SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2025
ORGANIZATION AND MANAGEMENT SERIES NO. 218

MEASUREMENT OF MUSIC STREAMERS' PREFERENCES USING BEST-WORST SCALING AND CONJOINT ANALYSIS METHODS

Tomasz BARTŁOMOWICZ

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

Purpose: The main aim of the research was to measure the music streamers' preferences using Best-Worst Scaling and conjoint analysis methods. The additional purpose is to compare the results obtained from both used methods, which should get similar conclusions. Finally, the cooperation of the `support.BWS3` and `conjoint` R packages as the one common tool for measurement of stated preferences was also examined.
Design/methodology/approach: Multi-profile Best-Worst Scaling method uses a modeling approach based on the conditional logit model, whereas traditional conjoint analysis method applies a linear regression model. Therefore, comparing the results of both methods was even more interesting.
Findings: In the paper, the results of measurement and analysis of music streamers' preferences were presented, calculations from different preference models were confronted and the correct use of R packages in the form of completed scripts was demonstrated.
Research limitations/implications: The limitations of one used method were compensated by the second one. The cooperation of both methods and used R packages was not only confirmed but also led to complete the research results.
Practical implications: The research results, as well as the combined use of some R packages may interest practitioners, researchers and students in the fields of marketing research, in the area of measurement of consumers' preferences. Streaming companies, manufacturers of playback equipment, artist and record labels as well as marketers should be interested in the research results.
Social implications: One of the used packages – authoring `conjoint` R package implements the traditional conjoint analysis method similarly to the module Conjoint IBM SPSS program. The statistics (over a half million of downloads) indicate that the non-commercial `conjoint` package is popular among R users.
Originality/value: In addition to the benefits in form of conclusions drawn from the research, the paper presents one more example confirming results of both used methods. There are no similar studies confirming the results of multi-profile Best-Worst Scaling (Case 3 BWS) and traditional conjoint analysis methods based on measurement of music streamers' preferences.
Keywords: stated consumers' preferences, Best-Worst Scaling, conjoint analysis, R program.
Category of the paper: research paper.
JEL Classification: C6, C8, D1.

1. Introduction

Streaming music began to gain importance in the early 2000s, but its real expansion started from 2010 and continued in the next years, when platforms such as Spotify, Apple Music, YouTube Music, Tidal and Deezer began to gain huge popularity. Today, streaming services have become the dominant way people listen to music, while traditional sources such as radio, records, and even MP3 downloads have become a thing of the past.

According to many studies, we live in the golden age of streaming. The global streaming industry market was worth about $21 billion in 2019, and according to report of the Recording Industry Association of America, as much as 80% of music industry revenues came from streaming in 2022. Forecasts suggest that the market will continue to grow at an average annual rate of 17,8% through 2027, almost eliminating music piracy.

The rise of streaming music has brought significant changes to the entire music industry. Previously, the main sources of income were record sales and concerts, whereas today revenue comes primally from music streaming services. The change in the music industry's business model means changes in the model of listening and creating music. Users can listen to song or playlist at any time, and the fee is not paid for songs on a physical medium but for a subscription, i.e. access to libraries. Creators earn mainly through streaming plays, which directly influences how music is produced and promoted.

Due to the features of streaming, it should be noted that music streaming is not only a way to listen to music, but it is also a tool for measurement all kinds of data including user activities and preferences. All users choices are monitored by algorithms, which then can recommend songs based on the diagnosed preferences.

The paper presents the results of measurement and analysis of stated preferences of music users, who are listening to music streaming from popular internet services and platforms. All calculations and visualizations of the obtained results were carried out using the R program and appropriate packages, in particular, the `support.BWS3` (Aizaki, 2024) package for the multi-profile Best-Worst Scaling and the `conjoint` (Bąk, Bartłomowicz, 2018a) package for the traditional conjoint analysis.

In the research, multi-profile Best-Worst Scaling (Case 3 BWS) method was applied first. Using a modeling approach based on discrete choice model (conditional logit model), the best and the worst attributes as well as attribute levels were identified. Then, conjoint analysis method was used. Using traditional conjoint model (linear regression model) next results of measurement of music streamers' preferences including attributes' importance were obtained. Additionally, having the results from both methods, a comparison of multi-profile Best-Worst Scaling and traditional conjoint analysis methods was made. Finally, the cooperation of the `support.BWS3` and `conjoint` R packages as the one common tool for measurement and analysis of stated preferences was also examined.

2. Literature review

The applications of Best-Worst Scaling as well as conjoint analysis methods, include the measurement of stated preferences in many areas. In food marketing research, examples of applications contain: chocolate (Thomson et al., 2010), wine (Cohen, 2009), coffee (Cohen, Neira, 2004), breakfast bars (Hein et al., 2008), restaurants (Chrzan, Golovashkina, 2006). In medical and healthcare the applications are as follows: healthcare system reform (Louviere, Flynn, 2010), residency programs (Wang et al., 2011), treatment decisions in rheumatoid artists, side effects of smoking (Marti, 2012). In values, research examples include: food safety issues (Finn, Louviere, 1992), food values (Luss, Briggeman, 2009), brand equity (Menictas et al., 2012), ethical beliefs (Auger et al., 2007) and energy-saving (Poortinga, 2003). There are also many examples in transportation, environmental, public policy and other research.

In the field of measurement and analysis of stated preferences of music users, who are listening to music streaming from popular internet services and platforms some research also were made.

Jones (2020) investigated users' loyalty to streaming platforms, focusing on the asset specificity of features and estimating users' willingness to pay (WTP) for each feature. A structural equation model based on survey data revealed that feature satisfaction positively influences both asset specificity and overall satisfaction with streaming platforms, thereby strengthening user loyalty. Using the conjoint analysis method Jones estimated that users are willing to pay at least $14,40 per month for platforms that offer recommendations, playlist and social features, and the ability to download music. Kim, Nam, & Ryu (2017) estimated and compared U.S. and Korean consumers' marginal willingness to pay (MWTP) for streaming services using the conjoint analysis method. The study examined attributes such as advertisements, streaming mode, exclusive content and offline usage. The results indicated that U.S. and Korean consumers have different preferences and MWTP for these product attributes. Based on the findings, the research suggests implications for both streaming services and the broader streaming industry.

Shin & Kim (2025) examined adolescent users, a highly influential demographic that rapidly adapts to new technology trends, to analyze the competitive dynamics among major music streaming services. The study selected 4 platforms (YouTube Music, Melon, Flo, Genie Music) and surveyed adolescent users aged 14 to 18. Grounded in niche theory, the research identified 5 gratification factors: price value, music diversity, ease of use, optional services and recommendation services. The competitiveness of each service was analyzed based on these factors. Through this analysis, the study offers strategic implications for Korean music streaming services to achieve sustainable success amid global competition.

Maftei, Gerogiannis & Papageorgiou (2016) identified the critical success factors of online music streaming services and examined the relationships between them. In the research they found that the core of online music streaming include free music streaming, the ability to purchase music in both digital and physical formats, the absence of advertisements and the satisfaction of supporting one's favorite artists.

Lopes & Coelho (2021) conducted a study that included interviews to gain a deeper understanding of the profile, behaviors, and motivations of the new music consumer. Their findings confirmed that habit, performance expectancy and price value play the most significant roles in influencing the intention to use a paid music streaming service. At the same time, new dimensions such as personalization, attitude toward piracy and perceived freemium-premium fit emerged as additional factors influencing the adoption of this type of service. The research provides valuable insights into consumer behavior in music streaming services, offering several theoretical and practical implications for music streaming service providers.

Allan & Leijonhufvud (2022) conducted a test to examine preferences of music streaming users, perceived sound quality and how the musical content affected them. Among the results, education and experience were found to influence preferences in some cases.

More research results covering streamers' music preferences using Best-Worst Scaling or conjoint analysis methods can be found in subject literature (Bamert et al., 2005; Breidert, Hahsler, 2007; Shin, Kim, 2014; Baek, 2023).

In the field of comparing the results of Best-Worst Scaling and conjoint analysis methods, research has been conducted such as the study by Cheng, Zhang, Lambert & Feuz (2023). This study compared consumer willingness to pay (WTP) derived from conjoint analysis and Best-Worst Scaling (Case 3 BWS) survey formats. Data on consumer preferences for single-use eating-ware products made from biobased materials were collected. The results suggest that for the most preferred attribute levels, WTP estimates are similar in magnitude and consistent in sign across both methods. However, for the least preferred attributes, WTP estimates from the conjoint analysis method are higher than those obtained from the multi-profile Best-Worst Scaling.

Hollin, Peay & Bridges (2015) compared Best-Worst Scaling (Case 2 BWS) and conjoint analysis methods, within a study measuring patients' muscular dystrophy treatment. Both methods were applied to 18 potential treatments, incorporating 6 attributes at 3 levels. The results showed that profile Best-Worst Scaling and conjoint analysis methods produced similar parameter estimates, conditional attribute importance and policy simulations. The highest concordance was observed for benefit and risk parameters, while differences emerged for nausea and knowledge about the drug, where a lack of monotonicity was noted in the conjoint analysis. Given the simplicity of combining Case 2 BWS and conjoint analysis for single profiles, the researchers suggested that a combined approach could be easily adopted. Potoglou, Burge, Flynn, Netten, Malley & Forder (2011) also compared Best-Worst Scaling (Case 2 BWS) and conjoint analysis methods using an in-person survey. Respondents were asked to answer both conjoint and Case 2 BWS questions. They found no significant difference between the methods regarding stated preferences.

More research results comparing the results of conjoint analysis and Best-Worst Scaling (or discrete choice) methods can be found in subject literature (van Dijk et al., 2016; Xie et al., 2013; Severin et al., 2013; Cheng et al, 2021; Himmler et al., 2021).

According to the presented sources, there are no similar studies in the literature measuring and confirming the results of multi-profile Best-Worst Scaling and traditional conjoint analysis methods based on measurement of music streamers' preferences.

3. Research methodology

Both methods used in the research are widely known, have a long history and hold a well-established position among methods of measurement and analysis of stated preferences. Historically, the older method is the traditional conjoint analysis, moreover the Best-Worst Scaling can be seen as a tool to eliminate some shortcomings of the conjoint analysis. Due to these all, although the multi-profile Best-Worst Scaling is used in the research as the basic method, in the next section the traditional conjoint analysis is discussed first.

3.1. Conjoint analysis

The first publication presenting conjoint measurement in psychometrics appeared in 1964 (Luce, Tukey, 1964), followed by additional works in the 1970s (Green, Rao, 1971; Green, Wind, 1973; Green, Srinivasan, 1978). Since then, numerous studies have discussed the methodological challenges and applications of conjoint analysis in marketing research. Nowadays, conjoint analysis is a group methods widely used for studying consumers' stated preferences for products and services, as well as political opinions and religious attitudes. A comprehensive review of the existing achievements and future development perspectives in conjoint analysis is provided in (Green et al., 2004; Gustafsson et al., 2007; Rao, 2014).

In the paper, the traditional conjoint analysis (TCA) was used. Two most important steps of TCA procedure are the measurement of consumers' preferences and the parameters estimation of conjoint analysis model.

The research material used in the traditional conjoint analysis method consists of marketing data on preferences declared by respondents obtained through survey research. Respondents evaluate product or service profiles[1] (real or hypothetical) described by a set of features

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

(attributes), thus expressing their (empirical) preferences. Based on the collected data, the total preferences are decomposed using statistical methods by calculating the share of each attribute in the estimated total utility value of the profile. Part-worth utilities are estimated for each respondent individually and as average value for the studied sample (Green, Wind, 1975; Walesiak, Bąk, 2000; Bąk, 2004).

In the traditional conjoint analysis the linear regression model is used, the parameters of which (part-worth utilities of the attribute levels) are estimated using the Ordinary Least Squares (OLS) method. The model for the selected respondent can be presented in the following form (Hair et al., 1995):

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

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

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

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

The total utility of i profile for s respondent (U_i^s) is calculated based on the following formula (Hair et al., 1995; Walesiak, 1996):

$$U_i^s = b_{0s} + \sum_{j=1}^{m} U_{l_j^i}^s \tag{2}$$

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

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

$$U_i = \frac{1}{S}\sum_{s=1}^{S}\left(b_{0s} + \sum_{j=1}^{m} U_{l_j^i}^s\right) \tag{3}$$

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

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

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

$$W_j^s = \frac{max\left\{U_{l_j^i}^s\right\} - min\left\{U_{l_j^i}^s\right\}}{\sum_{j=1}^m \left(max\left\{U_{l_j^i}^s\right\} - min\left\{U_{l_j^i}^s\right\}\right)} \times 100\% \tag{4}$$

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

$$W_j = \frac{1}{S}\sum_{s=1}^{S} W_j^s \tag{5}$$

where W_j^s – defined by a formula (4).

The traditional conjoint analysis method is a useful tool for studying consumer preferences, but it has certain limitations. First, it assumes that consumers make decisions in a fully rational and hierarchical manner, which does not always reflect real market behavior. Second, the number of attributes and their levels must be carefully selected because too many can lead to cognitive overload for respondents. Additionally, interpreting the results requires advanced statistical methods, which can be a challenge for companies lacking analytical resources. Finally, studies using this method can be costly and time consuming, especially when they require large respondent samples and complex experimental designs.

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

3.2. Best-Worst Scaling

The first publication discussing Best-Worst Scaling (BWS) presenting its theory, method and application was introduced in the 1990s (Louviere, 1988). More detailed descriptions of the method were provided in subsequent publications in 1991 and next years (Louviere, 1991), (Finn, Louviere, 1992). Since then, numerous studies were published on Best-Worst Scaling, showcasing various models for analyzing data based on stated preferences.

There are Case 1, Case 2 and Case 3 of Best-Worst Scaling method (Louviere et al., 2015), which differ in terms of the complexity of the items (options) under consideration. However, in each version of the method respondents are asked to choose the best (most important, most attractive) and the worst (least important, least attractive) items (objects, levels or profiles) comparing a finite and different set of alternatives.

In the object Best-Worst Scaling (Case 1 BWS) method, researchers examine the relative values assigned to each item in a given list of objects, considering only the items themselves without decomposing them into factors. A necessary number of distinct subsets of items is generated using an experimental design. Each subset is then presented to respondents as a choice set, and the process is repeated until all subsets have been evaluated (Finn, Louviere, 1992).

In the profile Best-Worst Scaling (Case 2 BWS) method, attributes levels replace objects. Respondents make their best-worst choices by comparing different levels of different attributes, what can be more difficult than comparing objects. It means that the alternatives in a Case 2 BWS question correspond to attribute levels presented within a choice set (Himmler et al., 2021; Flynn, 2010),

In the multi-profile Best-Worst Scaling (Case 3 BWS) method, each question corresponds to full profiles presented within a choice set. Respondents evaluate a sequence of choice sets based on the entire scenario rather than individual factor levels as in Case 2 BWS, or isolated factors as in Case 1 BWS (Louviere et al., 2000).

The most similar to traditional conjoint analysis method is multi-profile Best-Worst Scaling method because respondents are asked to select full profiles among a set of profiles. Case 3 BWS sometimes is also called conjoint BWS. However, unlike traditional conjoint analysis wherein respondents are asked to rank or rate profiles, the Case 3 Best-Worst Scaling tasks respondents with selecting best and worst profiles.

Regardless of the BWS method type, collecting two responses (best-worst choices) provides more specific data about respondents' preferences for items than can be obtained through conjoint analysis. The key assumption is that the best-worst choice captures the greatest difference in perceived importance among items on an underlying ranking of importance. Comparisons have shown (Cheng et al., 2023) that BWS methods offer advantages over other approaches, such as superior discriminatory power without increasing respondent burden and higher predictive validity, supporting empirical research using these methods.

Analyzing of responses in BWS methods can be made with 2 approaches – a counting approach and a modeling approach. The counting approach calculates several types of scores based on number of times (frequency) that item i is selected as the best (B_{in}) or the worst (W_{in}) in all the questions for respondent *n* (Finn, Louviere, 1992; Lee et al., 2007; Cohen 2009; Mueller et al., 2009). The scores are divided into two categories – disaggregated (individual level) and aggregated (total level) scores.

The disaggregated BW (best-worst) scores and its standardized scores are calculated based on the following formulas (Finn, Louviere, 1992; Lee et al., 2007; Louviere, Flynn, 2010):

$$BW_{in} = B_{in} - W_{in}, \qquad std.BW_{in} = \frac{BW_{in}}{r} \tag{6}$$

where *r* – the frequency with which item *i* appears across all questions.

The aggregated version of BW score and its standardized score as well the square root of the ratio and its standardized score are calculated based on the following formulas:

$$BW_i = B_i - W_i, \qquad std.BW_i = \frac{BW_i}{Nr} \tag{7}$$

$$sqrt.BW_i = \sqrt{\frac{B_i}{W_i}}, \qquad std.sqrt.BW_i = \frac{sqrt.BW_i}{max.sqrt.BW_i} \tag{8}$$

The modeling approach uses discrete choice models to analyze respondents' responses. The probability of selecting item *i* as the best and item *j* as the worst is expressed as a conditional logit model (CLM) and calculated as maxdiff (9), marginal (10) or marginal sequential (11) models using formulas:

$$Pr(i,j) = \frac{exp(v_i - v_j)}{\sum_{p,q;p \neq q} exp(v_p - v_q)} \tag{9}$$

$$P\,r(i,j) = \frac{exp(v_i)}{\sum_p exp(v_p)} \cdot \frac{exp(-v_j)}{\sum_p exp(-v_p)} \tag{10}$$

$$Pr(i,j) = \frac{exp(v_i)}{\sum_p exp(v_p)} \cdot \frac{exp(-v_j)}{\sum_q exp(-v_q)} \tag{11}$$

In the paper multi-profile Best-Worst Scaling (Case 3 BWS) method using maxdiff[2] model was used.

[2] Maxdiff (Maximum Difference Scaling) model assumes that respondents make selections because the difference in utility between *i* and *j* represents the greatest utility difference among m × (m − 1) possible utility differences, where m × (m − 1) is the number of possible pairs in which profile *i* is selected as the best profile and profile *j* is selected as the worst profile from *m* profiles.

Best-Worst Scaling has some limitations that should be considered. First, as the number of attributes per task increases (especially beyond 5-6), the decision-making process becomes more demanding, potentially leading to errors or random responses due to cognitive overload. Second, BWS method assumes that attributes are evaluated independently, while in reality, their importance may change depending on context and the interaction with other product features. Lastly, advanced statistical analysis is required to transform raw data into meaningful insights, making the interpretation of results more complex.

More information about multi-profile Best-Worst Scaling method and other BWS methods and their practical applications, can be found in subject literature (Lee et al., 2007; Louviere, 1991; Louviere, Flynn, 2010; Marley, Louviere, 2005; Flynn et al., 2007; Flynn, 2010; Fogarty, Aizaki, 2018; Aizaki et al., 2015; Marley, 2010; Cohen, 2023).

4. Results

In the measurement and analysis of stated preferences of users streaming the music from popular internet services and platforms (Spotify, Apple Music, YouTube Music, Tidal, Deezer, etc.) multi-profile Best-Worst Scaling (Case 3 BWS) method was used first. After collecting and then preparing data on respondents' preferences using modeling approach based on discrete choice model (conditional logit model), the best and the worst attributes as well as attribute levels were identified.

All calculations were carried out using the R program with `support.CEs` (Aizaki, 2023), `support.BWS` (Aizaki, 2023), `support.BWS3` (Aizaki, 2024) and `survival` (Therneau et al., 2024) packages for Best-Worst Scaling and the `conjoint` (Bąk, Bartłomowicz, 2018a) package for conjoint analysis. In the construction of experimental designs the `crossdes` (Sailer, 2022) and `DoE.base` (Groemping, 2023) packages were used. For the purpose of visualization of the results, the `broom` (Robinson et al., 2024), `fpc` (Hennig, 2024) and `ggplot2` (Wickham, 2016; Wickham et al., 2024) packages were applied:

```
> library(broom),
> library(conjoint),
> library(crossdes),
> library(DoE.base),
> library(fpc),
> library(ggplot2),
> library(support.BWS),
> library(support.BWS3),
> library(support.CEs),
> library(survival).
```

The following features (with the respective levels) were listed in the set of variables describing the examined product: quality of music (standard, high, lossless)[3], offline mode (disabled, playlists, full)[4], number of titles (under 10000, between 10000-40000, over than 40000) and subscription (student, multi, family)[5]:

```
>  stream.bws3.ffd<-list(
+ quality=c("standard","high","lossless"),
+ offline=c("disabled","playlists","full"),
+ titles=c("under_10k","b10_40k","over_40k"),
+ subscription=c("student","multi","family"))
```

Similar to discrete choice experiments, in Case 3 BWS method, question consists at least 3 (or more) profiles. Each profile has 2 (or more) attributes with each attribute having 2 (or more) levels. Consequently, the profile is expressed as a combination of attribute-levels. Respondents were asked to choose the best and worst profiles from a question.

The number of variables (4) combined with the number of their levels (3) enables the construction of 81 (full factorial design) different profiles of streaming music. Due to the respondents' limited capacity to evaluate a large number of profiles, a final set of 9 profiles using `oa.design` function from `DoE.base` R package was selected:

```
>  stream.oa.des<-oa.design(
+ nl=c(3,3,3,3),
+ randomize=FALSE)
>  stream.oa.des
  A B C D
1 1 1 1 1
2 1 2 3 2
3 1 3 2 3
4 2 1 3 3
5 2 2 2 1
6 2 3 1 2
7 3 1 2 2
8 3 2 1 3
9 3 3 3 1
class=design, type= oa
```

The design in a form of fractional factorial design meets the criterion of orthogonality:

```
>  ca<-as.numeric(unlist(stream.oa.des))
>  stream.ca.des<-as.data.frame(matrix(ca,nrow=9,ncol=4))
>  colnames(stream.ca.des)<-names(stream.bws3.ffd)
>  names(stream.ca.des)<-names(stream.bws3.ffd)
>  round(cor(stream.ca.des),5)

             quality offline titles subscription
quality            1       0      0            0
offline            0       1      0            0
titles             0       0      1            0
subscription       0       0      0            1
```

[3] Quality of music depends on bitrate. Standard quality means low bitrate (no more than 192 kbps). High bitrate is between 256 kbps - 320 kbps. Lossless music means the best bitrate (over 700 kbps) or quite lossless music (e.g. FLAC, ALAC).

[4] It is more comfortable to stream the music with full access to the libraries. Disabled mode means you can't download the music, playlists – only playlist are available to download, full – access to download all libraries.

[5] The student subscription means access for 1 device and lower fee, multi – max 3 devices and medium fee, family – access for max 4 persons and highest fee.

```
>  det(cor(stream.ca.des))
[1] 1
```

The resultant design is a matrix with 9 rows and 4 columns. The columns correspond to attributes, while the rows correspond to profiles (cf. tab. 1). For example, profile 3 consists the following attribute levels: standard quality of music (1), full access mode (3), between 10k and 40k titles (2) and family type of subscription (3).

Next, balanced incomplete block design (BIBD) is needed. The `find.BIB` function from `crossdes` package assigns profiles from fractional factorial design and creates BIBD design. All 9 profiles (trt = 9) have to be used, at least 3 in each question (k = 3), and such a number of questions (b = 12) that this design also meets the criterion of orthogonality:

```
>  stream.bibd.des<-find.BIB(trt=9,b=12,k=3)
>  stream.bibd.des
      [,1] [,2] [,3]
 [1,]    5    6    7
 [2,]    1    2    6
 [3,]    4    7    8
 [4,]    3    4    6
 [5,]    1    5    8
 [6,]    2    7    9
 [7,]    3    5    9
 [8,]    2    4    5
 [9,]    1    4    9
[10,]    2    3    8
[11,]    1    3    7
[12,]    6    8    9
>  isGYD(stream.bibd.des)

[1] The design is a balanced incomplete block design w.r.t. rows.
```

Table 1.
Fractional factorial design

Number of profile	Attributes of food service			
	Quality of music	**Offline mode**	**Number of titles**	**Type of subscription**
1	standard	disabled	under 10k	student
2	standard	only playlists	over 40k	multi
3	standard	full access	between 10k and 40k	family
4	high	disabled	over 40k	family
5	high	only playlists	between 10k and 40k	student
6	high	full access	under 10k	multi
7	lossless	disabled	between 10k and 40k	multi
8	lossless	only playlists	under 10k	family
9	lossless	full access	over 40k	student

Source: author's compilation using `DoE.base` R package.

In the BIBD design each row corresponds to question, while each column corresponds to profiles. For example, row 6 means a set of 3 profiles: 2, 7 and 9 (cf. Table 1). The result of executing the function `isGYD` (also from `crossdes` R package) indicates that the resultant design is a balanced incomplete block design.

The fractional factorial design of profiles (stream.oa.des) and balanced incomplete block design of questions (stream.bibd.des) allows using the `bws3.design` function to generate a questionnaire design (stream.bws3.des) for respondents in accordance with the assumptions of the Case 3 BWS method:

```
> stream.bws3.des<-bws3.design(
+   bibd=stream.bibd.des,
+   ffd=stream.oa.des,
+   attribute.levels=stream.bws3.fd)
> questionnaire(stream.bws3.des)

Block 1

Question 1
             alt.1       alt.2       alt.3
quality      "high"      "high"      "lossless"
offline      "playlists" "full"      "disabled"
titles       "b10_40k"   "under_10k" "b10_40k"
subscription "student"   "multi"     "multi"

Question 2
             alt.1       alt.2       alt.3
quality      "standard"  "standard"  "high"
offline      "disabled"  "playlists" "full"
titles       "under_10k" "over_40k"  "under_10k"
subscription "student"   "multi"     "multi"

Question 3
             alt.1      alt.2      alt.3
quality      "high"     "lossless" "lossless"
offline      "disabled" "disabled" "playlists"
titles       "over_40k" "b10_40k"  "under_10k"
subscription "family"   "multi"    "family"

Question 4
             alt.1      alt.2      alt.3
quality      "standard" "high"     "high"
offline      "full"     "disabled" "full"
titles       "b10_40k"  "over_40k" "under_10k"
subscription "family"   "family"   "multi"

Question 5
             alt.1       alt.2       alt.3
quality      "standard"  "high"      "lossless"
offline      "disabled"  "playlists" "playlists"
titles       "under_10k" "b10_40k"   "under_10k"
subscription "student"   "student"   "family"

Question 6
             alt.1       alt.2      alt.3
quality      "standard"  "lossless" "lossless"
offline      "playlists" "disabled" "full"
titles       "over_40k"  "b10_40k"  "over_40k"
subscription "multi"     "multi"    "student"

Question 7
             alt.1      alt.2       alt.3
quality      "standard" "high"      "lossless"
offline      "full"     "playlists" "full"
titles       "b10_40k"  "b10_40k"   "over_40k"
```

```
subscription "family"   "student"   "student"

...

Question 12
             alt.1       alt.2       alt.3
quality      "high"      "lossless"  "lossless"
offline      "full"      "playlists" "full"
titles       "under_10k" "under_10k" "over_40k"
subscription "multi"     "family"    "student"
```

In the research the respondents' responses from a survey conducted in 2024-2025 were used. The survey questionnaire included questions for Best-Worst Scaling method, questions for conjoint analysis method and questions about basic respondents' characteristics. The survey was prepared using Microsoft Forms and distributed using Microsoft Teams programs. The data were collected using employing the convenience sampling method, which involved selecting respondents based on their availability and willingness to participate in the study. In the research, 108 correctly completed questionnaires were used as a source of statistical data for both methods. The data containing respondents' preferences (stream.bws3.pref) for the Case 3 BWS method are as follows:

```
> stream.bws3.pref<-read.csv2("bws3.csv",header=TRUE)
> stream.bws3.pref<-stream.bws3.pref[,1:28]
> head(stream.bws3.pref);tail(stream.bws3.pref)
> stream.bws3.pref<-stream.bws3.pref[,1:26]
  id BLOCK B1 W1 B2 W2 B3 W3 B4 W4 B5 W5 B6 W6 B7 W7 B8 W8 B9 W9 B10 W10 B11 W11
1  1     1  3  2  2  1  2  3  2  3  3  2  2  1  3  1  3  2  3  1   2   1   1   2
2  2     1  2  1  3  1  2  1  2  1  2  3  3  2  1  3  2  1  2  1   3   1   3   1
3  3     1  3  1  3  1  2  1  3  1  3  1  3  1  3  1  3  1  3  1   2   1   3   1
4  4     1  2  1  2  1  3  1  3  1  3  1  2  3  1  3  3  2  3  1   2   1   2   3
5  5     1  3  1  2  1  1  3  2  3  1  2  2  3  1  2  2  1  2  1   2   1   3   1
6  6     1  3  2  1  3  2  1  1  2  3  2  2  3  3  1  3  2  3  2   2   3   3   2
  B12 W12 sex age
1   3   2   0  20
2   2   3   0  21
3   3   1   0  20
4   3   1   1  20
5   1   2   0  20
6   3   2   1  20
     id BLOCK B1 W1 B2 W2 B3 W3 B4 W4 B5 W5 B6 W6 B7 W7 B8 W8 B9 W9 B10 W10 B11
103 103     1  3  2  3  1  1  3  2  1  3  2  2  3  1  3  3  2  3  1   2   1   2
104 104     1  2  1  2  1  2  1  2  1  2  1  2  1  2  1  2  3  2  1   2   1   2
105 105     1  2  1  2  1  3  1  3  1  3  1  2  1  2  3  3  2  3  1   3   1   2
106 106     1  3  1  1  3  3  1  1  3  3  2  2  1  2  1  3  2  3  2   2   1   3
107 107     1  1  2  3  1  1  3  3  1  1  2  2  3  3  2  1  3  3  1   2   3   3
108 108     1  3  1  2  1  3  1  3  1  3  1  2  3  1  3  3  2  3  1   3   1   2
    W11 B12 W12 sex age
103   3   3   2   0  20
104   1   2   1   1  20
105   3   3   1   1  22
106   1   3   1   1  20
107   2   1   2   0  20
108   3   3   1   0  21
```

The multi-profile Best-Worst Scaling method employs a model approach requiring the definition of conditional logit model formula, whose parameters are then estimated. All necessary elements, presented in the form of an appropriate dataset, are provided by the `bws3.dataset` function:

```
> stream.bws3.dat<-bws3.dataset(
+   data=stream.bws3.pref,
+   response=colnames(stream.bws3.pref)[3:26],
+   choice.sets=stream.bws3.des,
+   categorical.attributes=names(stream.bws3.ffd),
+   optout=FALSE,
+   asc=c(0,0,0),
+   model="maxdiff")
> head(stream.bws3.dat);tail(stream.bws3.dat)
  id BLOCK QES PAIR BEST WORST RES.B RES.W RES ASC1 ASC2 ASC3 standard high
1  1     1   1    1    1     2     3     2   0    0    0    0        0    0
2  1     1   1    2    1     3     3     2   0    0    0    0        0    1
3  1     1   1    3    2     1     3     2   0    0    0    0        0    0
4  1     1   1    4    2     3     3     2   0    0    0    0        0    1
5  1     1   1    5    3     1     3     2   0    0    0    0        0   -1
6  1     1   1    6    3     2     3     2   1    0    0    0        0   -1
  lossless disabled playlists full under_10k b10_40k over_40k student multi
1        0        0         1   -1        -1       1        0       1    -1
2       -1       -1         1    0         0       0        0       1    -1
3        0        0        -1    1         1      -1        0      -1     1
4       -1       -1         0    1         1      -1        0       0     0
5        1        1        -1    0         0       0        0      -1     1
6        1        1         0   -1        -1       1        0       0     0
  family  STR
1      0 1010
2      0 1010
3      0 1010
4      0 1010
5      0 1010
6      0 1010
      id BLOCK QES PAIR BEST WORST RES.B RES.W RES ASC1 ASC2 ASC3 standard high
7771 108     1  12    1    1     2     3     1   0    0    0    0        0    1
7772 108     1  12    2    1     3     3     1   0    0    0    0        0    1
7773 108     1  12    3    2     1     3     1   0    0    0    0        0   -1
7774 108     1  12    4    2     3     3     1   0    0    0    0        0    0
7775 108     1  12    5    3     1     3     1   1    0    0    0        0   -1
7776 108     1  12    6    3     2     3     1   0    0    0    0        0    0
     lossless disabled playlists full under_10k b10_40k over_40k student multi
7771       -1        0        -1    1         0       0        0       0     1
7772       -1        0         0    0         1       0       -1      -1     1
7773        1        0         1   -1         0       0        0       0    -1
7774        0        0         1   -1         1       0       -1      -1     0
7775        1        0         0    0        -1       0        1       1    -1
7776        0        0        -1    1        -1       0        1       1     0
     family    STR
7771     -1 108120
7772      0 108120
7773      1 108120
7774      1 108120
7775      0 108120
7776     -1 108120
```

In the formula, for each attribute one attribute level should be omitted. In the research, the last attribute levels (lossless, full, over_40k, family) were assumed as reference levels for the corresponding attributes (BWS model 1). The structure of the conditional logit model is similar to that of a linear regression function:

$$\text{RES} \sim \text{standard} + \text{high} + \text{disabled} + \text{playlists} + \text{under\_10k} + \text{b\_10\_40k} + \text{student} + \text{multi} + \text{strata(STR)}$$

```
> stream.bws3.model<-RES ~ standard + high + disabled + playlists +
under_10k + b10_40k + student + multi + strata(STR)
> stream.bws3.clm<-clogit(
+   formula=stream.bws3.model,
+   data=stream.bws3.dat)
```

```
> stream.bws3.clm
> gofm(stream.bws3.clm)

Call:
clogit(formula = stream.bws3.model, data = stream.bws3.dat)

              coef exp(coef) se(coef)       z        p
standard  -0.54137   0.58195  0.04940 -10.959  < 2e-16
high      -0.27899   0.75655  0.04716  -5.915 3.31e-09
disabled  -0.34768   0.70632  0.04864  -7.147 8.84e-13
playlists -0.18235   0.83331  0.04688  -3.890 0.000100
under_10k -0.18413   0.83183  0.04813  -3.826 0.000130
b10_40k    0.16107   1.17477  0.04696   3.430 0.000604
student   -0.30042   0.74050  0.04834  -6.215 5.13e-10
multi     -0.03358   0.96698  0.04685  -0.717 0.473552

Likelihood ratio test=262.4  on 8 df, p=< 2.2e-16
n= 7776, number of events= 1296

Rho-squared = 0.05649267
Adjusted rho-squared = 0.05304754
Akaike information criterion (AIC) = 4397.875
Bayesian information criterion (BIC) = 4439.211
Number of coefficients = 8
Log likelihood at start = -2322.12
Log likelihood at convergence = -2190.938
```

The coefficients of omitted variables are normalized to 0. Therefore, other coefficients indicate how each variable affects the probability of selection (utility) compared to the reference category. The strata(STR) denotes that variable STR is used to identify each combination of respondent and Case 3 BWS question.

According to the obtained results, standard (-0,54) quality of streaming music significantly decreases the probability of selection, indicating that users prefer better quality – at least high (-0,28) or rather lossless (0) quality of music (cf. Figure 1). Downloading restrictions negatively impact the respondents' choices. Users significantly dislike the disabled (-0,35) offline mode (no music downloads available). Limiting downloads to only playlists (-0,18) reduces the dislike but it is still strong. The best (most preferred) offline mode is full (0) access to libraries.

Music libraries' size also influences streamers' preferences. The worst (least preferred) level is under 10k (-0,18) titles. The results show that number between 10k and 40k (0,16) titles is enough and even more preferred than over 40k (0) titles. For subscription type, the student (-0,31) plan restricted to only single device is the worst rated. This group may need additional incentives, such as discounted lossless streaming or expanded download options. The multi device (-0,03) plan is also not a key factor, meaning users may not strongly consider the number of devices allowed when choosing a subscription. Instead, family type of subscription (0) and factors like music quality and access to downloads play a more significant role.

Using traditional conjoint analysis method the results of multi-profile Best-Worst Scaling were confronted and next results were determined. Respondents evaluated the same 9 profiles[6]

[6] The profiles from the stream.oa.des design.

on an interval scale [0-10] considering the relative attractiveness of the profiles and assigning a higher value to the profile that was more attractive than the others. It means that the data was collected as a form of rating:

```
> stream.ca.pref<-read.csv2("conjoint.csv",header=TRUE)
> head(stream.ca.pref);tail(stream.ca.pref)
  profil1 profil2 profil3 profil4 profil5 profil6 profil7 profil8 profil9
1       4       6       5       6       8       6       4       5       7
2       0       4       6       6       2       7       6       6       3
3       2       4       5       2       6       9       2       7      10
4       2       4       5       5       5       4       4       5       5
5       2       4       5       6       7       7       0       0       2
6       4       3       2       2       7       4       3       2       9
    profil1 profil2 profil3 profil4 profil5 profil6 profil7 profil8 profil9
103       1       3       5       2       4       6       3       4       6
104       3       4       8       8       8       8       8       7       7
105       2       4       3       6       4       4       6       4       5
106       4       4       6       6       9       7       7       7      10
107       1       8      10       3       9       8       2       5       6
108       3       5       5       6       7       4       6       4       5
```

In the `conjoint` R package, the main function `Conjoint` needs 3 arguments. Apart from the data set on empirical respondents' preferences, the fractional factorial design and attribute levels' names are necessary:

```
> stream.ca.des
  quality offline titles subscription
1       1       1      1            1
2       1       2      3            2
3       1       3      2            3
4       2       1      3            3
5       2       2      2            1
6       2       3      1            2
7       3       1      2            2
8       3       2      1            3
9       3       3      3            1
> levn<-cbind(
+   stream.bws3.ffd$quality,
+   stream.bws3.ffd$offline,
+   stream.bws3.ffd$titles,
+   stream.bws3.ffd$subscription)
> stream.levn<-c(unlist(levn))
> stream.ca.levn<-stream.levn
> stream.ca.levn
 [1] "standard"  "high"      "lossless"  "disabled"  "playlists" "full"
 [7] "under_10k" "b10_40k"   "over_40k"  "student"   "multi"     "family"
```

The availability of data on empirical preferences (stream.ca.pref), the coded research design (stream.ca.des) and the names of attribute levels (stream.ca.levn) allow summarizing (in the cross-section of respondents) the most important results of measurement of preferences using the `Conjoint` function (TCA model):

```
> Conjoint(stream.ca.pref,stream.ca.des,stream.ca.levn)

Call:
lm(formula = frml)
Residuals:
   Min      1Q Median      3Q    Max
-4,926 -1,407 -0,037  1,296  4,741
```

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               4,971193   0,064358  77,243  < 2e-16 ***
factor(x$quality)1       -0,773663   0,091016  -8,500  < 2e-16 ***
factor(x$quality)2        0,534979   0,091016   5,878 5,72e-09 ***
factor(x$offline)1       -0,971193   0,091016 -10,671  < 2e-16 ***
factor(x$offline)2        0,251029   0,091016   2,758  0,00592 **
factor(x$titles)1        -0,823045   0,091016  -9,043  < 2e-16 ***
factor(x$titles)2         0,572016   0,091016   6,285 4,97e-10 ***
factor(x$subscription)1 -0,144033   0,091016  -1,582  0,11386
factor(x$subscription)2  0,004115   0,091016   0,045  0,96395
---
Signif. codes:  0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1
Residual standard error: 2,006 on 963 degrees of freedom
Multiple R-squared:   0,23,     Adjusted R-squared:  0,2236
F-statistic: 35,96 on 8 and 963 DF,  p-value: < 2,2e-16
 [1] "Part worths (utilities) of levels (model parameters for whole sample):"
      levnms    utls
1  intercept  4,9712
2   standard -0,7737
3       high   0,535
4   lossless  0,2387
5   disabled -0,9712
6  playlists   0,251
7       full  0,7202
8  under_10k  -0,823
9    b10_40k   0,572
10  over_40k   0,251
11   student  -0,144
12     multi  0,0041
13    family  0,1399
[1] "Average importance of factors (attributes):"
[1] 28,34 29,92 22,76 18,99
[1] Sum of average importance:  100,01
[1] "Chart of average factors importance"
```

According to the obtained results from conjoint analysis method, standard (-0,77) quality of music, disables (-0,97) offline mode, under 10k (-0,82) titles and student (-0,14) type of subscription are also (like for BWS method) the least preferred levels of the used variables (cf. Figure 1). The most preferred levels of music quality are high (0,53) and lossless (0,24), which switch their rankings between both methods. Next most preferred attribute levels are the same: full (0,72) access to libraries, between 10k and 40k (0,57) titles and family (0,14) subscription. Similar to BWS method, also rest of unmentioned attribute levels: only playlists (0,25), over 40k (0,25) titles and multi device (0,004) subscription take second place:

```
> df<-data.frame(names=stream.levn,coef=0)
> coef<-stream.bws3.clm$coef
> df[1,2]=coef[1];df[2,2]=coef[2];
> df[4,2]=coef[3];df[5,2]=coef[4];
> df[7,2]=coef[5];df[8,2]=coef[6];
> df[10,2]=coef[7];df[11,2]=coef[8];
> df$type<-ifelse(df$coef>=0,"above","below")
> df$names<-factor(df$names,levels=rev(df$names))
> ggplot(df,aes(x=names,y=coef))+
+ xlab("Atrributes' levels")+ylab("Probabilities")+
+ geom_bar(position='stack',stat='identity',width=.9,aes (fill=type))+
+ scale_fill_manual(values=c("above"="#9fc79a","below"="#C79a9a"))+
```

```
+ theme(legend.position='none',axis.title=element_text(size=12),
+ axis.text=element_text(size=12))+
+ geom_text(aes(label=round(coef,3),y=coef+.00),size=4)+
+ coord_flip()+geom_hline(yintercept=0)

> stream.ca.util<-caUtilities(stream.ca.pref,stream.ca.des, stream.ca.levn)
> util<-stream.ca.util[2:13]
> df<-data.frame(names=stream.ca.levn[1:12],util)
> df$type<-ifelse(df$util>=0,"above","below")
> df$names<-factor(df$names,levels=rev(df$names))
> ggplot(df,aes(x=names,y=util))+
+ xlab("Atrributes' levels")+ylab("Part-worth utilities")+
+ geom_bar(position='stack',stat='identity',width=.9,aes(fill=type))+
+ scale_fill_manual(values=c("above"="#9fc79a","below"="#C79a9a"))+
+ theme(legend.position='none',axis.title=element_text(size=12),
+ axis.text=element_text(size=12))+
+ geom_text(aes(label=round(util,3),y=util+.00),size=4)+
+ coord_flip()+geom_hline(yintercept=0)
```

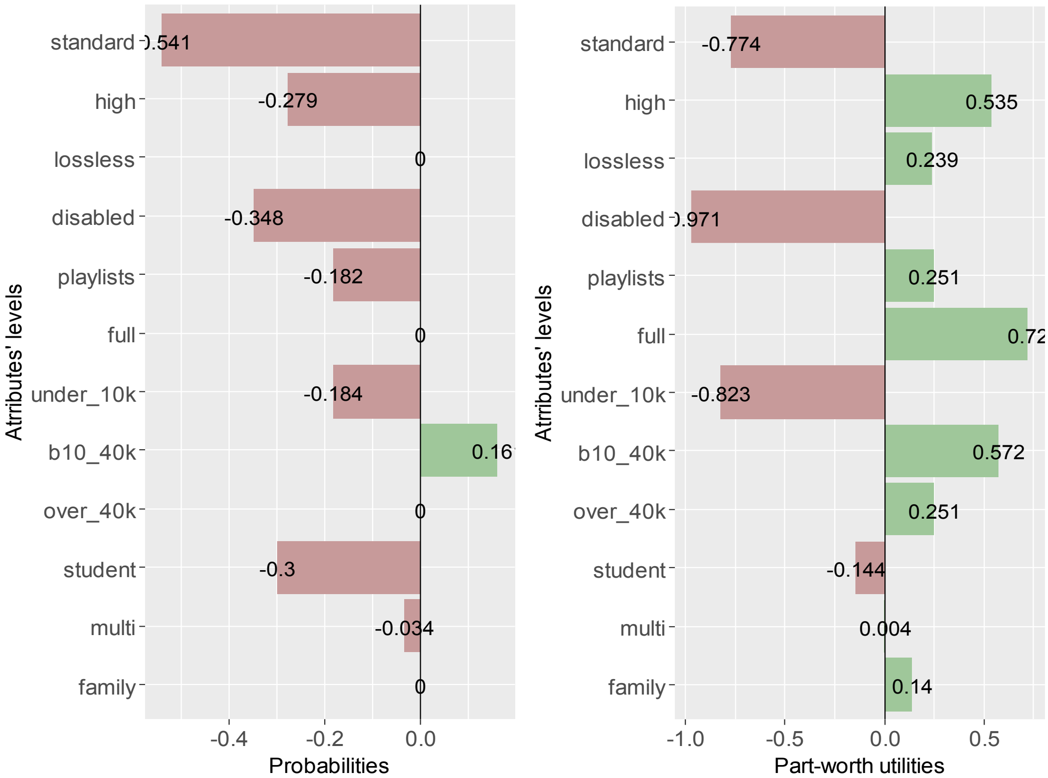

Figure 1. Probabilities (left) and parth-worth utilities (right) of attribute levels.

Source: author's compilation using `support.BWS3` and `conjoint` R packages.

Since the results of Best-Worst Scaling method are based on probabilities while the conjoint analysis method uses utilities, a direct comparison of these results is inappropriate. In this case, the character of the preferences should be rather compared. Figure 2 shows the probabilities and part-worth utilities of attribute levels according to BWS model 1 and TCA with the reference levels being the last attribute levels. Figure 3 presents the same for BWS model 2, where the reference levels are the first attribute levels (BWS model 2):

```
> stream.bws3.model<-RES ~ high + lossless + playlists + full + b10_40k +
                         over_40k + multi + family + strata(STR)
> stream.bws3.clm<-clogit(
+   formula=stream.bws3.model,
+   data=stream.bws3.dat)
> stream.bws3.clm
Call:
clogit(formula = stream.bws3.model, data = stream.bws3.dat)

             coef exp(coef) se(coef)      z        p
high      0.26238   1.30002  0.04823  5.440 5.33e-08
lossless  0.54137   1.71835  0.04940 10.959  < 2e-16
playlists 0.16533   1.17978  0.04824  3.427  0.00061
full      0.34768   1.41578  0.04864  7.147 8.84e-13
b10_40k   0.34520   1.41227  0.04845  7.124 1.05e-12
over_40k  0.18413   1.20217  0.04813  3.826  0.00013
multi     0.26685   1.30584  0.04841  5.512 3.55e-08
family    0.30042   1.35043  0.04834  6.215 5.13e-10

Likelihood ratio test=262.4  on 8 df, p=< 2.2e-16
n= 7776, number of events= 1296
```

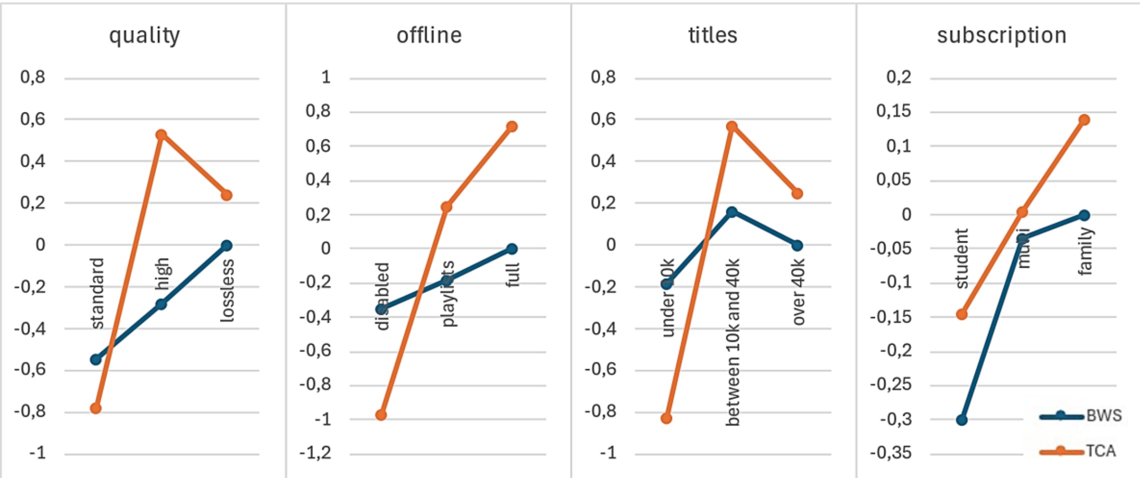

Figure 2. Probabilities and parth-worth utilities of attribute levels – model 1.

Source: author's compilation using `support.BWS3` and `conjoint` R packages.

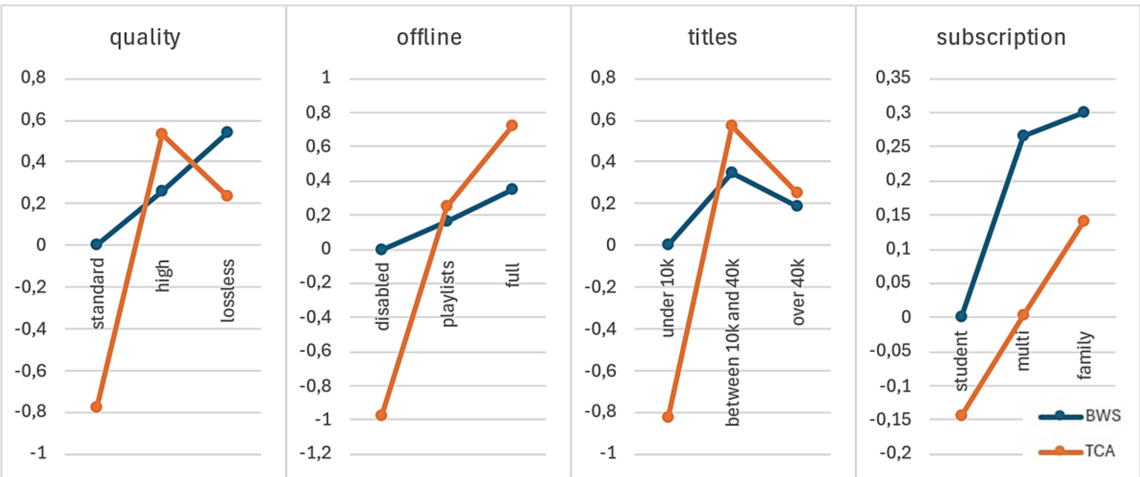

Figure 3. Probabilities and parth-worth utilities of attribute levels – model 2.

Source: author's compilation using `support.BWS3` and `conjoint` R packages.

The difference between Figures 2 and 3 lies in the vertical positioning of the BWS and TCA graphs, meaning that the shape of the graph is much more important than its location. As we can see, almost every pair of BWS and TCA preferences align closely, except for the previously noted switch between high and lossless levels of the music quality attribute.

What differs the most between the both methods is the importance of the attributes used in the study (cf. Figure 4). According to Best-Worst Scaling method, quality of music (40,42%) is the most important, the second place takes offline mode (26,11%). In contrast, traditional conjoint analysis identifies offline mode (29,92%) as the most important, with quality of music (28,34%) ranked on second place. The least important attribute in both methods is the type of subscription (16,46% in BWS / 18,99% in TCA):

```
> coef<-abs(stream.bws3.clm$coef)
> impo<-c(mean(coef[1:2]),mean(coef[3:4]),mean(coef[5:6]),mean(coef[7:8]))
> stream.bws3.impo=impo/sum(impo)*100
> stream.ca.impo=caImportance(stream.ca.pref,stream.ca.des)
> df=data.frame(names=names(stream.bws3.ffd),stream.bws3.impo)

> df$names<-factor(df$names,levels=c("quality","offline","titles","subscription"))
> ggplot(df,aes(x=names,y=stream.bws3.impo))+
+ xlab("Attributies")+ylab("Importance [%]")+
+ geom_bar(stat='identity',width=.9,fill="#9fc79a")+
+ theme(legend.position="none",axis.title=element_text(size=12),
+ axis.text=element_text(size=12))+
+ geom_text(aes(label=round(stream.bws3.impo,3),y=stream.bws3.impo+.00),size=4)+
+ geom_hline(yintercept=0)

> df=data.frame(names=colnames(stream.ca.des),stream.ca.impo)
> df$names<-factor(df$names,levels=c("quality","offline","titles","subscription"))
> ggplot(df,aes(x=names,y=stream.ca.impo))+
+ xlab("Attributies")+ylab("Importance [%]")+
+ geom_bar(stat='identity',width=.9,fill="#9fc79a")+
+ theme(legend.position="none",axis.title=element_text(size=12),
+ axis.text=element_text(size=12))+
+ geom_text(aes(label=round(stream.ca.impo,3),y=stream.ca.impo+.00),size=4)+
+ geom_hline(yintercept=0)
```

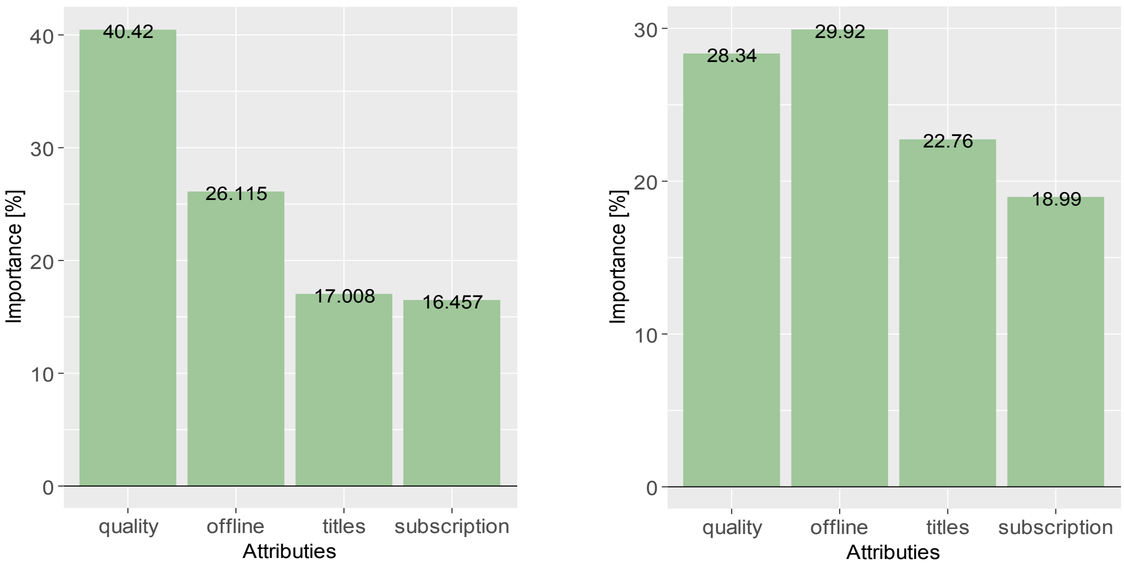

Figure 4. Importance of attributes according to BWS (left) and TCA (right).

Source: author's compilation using `support.BWS3` and `conjoint` R packages.

5. Discussion

Both conditional logit models show consistent trends, with streamers clearly preferring higher and lossless music quality, family subscription plan and access to full playlists. According to both models, a library between 10k and 40k titles is considered sufficient. The key distinction between the models is in how the results are presented. BWS model 1 emphasizes the negative aspects of less preferred options, while BWS model 2 highlights the positive aspects of more preferred choices.

The results indicate that both BWS and TCA models fit the data equally well, as confirmed by the identical likelihood ratio test (Likelihood ratio test = 262,4, p-value < 2,2e-16) and F-statistic test (F-statistic test = 35,96, p-value < 2,2e-16). It means that regardless of the chosen reference method, the analysis leads to similar conclusions regarding streamers' preferences.

A closer analysis of both models indicates that in BWS model 1, the multi device level of subscription type variable has a p-value of 0,47. It means that the effect is not statistically significant (p-value > 0,05). In TCA model both subscription factors – student (p-value of 0,12) and multi device (p-value of 0,96) are also not statistically significant.

A potential solution could be measurement of preferences after dividing respondents into homogeneous segments (Henig, 2024). The simple clustering music streamers using *k*-means method confirm 3 separately classes which can correspond with levels (student, multi, family) of subscription attribute (cf. Figure 5):

```
> stream.ca.segm<-caSegmentation(stream.ca.pref,stream.ca.des,c=3)
> summary(stream.ca.segm)
     Length Class  Mode
segm   9    kmeans list
util 972    -none- numeric
sclu 108    -none- numeric
> plotcluster(stream.ca.segm$util,stream.ca.segm$sclu)
> stream.ca.dcf<-discrcoord(stream.ca.segm$util,stream.ca.segm$sclu)
> assignments<-augment(stream.ca.segm$segm,stream.ca.dcf$proj[,1:2])
> ggplot(assignments)+geom_point(aes(x=X1,y=X2,color= .cluster))+
+ labs(color="Class Assignment",title="Clustering Results")
```

Figure 5. Segmentation of music streamers.

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

6. Conclusions

The paper explores the application of Best-Worst Scaling and conjoint analysis methods in the measurement and analysis of stated preferences. The objectives of the paper were to assess the music streamers' preferences, compare the results of both used methods and demonstrate the cooperation of `support.BWS3` and `conjoint` R packages as complementary analytical tool.

The research results showed that music streamers indicate different importance of streaming attributes and attribute levels. Although the importance of attributes differs in percentage terms, according to both methods, the most important for respondents are quality of music and offline mode, while the least important are the number of titles and type of subscription. In the case of attribute levels, the most attractive are full access to libraries of 10-40 thousand songs, family subscription type and high or lossless music quality. In the case of these two levels, the results switch their rankings between both methods. According to the results of all models, respondents indicate standard music quality, disabled offline mode, number of songs under 10k and student subscription type as the least attractive levels.

Using the R packages the probabilities and part-worth utilities of attribute levels according to both BWS and TCA models with different reference levels were calculated. The character of almost all pair of attribute levels are similar. It means that the confrontation of the results confirmed the compatibility of conclusions from both used methods.

The paper also demonstrates that combining used R packages allows for measurement and analysis of preferences, making it useful for practitioners, researchers and students in the fields of marketing research, in particular in the area of measurement of consumers' preferences. In particular, streaming companies, manufacturers of playback equipment, artist and record labels as well as marketers should be interested in the research results.

Streaming platforms can use the research results to adjust their subscription plans, focusing on higher music quality and comprehensive offline access. Offering lossless music as a standard or premium option could serve as a competitive advantage. Additionally, emphasizing family subscription plans over student or multi device options may increase customer retention.

Manufacturers can capitalize on the demand for high quality music by promoting devices that support lossless formats. Partnering with streaming services to offer promotional packages with access to high quality music can increase user adoption and brand loyalty.

Record labels should prioritize offering catalogs in lossless quality rather than simply expanding their song libraries. Investing in high quality recording and production can provide a more compelling listening experience, aligning with consumer demand for superior sound quality.

Marketing strategies should emphasize user convenience, premium music quality, and offline access as key selling points. Campaigns should target families rather than just individual users, highlighting cost savings and shared access benefits. Furthermore, leveraging social media influencers and music enthusiasts to demonstrate the advantages of high quality music can increase engagement and consumer trust.

The paper leaves room for further exploration. There are more methods of measurement of stated preferences and more R packages for them. The research suggests potential for future studies to explore more about the music streamers' preferences.

References

1. Aizaki, H. (2023). *support.BWS: Tools for Case 1 Best-Worst Scaling*. https://CRAN.R-project.org/package=support.BWS
2. Aizaki, H. (2024). *support.BWS3: Tools for Case 3 Best-Worst Scaling*. https://CRAN.R-project.org/package=support.BWS3
3. Aizaki, H. (2024). *support.CEs: Basic Functions for Supporting an Implementation of Choice Experiments,* https://CRAN.R-project.org/package=support.CEs
4. Aizaki, H., Nakatani, T., Sato, K. (2015). *Stated Preference Methods Using R.* Chapman & Hall/CRC.
5. Allan, J., Leijonhufvud, S. (2022). *Listener Preferences in Streamed Music*. Sweden. Paper available from: https://aes2.org/publications/elibrary-page/?id=21557

6. Auger, P., Devinney, T.M., Louviere, J.J. (2007). Using best-worst scaling methodology to investigate consumer ethical beliefs across countries. *Journal of Business Ethics*, *70*, pp. 299-326.
7. Baek, B.H. (2023). Using conjoint analysis for a value estimation of the consumption attributes of online performances. *Cogent Business & Management*, *10(3).*
8. Bąk, A. (2004). *Dekompozycyjne metody pomiaru preferencji w badaniach marketingowych.* Wrocław: Wydawnictwo Akademii Ekonomicznej we Wrocławiu.
9. Bąk, A. (2013). *Mikroekonometryczne metody badania preferencji konsumentów z wykorzystaniem programu R.* Warszawa: C.H. Beck.
10. 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.* C.H. Beck, pp. 239-248.
11. Bąk, A., Bartłomowicz, T. (2018a). *conjoint: An Implementation of Conjoint Analysis Method.* https://CRAN.R-project.org/package=conjoint
12. Bąk, A., Bartłomowicz, T. (2018b). *Conjoint R package.* https://pl.wikipedia.org/wiki/Conjoint_R.
13. Bamert, T., Meier-Bickel, T.S., Rudt, C. (2005). *Pricing music downloads: A conjoint study.* Proceedings of the 2005 European Academy of Management annual conference*,* pp. 409-416.
14. Bartłomowicz, T., Bąk, A. (2021). *Badania preferencji i zachowań konsumentów z wykorzystaniem metod mikroekonometrii i programu R.* Wrocław: Wydawnictwo UE.
15. Breidert, C., Hahsler, M. (2007). Adaptive Conjoint Analysis for Pricing Music Downloads. In: R. Decker, H.J. Lenz (eds.), *Advances in Data Analysis. Studies in Classification, Data Analysis, and Knowledge Organization*. Springer.
16. Cheng, H., Lambert, M.D., DeLong, K.L., Jensen, K.L. (2021). Inattention, availability bias, and attribute premium estimation for a biobased product. *Agricultural Economics*, *53(2),* pp. 274-288.
17. Cheng, H., Zhang, T., Lambert, M.D., Feuz, R. (2023). An empirical comparison of conjoint and best-worst scaling case III methods. *Journal of Behavioral and Experimental Economics, Vol. 106*.
18. Chrzan, K., Golovashkina, N. (2006). An empirical test of six stated importance measures. *International Journal of Market Research*, *48(6),* pp. 717-740.
19. Cohen, E., Neira, L. (2004). Measuring preference for product benefits across countries: Overcoming scale usage bias with maximum difference scaling. *Excellence in International Research*, pp. 1-22.
20. Cohen, E. (2009). Applying best-worst scaling to wine marketing. *International Journal of Wine Business Research*, *21(1),* pp. 8-23.
21. Cohen, S.H. (2003). Maximum Difference Scaling: Improved Measures of Importance and Preference for Segmentation. *Sawtooth Software Conference Proceedings.*

22. Coombs, C.H., Dawes, R.M., Tversky, A. (1970). *Mathematical Psychology: An Elementary Introduction.* Englewood Cliffs: Prentice-Hall.
23. Finn, A., Louviere, J.J. (1992). Determining the Appropriate Response to Evidence of Public Concern: The Case of Food Safety. *Journal of Public Policy & Marketing*, *11(2),* pp. 12-25.
24. Flynn, T.N. (2010). Valuing citizen and patient preferences in health: Recent developments in three types of best–worst scaling. *Expert Review of Pharmacoeconomics & Outcomes Research*, *10(3),* pp. 259-267.
25. Flynn, T.N., Louviere, J.J., Peters, T.J., Coast, J. (2007). Best–worst scaling: What it can do for health care research and how to do it. *Journal of Health Economics*, *26(1),* pp. 171-189.
26. Fogarty, J., Aizaki, H. (2018) An Illustrative Example of Contingent Valuation. NMVR Project Team, *Non-Market Valuation with R*. http://lab.agr.hokudai.ac.jp/nmvr/
27. Green, P.E., Rao, V.R. (1971). Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research*, *8(September),* pp. 355-363.
28. Green, P.E., Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research*, *5(2),* pp. 103-123.
29. Green, P.E., Srinivasan, V. (1990). Conjoint analysis in marketing: New developments with implications for research and practice. *Journal of Marketing*, 54(October), pp. 3-19.
30. Green, P.E., Wind, Y. (1973). *Multiattribute decisions in marketing. A measurement approach*. Hinsdale, Illinois: Dryden Press.
31. Green, P.E., Wind, Y. (1975). New way to measure consumers' judgments. *Harvard Business Review*, *53(July/August),* pp. 107-117.
32. Groemping, U., Amarov, B., Xu, H. (2023). *DoE.base: Full Factorials, Orthogonal Arrays and Base Utilities for DoE Packages*, https://CRAN.R-project.org/package=DoE.base
33. Gustafsson, A., Herrmann, A., Huber, F. (Eds.) (2007). *Conjoint measurement. Methods and applications*. Berlin/Heidelberg/New York: Springer-Verlag.
34. Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C. (1995). *Multivariate data analysis with readings*. Englewood Cliff s: Prentice-Hall.
35. Hein, K.A., Jaeger, S.R., Carr, B.T., Delahunty, C.M. (2008). Comparison of five common acceptance and preference methods. *Food Quality and Preference*, *19(7),* pp. 651-661.
36. Henig, C. (2024). *fpc: Flexible Procedures for Clustering*, https://CRAN.R-project.org/package=fpc
37. Himmler, S., Soekhai, V., van Exel, J., Brouwer, W. (2021). What works better for preference elicitation among older people? Cognitive burden of discrete choice experiment and case 2 best-worst scaling in an online setting. *Journal of Choice Modelling*, *Vol. 38*.
38. Hollin., I.L., Peay., H.L., Bridges J.F. (2015). Caregiver Preferences for Emerging Duchenne Muscular Dystrophy Treatments: A Comparison of Best-Worst Scaling and Conjoint Analysis. *Patient, Vol. 8*, pp. 19-27.

39. IBM, SPSS (2023). *IBM SPSS Conjoint*. https://www.ibm.com/us-en/marketplace/spss-conjoint
40. Jones, N. (2020). *User Loyalty and Willingness to Pay for a Music Streaming Subscription Identifying Asset Specificity in the Case of Streaming Platforms*. Paper available from: https://api.semanticscholar.org/CorpusID:221354391
41. Kim, J., Nam, C., Ryu, M.H. (2017). What do consumers prefer for music streaming services?: A comparative study between Korea and US. *Telecommunications Policy*, *41(4),* pp. 263-272.
42. Lee, J.A., Soutar, G.N., Louviere, J.J. (2007). Measuring Values Using Best-Worst Scaling: The Lov Example. *Psychology & Marketing, 24(12),* pp. 1043-1058.
43. Lopes., M., Coelho P.S. (2021). Music streaming services: understanding the drivers of customer purchase and intention to recommend. *Heliyon*, *7(8).*
44. Louviere, J.J., Flynn, T.N. (2010). Using best-worst scaling choice experiments to measure public perceptions and preferences for healthcare reform in Australia. *The Patient: Patient-Centered Outcomes Research*, *3(4),* pp. 275-283.
45. Louviere, J.J. (1988). *Analyzing Decision Making: Metric Conjoint Analysis*. Newbury Park, CA: SAGE Publications.
46. Louviere, J.J. (1991). Best-Worst Scaling: A Model for the Largest Difference Judgments. *Working Paper.* University of Alberta.
47. Louviere, J.J., Flynn, T.N., Marley, A.A.J. (2015). *Best-Worst Scaling: Theory, Methods and Applications*. Cambridge: Cambridge University Press.
48. Louviere, J.J., Hensher, D.A., Swait, J.D. (2000). *Stated Choice Methods: Analysis and Application*. Cambridge: Cambridge University Press.
49. Lu, B., Zhang, S. (2020). A conjoint approach to understanding online buyers' decisions towards online marketplaces. *Journal of Theoretical and Applied Electronic Commerce Research*, *15(3),* pp. 69-83.
50. Luce, R.D., Tukey, J.W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, *1(1),* pp. 1-27.
51. Lusk, J.L., Briggeman, B.C. (2009). Food values. *American Journal of Agricultural Economics*, *91(1),* pp. 184-196.
52. Maftei V.A., Gerogiannis V.C., Papageorgiou E.I. (2016). Critical success factors of online music streaming services – a case study of applying the fuzzy cognitive maps method. *International Journal of Technology Marketing*, *11(3),* pp. 276-300.
53. Marley, A.J. (2010). The best-worst method for the study of preferences. In: P.A. Frensch, R. Schwarzer (eds.), *Cognition and Neuropsychology: International Perspectives on Psychological Science* (pp. 147-157).
54. Marley, A.J., Louviere, J.J. (2005). Some probabilistic models of best, worst, and best-worst choices. *Journal of Mathematical Psychology*, *49(6),* pp. 464-480.

55. Marti, J. (2012). A best worst scaling survey of adolescents' level of concern for health and non-health consequences of smoking. *Social Science & Medicine*, *75(1),* pp. 87-97.
56. Menictas, C., Wang, P.Z., Louviere J.J. (2012). Assessing the validity of brand equity constructs. *Australasian Marketing Journal*, *20(1),* pp. 3-8.
57. Mueller, S., Francis, I.L., Lockshin, L. (2009). Comparison of Best-Worst and Hedonic Scaling for the Measurement of Consumer Wine Preferences. *Australian Journal of Grape and Wine Research, 15(3),* pp. 205-215.
58. Poortinga, W., Steg, L., Vlek, C., Wiersma, G. (2003). Household preferences for energy-saving measures: A conjoint analysis. *Journal of Economic Psychology*, *24(1),* pp. 49-64.
59. Potoglou, D., Burge, P., Flynn, T., Netten, A., Malley, J., Forder, J.E. (2011). Best-worst scaling vs. discrete choice experiments: An empirical comparison using social care data. *Social Science & Medicine*, *72(10),* pp. 1717-1727.
60. R Development Core Team (2023). *R: A language and environment for statistical computing*. https://cran.r-project.org/. Vienna, Austria. https://www.R-project.org
61. Rao, V.R. (2014). *Applied Conjoint Analysis*. Berlin/Heidelberg: Springer-Verlag.
62. Robinson, D., Hayes, A., Couch, S. (2024). *broom: Convert Statistical Objects into Tidy Tibbles*, https://CRAN.R-project.org/package=broom
63. Sailer, S. (2022). *crossdes: Construction of Crossover Designs*, https://CRAN.R-project.org/package=crossdes
64. Severin, F., Schmidtke, J., Mühlbacher, A., Rogowski, W.H. (2013). Eliciting preferences for priority setting in genetic testing: A pilot study comparing best-worst scaling and discrete-choice experiments. *European Journal of Human Genetics*, *21(11),* pp. 1202-1208.
65. Shin, D., Kim, S. (2014). Consumption attribute value estimation of digital music contents service by conjoint analysis. *Journal of Korea Contents Association*, *14(12),* pp. 924-934.
66. Shin, S., Kim, S. (2025). Niche analysis of competition among music streaming services in Korea. *Telecommunications Policy*, *49(2).*
67. Therneau, T.M., Lumley, T., Atkinson, E., Crowson, C. (2024). *survival: Survival Analysis,* https://CRAN.R-project.org/package=survival
68. Thomson, D.M., Crocker, C., Marketo, C.G. (2010). Linking sensory characeristics to emotions: An example using dark chocolate. *Food Quality and Preference*, *21(8),* pp. 1117-1125.
69. van Dijk, J.D., Groothuis-Oudshoorn, C.G., Marshall, D.A., IJzerman, M.J. (2016). An empirical comparison of discrete choice experiment and best-worst scaling to estimate stakeholders' risk tolerance for hip replacement surgery. *Value in Health*, *19(4),* pp. 316-322.
70. Walesiak, M. (1996). *Metody analizy danych marketingowych*. Warszawa: PWN.
71. Walesiak, M., Bąk, A. (2000). *Conjoint analysis w badaniach marketingowych.* Wrocław: Wydawnictwo Akademii Ekonomicznej.

72. Wang, T., Wong, B., Huang, A., Khatri, P., Forgie, M., Lanphear, J.H. O'Neill, P.J. (2011). Factors affecting residency rank-listing: A Maxdiff survey of graduating Canadian medical students. *BMC Medical Education*, *11(61).*
73. Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* New York: Springer-Verlag.
74. Wickham, H., Winston, C., Henry, L., Pedersen, T.L., Takahasi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., van der Brand, T. (2024). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. https://cran.r-project.org/web/packages/ggplot2.
75. Wilkinson, L. (1998). *Conjoint analysis*. Chicago: SYSTAT.
76. Xie, F., Pullenayegum, E., Gaebel, K., Oppe, M., Krabbe, P.F.M. (2013). Eliciting preferences to the EQ-5D-5 L health states: Discrete choice experiment or multi-profile case of best–worst scaling? *The European Journal of Health Economics*, *15(3),* pp. 281-288.
77. Zwerina, K. (1997). *Discrete choice experiments in marketing*. Heidelberg/New York: Physica-Verlag.