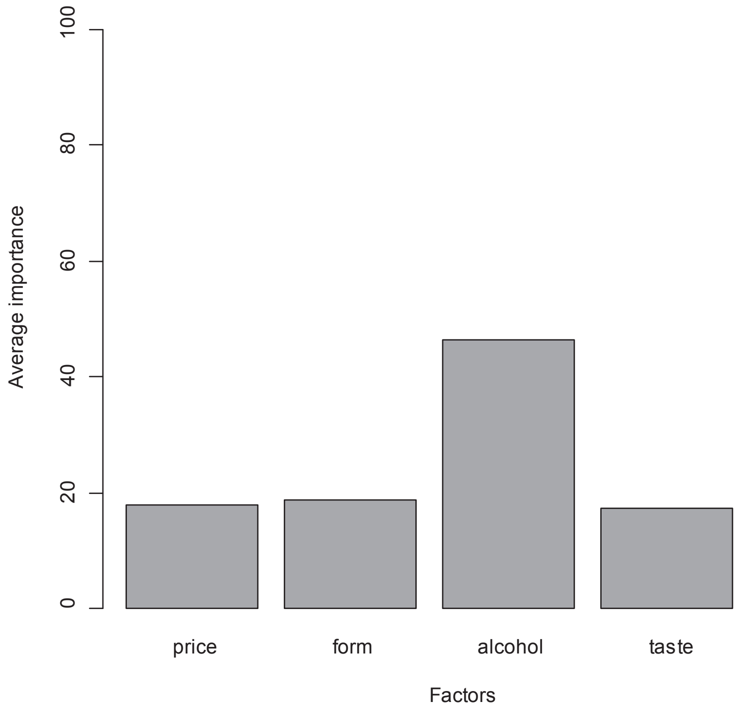

Fig. 5. The graph of the attributes' importance

Source: the authors' compilation using the `conjoint` R package.

The conjoint package offers the possibility of estimating measurement results at an individual level. The following commands, based on the selected `conjoint` R package functions and the selected respondents:

```
> caModel(prefer[1, ], profnr)
> caUtilities(prefer[2, ], profnr, nazpoz)
> caPartUtilities(prefer[3, ], profnr, nazpoz)
> caTotalUtilities(prefer[4, ], profnr)
> caImportance(prefer[5, ], profnr)
> ShowAllUtilities(prefer[6, ], profnr, nazpoz)
```

allow estimating, respectively: model parameters for respondent No. 1 (the intercept and the model parameters without reference levels), model parameters for respondent No. 2 (the intercept and the model parameters with reference levels, without names of the levels), model parameters for respondent No. 3 (model parameters with reference levels and names of the levels), the theoretical values of total utilities for respondent No. 4, the attributes' importance for respondent No. 5 and the aggregate results of part-worths utilities for respondent No. 6. Other functions of the `conjoint`

R package also offer the possibility of referring to the selected respondents, including the Conjoint() function.

The conjoint package allows estimating market shares of the so-called simulation profiles, i.e., the profiles not ranked by the respondents before. Based on the analysis of the obtained importance of the features indicated by the consumers, four beer variants were selected for the simulation analysis. 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 (a glass bottle, alcohol level ranging 4%-6% and no taste additions) along with a price of more than PLN 5. Profiles No. 1 and 2 offer two desired features (price of PLN 3-5 and no taste additions), but differ in terms of the remaining features – a bottle and no alcohol in profile No. 1 is replaced by a mug and alcohol level ranging 4%-6% in profile No. 2. Profile No. 4 offers only a mug from the preferred feature levels:

```
> profsm=read.csv2("beer_simulations.csv", header=TRUE)
> print(profsm)
  price form alcohol taste
1     2    2       1     2
2     2    1       4     2
3     3    2       3     2
4     1    3       2     1
```

The total utility (attractiveness) of the simulation variants for all respondents was calculated using maximum utility models – caMaxutility() function, the probabilistic BTL (Bradley-Terry-Luce) model – caBTL() function and the logit model – caLogit() function. The ShowAllSimulations() function allows obtaining the aggregate simulation results :

```
> ShowAllSimulations(profsm, prefer, profnr)
  TotalUtility MaxUtility BTLmodel LogitModel
1         3,09       4,38    18,03       7,45
2         4,48      18,12    25,67      21,22
3         6,17      55,62    34,39      51,15
4         3,81      21,88    21,92      20,18
```

Out of the selected beer variants covered by the simulation analysis, the largest market share is expected for profile No. 3 (based on all – maximum utility, BTL, logit and total utility models). The lowest market share (also according to the all models) is expected for profile No. 1. The comparison of relevant profiles confirms the respondents' preferences regarding the desired features and indicates that the respondents are able to accept the levels of some features, e.g. higher price than the preferred one – more than PLN 5 (profile No. 3) and are not able to accept certain features, e.g. zero alcohol level (profile No. 1).

In order to carry out the respondents' segmentation the caSegmentation() function from the `conjoint` R package can be applied, which uses the *k*-means method. In the case of the default parameters the caSegmentation() function is used for the sample division into three segments (in the presented example into 46, 39 and 75 respondents – see Figure 6) presenting the following composition (numbers 1, 2, 3 stand for the respondent's inclusion in a given segment):

```
> segments<-caSegmentation(prefer, profnr, c=3)
> print(segments$seg)
K-means clustering with 3 clusters of sizes 46, 39, 75

Cluster means:
       [,1]       [,2]        [,3]      [,4]      [,5]      [,6]      [,7]
[,8]      [,9]     [,10]     [,11]     [,12]     [,13]     [,14]
1 5.4581522  5.4853261  6.94130435 6.854348 5.746739 5.208152 5.740761
3.8027174 4.449457 5.008696 4.825000 4.623913 3.552717 3.694022
2 0.1519231 -0.3865385 -0.08461538 1.300000 2.119231 1.600641 2.017308
0.5532051 2.553205 2.528205 6.642308 6.443590 2.001923 3.046795
3 1.4496667  1.1713333  3.62366667 3.862000 5.372333 5.559667 5.974667
1.3336667 2.592000 4.501000 7.636333 7.031333 5.443667 6.008667

Clustering vector:
  [1] 3 2 2 3 3 3 3 3 3 1 1 3 1 2 3 3 3 3 3 3 1 3 3 3 1 3 3 3 3 1 3 3 1 1 3
3 1 3 2 3 2 2 2 3 1 1 1 3 2 1 2 3 3 2 1 2 2 1 3 2 1 1 2 2 2 2 1 3 2 3 2 3
 [73] 3 3 3 1 1 1 1 1 1 3 1 2 3 3 3 2 3 1 3 3 3 1 3 3 1 2 1 3 3 3 3 1 2 3 3
3 3 1 1 1 1 1 3 1 1 2 3 3 2 3 3 3 1 2 3 3 3 1 1 2 2 1 1 1 2 2 3 2 3 2 3 3
[145] 2 2 3 3 1 3 2 2 3 2 2 3 1 2 3 1

Within cluster sum of squares by cluster:
[1] 3637.629 2011.596 4102.799
 (between_SS / total_SS =  38.0 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"
"tot.withinss" "betweenss"    "size"         "iter"         "ifault"

> summary(segments)
     Length Class  Mode
segm    9   kmeans list
util 2240   -none- numeric
sclu  160   -none- numeric
> require(fpc)
> plotcluster(segments$util, segments$sclu)
```

Full versions of other examples illustrating the application of the `conjoint` package in the analysis of stated preferences (on both measurement scales – ranking

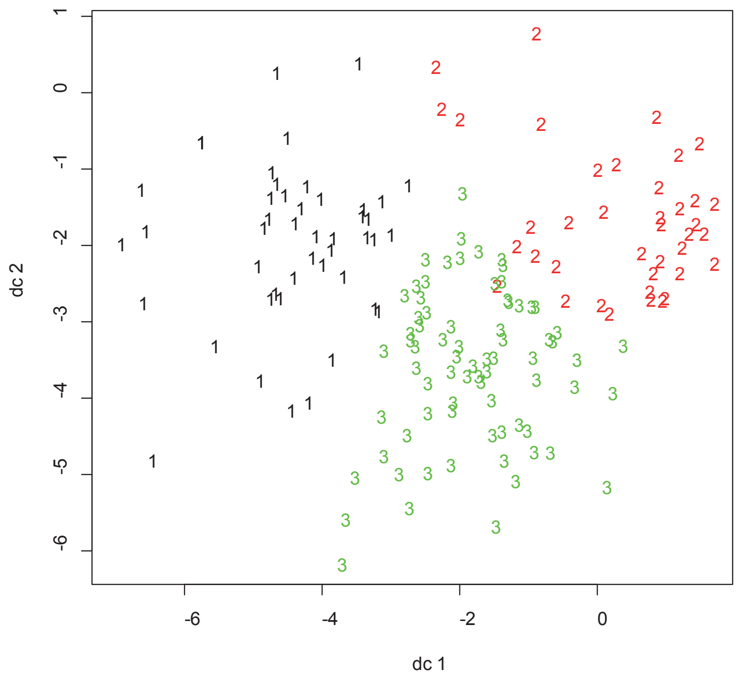

Fig. 6. The division of respondents into three segments

Source: the authors' compilation using the conjoint R package.

and rating), including the simulation analysis and consumer segmentation are available, e.g. on the following websites: https://pl.wikipedia.org/wiki/Conjoint_R (Polish) (Bartłomowicz and Bąk, 2018a) and http://keii.ue.wroc.pl/conjoint/Conjoint_R.html (English) (Bartłomowicz and Bąk, 2018b).

5. Conclusion

The conjoint R package represents a non-commercial alternative for the commercial IBM SPSS Conjoint module (SPSS, 1994; IBM SPSS, 2018), and also other commercial packages supporting the conjoint analysis, e.g. Sawtooth software (Sawtooth and Software, 2020), Statistica (Statistica, 2020), SAS (SAS, 2020) as well as websites which allow conducting the stated preference research online, e.g. Conjoint.ly (Conjoint.ly, 2020), Survey Analytics (Survey Analytics, 2020) and others.

The package functions support all stages of the research procedure carried out using the traditional conjoint analysis method, including the development of a questionnaire survey using factorial designs (orthogonal and effective) and the conjoint model estimation (linear regression model with dummy variables).

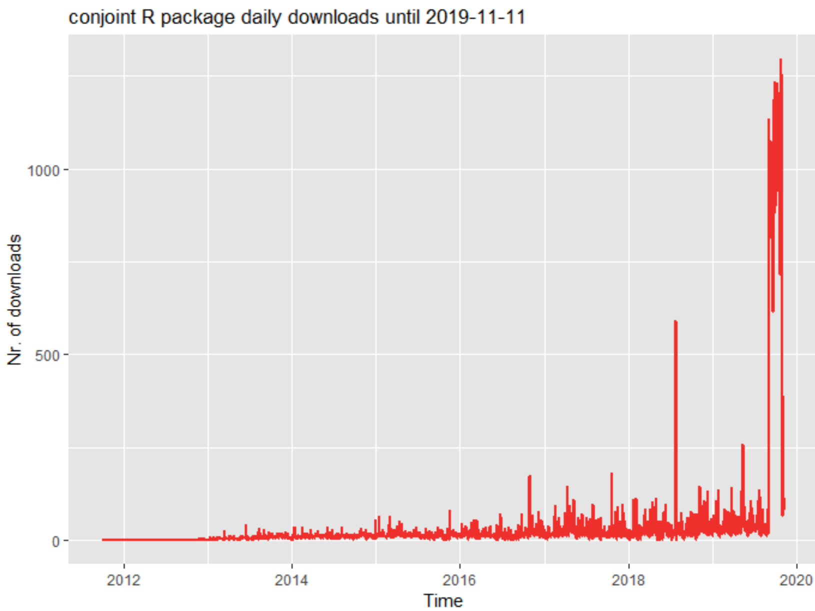

Fig. 7. The number of `conjoint` R package installations per day

Source: the authors' compilation using the `cranlogs` and `ggplot2` packages.

The general access to the package (including the access to the source code and documentation), its standard installation as well as the application of the `conjoint` R package functions represent the typical features of the `conjoint` package and the other CRAN R packages. The `conjoint` package has been available in the CRAN R repository (R Core Team, 2020) since 2011. From then the number of package installations by its users exceeded 100 000 downloads and is still showing an increasing tendency. It also confirms the growing interest of students and researchers in the field of microeconomics and marketing research regarding the practical application of conjoint analysis in the research covering consumer stated preferences. This is illustrated in Figure 7 developed using the cran_downloads() function from the `cranlogs` R package and the ggplot() function from the `ggplot2` R package – script:

```
library(cranlogs)
library(ggplot2)
x<-cran_downloads("conjoint", from="2011-10-01",to=Sys.Date()-2)
g<-ggplot(x, aes(x$date, x$count))+
geom_line(colour="red", size=1)
g+xlab("Time")+ylab("Nr. of downloads")+
labs(title=paste0("conjoint R package daily downloads until ",Sys.
Date()-2))
sum(x$count)
```

References

Bartłomowicz, T., and Bąk, A. (2018a). *Conjoint R – Wikipedia, wolna encyklopedia.* Retrieved from: https://pl.wikipedia.org/wiki/Conjoint_R

Bartłomowicz, T., and Bąk, A. (2018b). *Conjoint R.* Retrieved from: http://keii.ue.wroc.pl/conjoint/Conjoint_R.html

Bąk, A. (2004). Dekompozycyjne metody pomiaru preferencji w badaniach marketingowych. *Research Papers of the University of Economics in Wrocław*, (1013).

Bąk, A. (2013). *Mikroekonometryczne metody badania preferencji konsumentów z wykorzystaniem programu R.* Warsaw: C.H. Beck Publishers.

Bąk, A., and Bartłomowicz, T. (2012). Conjoint analysis method and its implementation in conjoint R package. In J. Pociecha, R. Decker, *Data Analysis Methods and Its Applications* (pp. 239-248). C.H. Beck Publishers, Warsaw.

Bąk, A., and Bartłomowicz, T. (2018). *Conjoint analysis – conjoint package*. Retrieved from https://cran.r-project.org/package=conjoint

Conjoint.ly. (2020). *Conjoint.ly. Tools and support for product and pricing research. Conjoint.ly.* Retrieved from: https://conjoint.online

Coombs, C. H., Dawes, R. M., Tversky, A. (1977). *Wprowadzenie do psychologii matematycznej*. Warsaw: Polish Scientific Publishers.

Fenwick, I. (1978). A user's guide to conjoint measurement in marketing. *European Journal of Marketing*, *12*(2), 203-211.

Green P. E., and Wind Y. (1975). New way to measure consumers' judgments. *Harvard Business Review*, 53 (July-August), 107-117.

Green, P. E., and Srinivasan, V. (1978). Conjoint analysis in consumer research: issues and outlook. *Journal of Consumer Research*, (5), 103-123.

Green, P. E., and Srinivasan, V. (1990). *Conjoint analysis in marketing: new developments with implications for research and practice*. Journal of Marketing, (54), 3-19.

Green, P. E., Krieger, A. M., and Wind, Y. (2001). *Thirty years of conjoint analysis: reflections and prospects*. Retrieved from http://www-marketing.wharton.upenn.edu/ideas/pdf

Green, P. E., Wind, Y. (1973). *Multiattribute decisions in marketing. A measurement approach*. Hinsdale, Illinois Dryden Press.

Großmann, H., Holling, H., and Schwabe, R. (2002), *Advances in optimum experimental design for conjoint analysis and discrete choice models*. In Franses P.H., Montgomery A.L. (ed.), *Econometric Models in Marketing*, *vol. 16* (pp. 93-117). Oxford: Elsevier Science.

Gustafsson A., Herrmann A., and Huber F. (ed.). (2007), *Conjoint measurement. Methods and applications*. 4th Edition, Springer-Verlag, Berlin-Heidelberg-New York.

Hair J. F., Anderson R. E., Tatham R. L., and Black W. C. (1995). *Multivariate data analysis with readings*, Prentice-Hall, Englewood Cliffs.

Hooley, G. J., and Lynch, J. E. (1981). *Modelling the student university choice process through the use of conjoint measurement techniques*. European Research, (4), 158-170.

Huber, J., and Zwerina, K. (1996). *The importance of utility balance in efficient choice designs*. Journal of Marketing Research, (33), 307-317.

Kendall, M. G., and Buckland, W. R. (1986). *Słownik terminów statystycznych*. Warsaw: Polish Economic Publishers,.

Kruskal, J. B. (1964), *Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis*. Psychometrika, *2*(91), 1-27.

Kruskal, J. B. (1964a). *Nonmetric multidimensional scaling: a numerical method*. Psychometrika, *29*(2), 115-129.

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.
Kuhfeld, W. F. (1997). *Efficient experimental design using computerized searches*. Retrieved from: https://www.sawtoothsoftware.com/download/techpap/1997Proceedings.pdf
Ott, L. (1984). *An introduction to statistical methods and data analysis* (2nd Edition). Boston: PWS Publishing Co.
R Core Team. (2020). *R: a language and environment for statistical computing. R foundation for statistical computing*. Vienna. Retrieved from https://www.R-project.org/
Rasch, D., and Herrendörfer, G. (1991). *Statystyczne planowanie doświadczeń.* Warsaw: Polish Scientific Publishers.
RStudio Team. (2018). *RStudio: integrated development for R*. Boston: MA: RStudio. Inc. Retrieved from: http://www.rstudio.com/
SAS. (2020). *Statistical analysis system*. NC: SAS Institute, Cary. Retrieved from: https://www.sas.com
Sawtooth Software. (2020), *Sawtooth Software, Inc.,* Provo. Retrieved from: https://www.sawtoothsoftware.com
Smith, S. L. J. (1989). *Tourism analysis: a handbook*. Harlow: Longman Scientific & Technical.
Statistica. (2020). *Statistica*. Palo Alto, CA: TIBCO Software Inc. Retrieved from: http://statistica.io
Survey Analytics. (2020). *Conjoint analysis free survey. Survey analytics*. Retrieved from: https://www.surveyanalytics.com/conjoint-survey-free.html
Vriens, M. (2001). *Market segmentation. Analytical developments and application guidelines*. Millward Brown IntelliQuest.
Vriens, M., and Wittink, D. R. (1994). *Conjoint analysis in marketing* (maszynopis powielony *[typewritten copy])*.
Walesiak, M. (1996). *Metody analizy danych marketingowych*. Warsaw: Polish Scientific Publishers.
Walesiak, M., and Bąk, A. (2000). *Conjoint analysis w badaniach marketingowych.* Press Wrocław: University of Economics in Wrocław.
Wheeler, R. E. (2015). *Package AlgDesign*. *Algorithmic Experimental Design*. Retrieved from: https://cran.r-project.org/package=AlgDesign
Wilkinson, L. (1998). *Conjoint analysis*. In *SYSTAT 8.0* (pp. 87-114). Chicago: SPSS Inc.
Zwerina, K. (1997). *Discrete choice experiments in marketing*. Heidelberg-New York: Physica-Verlag.
Zwerina, K., Huber, J., and Kuhfeld, W. F. (2000). *A General Method for Constructing Efficient Choice Design*. Retrieved from: https://faculty.fuqua.duke.edu/~jch8/bio/Papers/