Practical Assessment, Research, and Evaluation

Volume 27 Article 7

2022

An introduction to the analysis of ranked response data

Holmes Finch
Ball State University

Recommended Citation

Finch, Holmes (2022) "An introduction to the analysis of ranked response data," *Practical Assessment, Research, and Evaluation*: Vol. 27, Article 7.
DOI: https://doi.org/10.7275/tgkh-qk47
Available at: https://scholarworks.umass.edu/pare/vol27/iss1/7

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 27 Number 7, April 2022 ISSN 1531-7714

An Introduction to the Analysis of Ranked Response Data

Holmes Finch
Ball State University

Researchers in many disciplines work with ranking data. This data type is unique in that it is often deterministic in nature (the ranks of items *k*-1 determine the rank of item *k*), and the difference in a pair of rank scores separated by *k* units is equivalent regardless of the actual values of the two ranks in the pair. Given its unique qualities, there are specific statistical analyses and models designed for use with ranking data. The purpose of this manuscript is to demonstrate a strategy for analyzing ranking data from sample description through the modeling of relative ranks and inference regarding differences in ranking patterns between groups. An example dataset of university faculty ratings of job characteristics was used to demonstrate these various methods, and the ways in which they can be tied together to obtain a comprehensive understanding of a ranking dataset. The analyses were carried out using libraries from the R software package, and the code for this purpose is included in the appendix to the manuscript.

Introduction

Ranking data arises from situations in which a finite number of entities, such as sports teams, product brands, political candidates, television programs, or job qualities, are ranked relative to one another. There are many examples of ranking data in an array of academic disciplines, including education (Acuna-Soto, Liern, & Perez-Gladish, 2021) psychology (Regenwetter, et al., 2007), health care (Hackert, et al., 2019; Bothung, et al., 2015; Craig, et al., 2009), quality of life (Peiro-Palomino & Picazo-Tadeo, 2017), sociology (Harakawa, 2021), market research (Kamishima & Akaho, 2006), and political science (Moors & Vermunt, 2007; Gormley & Murphy, 2008). The breadth of these examples demonstrates the great utility of rankings as a tool for understanding human behavior and other scientific phenomena. Throughout this manuscript, the entities being ranked will be referred to as items.

The mechanism for rankings can come in the form of a sample of raters, television viewers, voters, or professional sports competitions. Whichever mechanism is used to rank the items, this type of data share some common qualities. By its very nature, ranking data has a deterministic quality that is not found in most other data situations. Determinism in this context refers to the fact that given the first *k*-1 of *k* rankings, the k^{th} item can only take a specific value. For example, if we know that among a set of 4 tennis players, Novak Djokovic is ranked first, Rafael Nadal second, and Roger Federer third, Andy Murray must be ranked fourth. It should be noted that this deterministic quality is not present if ties are allowed. In that case, it is possible for two or more of the items to have the same rank, and thus the ranking pattern of items *k*-1 does not dictate the ranking of item *k*. In addition to the deterministic nature of the scores, a second signal feature of ranking data is that typically the difference in scores between any pair of items with adjacent rankings is equivalent, regardless of the actual values. For example, the difference between rankings 4 and 5 is equal to the difference between rankings 1 and 2. A third unique quality of ranking data is with respect to their correspondence with the set of permutations of the data. Specifically, common analytic approaches such as histograms or analysis of variance (ANOVA) are not appropriate for use with ranking data because the set of all possible permutations from which the ranks are drawn do not have a natural linear ordering (Fischer, et al., 2019; Alvo & Yu, 2014). Therefore these commonly used techniques will not yield meaningful results and alternative methods, such as those described in this paper, are needed. As described above, all items are ranked by all raters. However, this design is not always used, and in some cases raters are asked to rank only a subset of the *k* items. For example, individuals may be asked to rank their three top candidates for office from a set of 10 in an election. This data structure presents the researcher with unique data analysis challenges, and though interesting, will not be addressed in this manuscript.

Study purpose

The purpose of the current work is to describe and to demonstrate a strategy for analyzing a set of ranking data, from the initial description of the sample through inferential models for characterizing the ranking patterns and investigating relationships between one or more covariates and these patterns. The goal in this demonstration is to provide researchers with a complete example for how to consider ranking data from an analytic perspective, and how to synthesize the results from these multiple techniques in order to gain a full picture of the ranked phenomena being studied. The data analyses include a description of the rankings, as well as model based explorations of the rankings, and investigations of relationships between the rankings and substantively relevant covariates. The example analyses were conducted using the R software package, with an eye to providing the reader with the tools necessary to successfully investigate their own ranking data. Therefore, the R code for conducting these analyses appears in the appendix and the example data are available as supplementary materials to the manuscript.

Sample description

A first step in most data analyses involves an exploration of the sample using descriptive statistics. This is certainly true of ranking data for which we are interested in the mean ranks of the items, as well as the pairwise comparisons of the items and the distribution of ranks for each of the items. The mean rank for item *i* (m_i) is defined as

$$m_i = \frac{\sum_{j=1}^{t!} n_j v_j(i)}{n} \quad (1)$$

Where

v_j =All possible rankings of the *t* objects

$xxv_j(i)$ =Rank score given to object *i* in ranking *j*

n_j =Observed frequency of ranking *j*

$n = \sum_{j=1}^{t} n_j$

A lower value for m_i indicates that the item is more favored by the members of the sample; i.e., has received a higher ranking with 1 being most favorable. For example, if item 1 has a mean rank of 2.4 and item 2 has a mean rank of 3.9, we would conclude that item 1 was typically ranked higher than item 2.

Another useful description of the sample is the frequency of pairwise comparisons of the item rankings. In other words, how frequently was item A preferred over item B? Table 1 includes a pairwise matrix for a simple example of 3 items that were ranked by 10 individuals. In this example, we can see that item 1 was ranked above item 2 five times, and above item 3 three times. In contrast, item 2 was ranked above item 1 8 times, and above item 3 10 times. Another way in which the rankings can be described is based on the marginal frequency of each rank for each of the items. These results can be presented in a marginal ranking matrix, as in Table 2. For this hypothetical example, item 2 most frequently received a top ranking, followed by item 1, and then item 3. Item 3 was most frequently the lowest ranked.

In addition to describing the sample in terms of central tendency and relative ranking, we may also want to ascertain whether the pattern of rankings is random in nature. One way to do that is to test the null hypothesis that the mean rank is equal to $\frac{(t+1)}{2}$ for *t* ranked items. For the three ranked items, the mean under the null hypothesis of a random ranking would be $\frac{(3+1)}{2} = 2$. In other words, if the rankings provided by the members of the sample had no systematic pattern (i.e., were random in nature), then the mean

Table 1. Example pairwise ranking matrix for three ranked items

	Item 1	Item 2	Item 3
Item 1	0	5	3
Item 2	8	0	10
Item 3	2	2	0

Table 2. Example marginal ranking matrix for three ranked items

	Item 1	Item 2	Item 3
Item 1	2	5	3
Item 2	8	1	1
Item 3	0	2	8

ranking would be 2. The test statistic for this null hypothesis is

$$Q = \frac{12n}{t(t+1}\sum_{j=1}^{t}\left(m_i - \frac{t+1}{2}\right)^2 \qquad (2)$$

Where

m_i =Mean rank for item i

$\frac{t+1}{2}$ =Mean under null hypothesis of a random ranking

Q is distributed as a Chi-square statistic with degrees of freedom of t-1. If the p-value associated with Q and t-1 degrees of freedom is ≤ 0.05, we would reject the null hypothesis that all items have the same mean rank. In rejecting the null hypothesis, we would conclude that the rankings provided by members of the sample were not random in nature.

Together, these descriptive statistical methods provide the researcher with important information about the general patterns of the rankings produced by members of the sample. The mean ranking for each item reflect how popular/positive (or not) the members of the sample found each of them. An understanding of the popularity of individual items can be deepened through a consideration of their relative and marginal ranks. In other words, how likely was one item likely to be preferred to each of the others, and how frequently was each item given each ranked value? Answering these questions provides the researcher with a deeper understanding of the relative rank of each item than does the mean ranking alone. Finally, the statistical test of the null hypothesis of randomness yields information regarding the systematic nature of the ranking process used by the sample. If the null hypothesis of this test is rejected, we would conclude that in the population, the items are ranked in a systematic fashion; i.e., some are given higher ranks on average than are others.

Multidimensional scaling

Another aspect of the ranking process that may be of interest to researchers are patterns of relationships among item rankings. One approach for investigating these patterns involves the application of unfolding multidimensional scaling (UMDS), which has been recommended for use with ranking data (Alvo & Yu, 2014; Armstrong, et al., 2014). MDS with ranked data has been used in a variety of research areas including health sciences (e.g., Krabbe, Salomon, & Murray, 2007), marketing (e.g., Adlakha & Sharma, 2020; Jeong & Kwon, 2016), and psychology (Askell-Williams & Lawson, 2002) among others.

The goal of this analysis is to reduce the dimensionality present in a set of data (e.g. j raters by i rated items) to two or three dimensions, and then to examine the proximity of objects of interest (e.g., raters to ranked items, ranked items to one another) in this low dimensional space. The goal when reducing the dimensionality of the data in this fashion is to retain the essential underlying relationships among the raters and items while also simplifying it so that these relationships are easier to discern.

The data upon which UMDS operates is in the form of an item by rater rectangular distance matrix, where distances express proximity of a rater to an item. A pair that is ranked more closely together will be associated with a smaller distance value. A commonly used measure of distance/dissimilarity is Euclidean distance:

$$d_{ij} = \sqrt{\left(\boldsymbol{y_i} - \boldsymbol{x_j}\right)'\left(\boldsymbol{y_i} - \boldsymbol{x_j}\right)} \qquad (3)$$

Where

$\boldsymbol{y_i}$ =Set of rankings given by judges to item *i*

$\boldsymbol{x_j}$ =Set of item rankings given by judge *j*

The value of d_{ij} represents the relative preference of item *i* for judge *j*. Therefore, smaller values reflect that judge *j* prefers item *i*. It should be noted that there are a number of other distance measures that can be used with UMDS including the Hamming distance (Hamming, 1950), the Kendall distance (Kidwell, et al., 2008; Alvo & Cabilio, 1995), and the Cayley distance (Fligner & Verducci, 1986), among others.

UMDS works by finding a small number (e.g., 2) of dimensions that contains most of the information available in the raw ranking data. The optimal solution minimizes the Stress function, which is expressed as:

$$Stress = \sum_{i<j}\left(\delta_{ij} - d_{ij}\right)^2 \qquad (4)$$

Where

δ_{ij} =Estimated distance between item *i* and rater *j* based on UMDS

d_{ij} =Observed distance between item *i* and rater *j*

The resulting model yields coefficients for each rater and each item for each of the dimensions. Therefore, if a 2-dimensional solution is used, UMDS finds coefficients for each rater and item for each of the dimensions that yields values of δ_{ij} that minimize the Stress value in equation (4). The performance of the model is represented as the percent of the variance in the observed data that is accounted for by the MDS model.

As an example to demonstrate UMDS consider Table 3, which displays the ranks provided by three of the raters and the corresponding estimates of d_{ij} based on a 2-dimensional UMDS. Rater 1 provided the following rank ordering of the job qualities: Contract, Salary, Chair Support, Travel budget, Health care, Workload. The estimated distances for this rater from each of these qualities was: 0.34, 0.79, 0.88, 1.24, 1.04, 1.60. These results confirm that, with the exception of Travel budget, the model estimated distances conform to the rank ordering provided by the rater. More specifically, we see that based on the distance estimates Rater 1 valued Contract most strongly, followed by Salary, and then Chair support. They were least concerned with Workload, as reflected both in the observed rank and the UMDS estimated distance.

Table 4 includes the UMDS coefficients for each of the first three raters as well as for the on the rated items, with respect to each of the dimensions. From these results, we can see that raters 1 and 3 are relatively far apart on the first dimension; i.e., their coefficients are further from one another than either is from that of rater 2. An examination of the rankings illuminates this spread, in that the rank ordering of the job qualities for Raters 1 and 3 were quite different, with the exception that they both valued Salary relatively highly. The coefficients for the items also show that Salary and Health care were most closely associated with one another, as were Chair support and Travel budget. Contract and Workload had coefficients that differed from the other items and from one another. Most often, the results of a MDS analysis are expressed in the form of a graph displaying the location of the items and raters in 2-dimensional space, as demonstrated in the results below.

When applied to ranking data, UMDS provides the researcher with insights into the relative popularity of the ranked items to one another. By viewing the relative locations of the items to one another on a 2-dimensional scatterplot of the dimension weights, it is possible to see which items tended to be ranked similarly by members of the sample. For example, if two items appear close together in the plot, we can conclude that members of the sample tended to rank them close together (e.g., 1, 2 or 4,5). In addition, by

Table 3. Rankings and UMDS estimated d_{ij} values for the first three raters: 2-dimensional solution

Rater*	Contract+	Salary	Health care	Workload	Chair support	Travel budget
Rankings						
1	1	2	5	6	3	4
2	3	1	2	4	5	6
3	6	1	2	3	5	4
Estimated Distance						
1	0.34	0.79	1.04	1.60	0.88	1.24
2	1.01	0.21	0.53	1.03	0.99	1.32
3	1.42	0.30	0.35	0.91	1.35	1.40

*Rater refers to the individual providing the rankings. Here they are numbered 1, 2, and 3

+Columns include the items being ranked from 1 to 6 by each rater.

Table 4. UMDS coefficients for items and the first three raters: 2-dimensional solution

Rater or item	Dimension 1	Dimension 2
1	-0.53	0.05
2	0.15	0.04
3	0.54	-0.12
Contract	-0.86	0.15
Salary	0.24	-0.14
Health care	0.39	-0.44
Workload	0.94	0.69
Chair support	-0.33	0.91
Travel budget	-0.38	-1.17

plotting the locations of the items on the same graph along with the rater locations, we can gain insights into the relative popularity of the individual items. Those items that are more centered among the persons are more popular than are those items that appear at the periphery of the cloud of participants. Finally, an examination of the participant locations can reveal the extent to which there may be different subgroups among the raters. For example, if the plot reveals two distinct groupings of individuals based on their MDS weights, we can conclude that there are two separate ways in which the individuals ranked the items. Conversely, if the plot reveals a single group of points based on the MDS weights, we would conclude that most of the participants ranked the items in a similar fashion. Finally, the performance of the UMDS is typically evaluated in terms of the proportion of variability in the observed rankings that the model accounts for. A higher proportion of explained variance indicates that the model better accounts for the variability in the observed rankings. These issues will be revisited in the extended example below.

Plackett-Luce model

One of the more useful modeling tools available specifically for use with ranking data is the Plackett-Luce model (PLM; Plackett, 1975). This modeling approach has been used in a variety of settings involving ranked data, such as in medicine (Mollica & Tardella, 2014), management science (Farias, Jagabuthla, & Shah, 2013), wine tasting (Bodington & Malfeito-Ferreira, 2017), school psychology (Bargogliotti, et al., 2021), and ecological research (Lohr, Cox, & Lepczyk, 2012). There have also been a number of developments and extensions to the PLM, including a mixture model version (Collins & Tardella, 2017), a Bayesian estimator (Mollica & Tardella, 2020), a mixed effects PLM (Bockenholt, 2001), a robust estimator for crowd sourced preference data (Han, Pan, & Tsang, 2018), and a nonparametric PLM (Caron, Teh, & Murphy, 2014). In addition, there are a number of in-depth treatments focused on the application of the PLM, to ranked data (e.g., Turner, et al., 2020; Yu, Gu, & Xu, 2019; Turner, van Etten, Firth, & Kosmidis, 2018; Alvo & Yu, 2014; Glickman & Hennessey, 2015).

The PLM is designed to model the probability of a specific rank ordering for a set of *I* items and is based on Luce's axiom (Luce, 1959), which states that for a set of items, *S*, the probability of selecting item *i* from the set is given by:

$$P(i|S) = \frac{\alpha_i}{\sum_{i \in S} \alpha_i} \quad (5)$$

Where

α_i =Worth of item *i*

Based on this axiom, we can view the rank ordering of the *I* items as a sequence of choices from the items remaining in set *S*. In other words, when an individual ranks a set of options from most to least favored, they are selecting the most favored item from the set of *S* items that have yet to be ranked. Once the first ranking is made, the individual selects the next item to rank from the remaining *S*-1 item, and so on. The probability of ranking the items in a particular order (π) can then be expressed in the PLM as:

$$P(\pi) = \prod_{i=1}^{I} \frac{\alpha_i}{\sum_{i \in A_i} \alpha_i} \quad (6)$$

Where

A_i =Set of alternatives from which item *i* is chosen.

The model parameters can be estimated using either maximum likelihood or Bayesian methods. For the example described below, maximum likelihood was used.

The key parameter in the PLM is item worth, which reflects the importance of the item and corresponds to the ranks provided by the subjects. Higher values of the worth reflects greater importance of the item as reflected in the rankings. In other words, items that are given a higher rank will also have a higher worth value. In order for the model to be identified (i.e., for the item parameters to be estimable), the worth parameter for one of the items is typically set to 0 so that the other item worth values reflect the relative importance of each non-reference item vis-à-vis the reference. An alternative method for identifying the model is to set the mean of the worth parameters as the reference, in which case the individual item worth values reflect the importance of the item to the average item worth. Finally, we can also view the relative importance of the items in terms of the probability of its being selected as the most important. This conversion can be done using the axiom expressed in equation (5). The PLM also produces standard errors for the item parameters. These standard errors can be used to construct a test statistic for testing the null hypothesis that $\alpha_i = 0$. When a single item serves as the reference, this statistic tests whether a given item has equal worth to the reference, whereas when the item worth mean is the reference the test would assess whether a specific item's worth differs from the mean worth across items.

Fit of the PLM to the data is reflected in the residual deviance statistic, which is part of the standard output of the model. The residual deviance reflects the difference between the rankings predicted by the PLM and those actually observed in the data. When this difference is small, the model is said to fit the data well, whereas large values of the deviance indicate poor model-data fit. Determining whether the deviance is large can be done by comparing it to the degrees of freedom either using the Chi-square statistic or the ratio of the deviance to the residual degrees of freedom (Agresti, 2013). When the model fits well, the ratio of the deviance to degrees of freedom is approximately 1. In addition, the deviance is distributed approximately as a Chi-square statistic with residual degrees of freedom, and can therefore be used for statistical hypothesis testing (Turner, et al., 2021a; Agresti). The null hypothesis of this test is that the model fits the data well. Therefore, a non-significant Chi-square test result would mean that the model provides good fit to the observed data. It should be noted that the users-manual for the `PlackettLuce` R package indicates that the residual deviance can be used for inference regarding model-data fit (Turner, et al.). However, it is also true that there is not a great deal of literature investigating the distribution of this statistic in the context of the PLM. Therefore, it is recommended that interpretation of the deviance in this context be undertaken with some care, and the reader continue to read new research regarding the use of the deviance statistic and other methods for assessing model fit.

Plackett-Luce model with covariates

One of the primary advantages to researcher using the PLM is that it can be extended to investigate relationships between the rankings and other variables associated either with the item or the rater. This PLMC model is particularly useful when the researcher is interested in ascertaining the extent to which specific qualities of the raters (or of the items) is related to the item worth parameters. For example, it may be of interest to know the extent to which an employee's years of experience is related to how they rank the importance of various aspects of their job. The PLMC allows the researcher to include years of experience as a covariate for the set of ranks, and provides an estimate of this relationship in the form of a coefficient very similar to what is obtained using linear regression. Mathematically, the PLMC is written as

$$log(\alpha_i) = \beta_0 + \beta_1 x_{j1} + \cdots + \beta_p x_{jp} \qquad (7)$$

Where

β_0 =Intercept that is fixed by the constraint $\sum_i \alpha_i = 1$.

β_p =Coefficient for covariate *p*

x_{jp} =Value of covariate *p* for rater *j*

If the hypothesis test for β_p is statistically significant, we would conclude that there is a relationship between the covariate and the worth of item *i*. The sign of β_p reflects the nature of this relationship (positive or negative). Referring back to the years of experience example, a statistically significant positive coefficient with respect to the item number of vacation days would indicate that employees with more experience are more likely to rate the number of vacation days provided by their employer as being a more important aspect of their job.

The fit of the PLMC can be compared with that of the PLM using the difference between the deviance statistics for the two models, which is distributed as a Chi-square statistic with degrees of freedom equal to the difference in degrees of freedom for the two models. The null hypothesis of this test is that the two models fit the data equally well. Therefore, a statistically significant Chi-square difference test would indicate that the models provided different degrees of model fit. In addition to the Chi-square, fit of the models can be compared using the Akaike Information Criterion (AIC), which penalizes the deviance for model complexity (i.e., the number of parameters estimated by the two models). The model with the smaller AIC is considered to be the better fitting, once model complexity is taken into consideration.

Plackett-Luce tree (PLT)

An alternative approach for investigating relationships between rater covariates and the PLM worth parameters is with a Plackett-Luce tree (PLT). Like the PLMC, the PLT is designed to assess whether specific rater traits are associated with rater characteristics. However, rather than expressing this relationship in the form of a linear model, as in equation (7), the PLT is based upon a recursive partitioning algorithm (Strobl, Wickelmaier, & Zeileis, 2011). This algorithm can be used to automatically identify interactions among the covariates with respect to item worth values. The PLT algorithm works using the following steps:

1. Fit the PLM to the full sample.
2. Assess the stability (statistically significant differences) of worth parameters across differing values of each possible split for each covariate.
3. If instability is found, split the sample using the covariate with the largest statistically significant difference at the cut-point where the model fit improves the most.
4. Repeat steps 1-3 within each split until no statistically significant instability (differences in

the covariates) is found for the worth parameter.

The results of the PLT comes in the form of a graph displaying the tree resulting from the algorithm described above.

Taken together, the PLM, PLMC, and PLT provide the researcher with information regarding not only the relative importance of each item, but also how rankings of these items might (or might not) be related to traits associated with members of the sample. In addition, the specific item parameters and their associated standard errors can be used to develop statistical tests comparing the relative ranking of items versus one another. Thus, whereas descriptive statistics such as the means and relative frequencies of ranks provide a general description of the sample rankings, the results of Plackett-Luce family of models give researchers insights into individual factors that are associated with ranking behavior, and whether two items are likely to have different rankings in the population. If the comparison of the worth parameters for two items are found to differ statistically, we would conclude that indeed respondents viewed them as having different levels of importance. This comparison, along with the model examining relationships between covariates and rankings are demonstrated in the example below.

Methods

As stated above, the goal of this manuscript is to provide the reader with a full example of how ranked data can be modeled using libraries in the R software package. Therefore, a real world example dataset was used, with the R code included in the appendix and the data available at https://scholarworks.umass.edu/pare/vol27/iss1/7/. In this section, the sample, measures, and data analytic techniques are described. The results of these analyses are then presented in the next section of the manuscript.

Study participants

The participants in this study were 41 non-tenure track higher education faculty members from whom data were collected using a survey administered by a researcher who was independent of the university. The faculty were employed in a variety of departments across a single university, with varying levels of experience and education. Participants completed informed consent forms and their data was completely anonymous.

Methods

Study participants were presented with a set of six job qualities and asked to rank them from most to least important. The items to be ranked were:

Contract length

Salary

Health care plan

Workload (number of classes taught)

Chair support

Travel budget

Respondents were asked to rank the data from most to least important (1=most important to 6=least important), with tied rankings not allowed. In addition, each respondent was asked the number of years that they had been teaching and their highest degree. The years of experience were collected as 1=0-5 years, 2=6-10 years, 3=11-15 years, 4=16-20 years, and 5=21+ years. The degree data were classified as 1=BA/BS, 2=MA/MS, 3=Specialist/Masters+, 4=PhD/EdD. Data were collected using the Qualtrics (Qualtrics, 2020) online platform.

Data analysis

A variety of analyses were used to explore the rankings of the job qualities listed above. These analyses correspond to those described in the previous section of the manuscript, and were conducted using libraries from the R software package (R Core Team, 2021). Descriptive statistics, including the sample mean ranks, as well as the pairwise and marginal ranks were used to provide insights into typical behavior of the respondents. These descriptive statistics were obtained using the `destat` function within the `pmr` R library (Lee & Yu, 2015). The null hypothesis of random ranking behavior was also tested using the Chi-square statistic, which was calculated using statistics obtained from the `destat` function. In addition, UMDS, using the `smacofRect` function from the `smacof` R library (Mair, de Leeuw, Groenen, & Borg, 2021), was employed in order to gain insights into relationships of the ranks among the six items, and the respondents. UMDS was fit to the data using the Euclidean distance, as well as the Kendall and Hamming distance measures. Results for all three approaches were quite similar, and only those for the Euclidean distance are reported below. The PLM was fit to the data using the `PlackettLuce` function from the R `PlackettLuce` library (Turner, Kosmidis, Firth, & van Eten, 2021b) in conjunction with the `prefmod` library (Hatzinger & Maier, 2017), with quasi-standard errors for the worth parameter estimates obtained using the `qvcalc` R library (Firth, 2020). The PLMC with both experience and highest degree serving as covariates was fit to the data using the `rol` function from the `pmr` R library. Finally, a PLT was used to investigate the possibility of interactions between highest degree and years of teaching experience in terms of the ranking behavior. This tree model was employed using the `pltree` library from the `PlackettLuce` R library.

Results

Sample description

The mean ranks for the six items appear in Table 5. Salary was the highest ranked job quality on average, followed by health care. The least favored (lowest sample means) items were travel budget and workload. Table 5 also includes the pairwise rank comparisons for the set of items. Recall that these values reflect the number of times that the row item was ranked higher than the column item. For example, Salary was ranked higher than contract by 30 of the 41 study participants. From these results, we can confirm that salary was the most popular (highest ranked) job quality, with pairwise comparison values ranging between 30 and 38 when compared to the other items; i.e., it received a higher rank than each of the other qualities from between 30 and 38 of the study participants. In contrast, travel budget was not ranked higher than any of the other items by a majority of the respondents. It performed best compared to workload, against which it was given a higher rank by 13 individuals.

The marginal frequencies, which appear at the bottom of Table 3, provide more evidence regarding the most and least popular items. Salary received the highest rank 24 times, and the second highest rank 7 times, and was never the lowest ranked item. Health care was the highest ranked item for 4 respondents, and the second highest for an additional 19 respondents. In contrast to these popular items, the travel budget was the least valued by study participants, with 31 of them ranking it either lowest or next to lowest. Workload yielded a bimodal distribution of ranks with 11 individuals placing it third, and 13 placing it fifth.

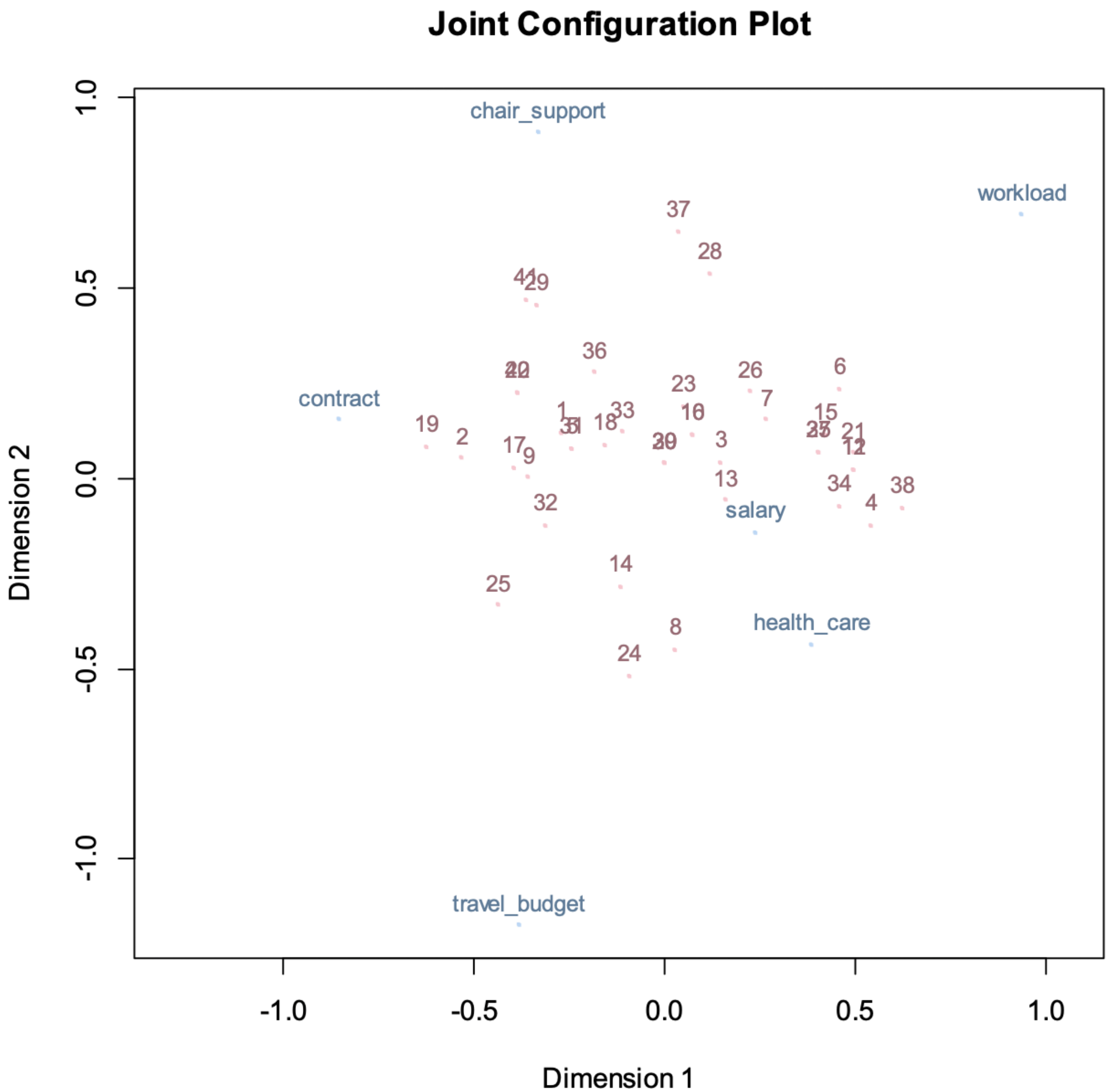
In order to assess whether the pattern of ranks departed from what we would expect were they completely random, the Chi-square test was used, as described above. The mean rank under the null hypothesis for this calculation was $\frac{t+1}{2}=\frac{6+1}{2}=3.5$. The Chi-square statistic for this problem was 78.99, with degrees of freedom of 5 (6-1), yielding a *p*-value less than 0.001. Thus, if □ =0.05, we would reject the null hypothesis and conclude that there was a nonrandom pattern to the ranks provided by the participants. In other words, we would conclude that in the population some of the job qualities are ranked as more important than are others.

UMDS

In order to gain insights into how the ranked items are related to one another, UMDS with 2 dimensions was fit to the data using the `smacofRect` function from the `smacof` R library. The plot was created using the `mdpref` function from the `pmr` R library. This model explained approximately 55% of the variance in the rankings. Figure 1 displays the locations of the 6 items and 41 respondents on dimensions 1 and 2. First, we note that salary is most central with respect to the study participants, which reflects that it was the highest ranked of the items by many individuals. In contrast, travel budget and workload lay furthest from the cloud of participant points, which is expected given that they were the lowest ranked items for most raters. The locations of health care, contract and chair support relative to the participants indicates their midlevel rankings as also shown in Table 5.

Based on the distribution of job categories in Figure 1, dimension 1 appears to reflect the contrast between workload and contract, such that those who ranked workload relatively more highly were also more likely to rank contract terms relatively lower. In addition, dimension 1 also reveals that ranks for salary and health care were closely related to one another; i.e., those who ranked salary highly also tended to rank health care highly. The second dimension displays the

Figure 1. Plot for ranked items and study participants for the 2-dimensional MDS solution



contrast between travel budget versus contract, chair support, and workload. In other words, within individual respondents, those who ranked travel budget relatively higher tended to rank the other three items somewhat lower. Once again, on the second dimension salary and health care were located in relatively close proximity to one another.

Plackett-Luce model

Two versions of the PLM were fit to the data, with the first treating the first item in the list (contract) as the reference and the second using the mean worth as the reference. Both models were fit using the `PlackettLuce` function from the `PlackettLuce` R library. Table 6 includes the worth estimates, standard errors, Z test statistics

Table 5. Descriptive statistics for ranks of six employment items

	Contract	Salary	Health care	Workload	Chair support	Travel budget
Mean	3.42	1.81	2.73	4.34	3.61	5.10
(SD) rank	(1.66)	(1.45)	(1.23)	(1.30)	(1.55)	(1.18)
Pairwise rank comparisons						
Contract	0	11	13	26	23	33
Salary	30	0	34	38	32	38
Health care	28	7	0	36	28	35
Workload	15	3	5	0	17	28
Chair support	18	9	13	24	0	34
Travel budget	8	3	6	13	7	0
Marginal rank frequencies						
	1	2	3	4	5	6
Contract	8	5	6	11	6	5
Salary	24	7	5	4	1	0
Health care	4	19	7	7	3	1
Workload	0	3	11	5	13	9
Chair support	5	5	9	9	8	5
Travel budget	0	2	3	5	10	21

Table 6. Plackett-Luce model parameter estimates for job characteristics data with reference item

Item	Worth	Standard error	*Z*	*p*
Contract*	0	NA	NA	NA
Salary	1.57	0.30	5.20	<0.001
Health care	0.77	0.28	2.68	0.007
Workload	-0.48	0.28	-1.73	0.08
Chair support	-0.09	0.27	-0.35	0.73
Travel budget	-1.17	0.30	-3.92	<0.001

*Contract is the reference item

(ratio of estimate to standard error), and the *p*-values associated with the test statistic. From these results, we see that salary and health care both had statistically significant positive worth values, meaning that they were higher ranked (more valued) than contract terms by the participants. Conversely, travel budget had a statistically significant negative coefficient, indicating that it was rated as less valuable than contract terms. The worth estimates for workload and chair support were not significantly different than that of contract.

If we are interested in comparing the worth values for items other than the reference, we can obtain the covariance matrix for the parameter estimates and use them with the standard errors to construct a test statistic. For example, we may wish to compare the worth of the two most popular items, salary and health care. The covariance between the parameter estimates is 0.05, a value which can be obtained using R. The test statistic for comparing the two coefficients is then calculated as:

$$z = \frac{\alpha_{salary} - \alpha_{health\ care}}{\sqrt{SE^2_{salary} + SE^2_{health\ care} - 2COV(\alpha_{salary}, \alpha_{health\ care})}} = \frac{1.57 - 0.77}{\sqrt{0.30^2 + 0.28^2 - 2(0.05)}} = 3.06$$

Because this value is greater than 2 (which is associated with a 2-tailed α of 0.05), we would reject the null hypothesis that the two items have equivalent worth values and conclude that the worth of salary was greater than that of health care. In other words, the raters values salary more than they did health care.

Table 7 includes the worth estimates, standard errors, and associated hypothesis test statistics and *p*-values for each item when the mean worth served as the reference. Recall that in this case, the worth estimates reflect the importance of an item relative to the mean ranking across the items. Thus, salary and health care were both ranked significantly higher than average, whereas workload and travel budget were ranked significantly lower than average by the participants. Contract and chair support had ranks that were statistically equivalent to the overall average. The estimated probabilities that each item received the highest rank appear in Table 8. We can see that salary clearly was most likely to be ranked first, followed by health care. Each of the other items had probabilities of being top ranked at or below 0.1.

As described above, we can evaluate the performance of the PLM using the residual deviance statistic, which for this model was 382.72 with 516 degrees of freedom. The ratio of the two was 0.74, which based on the commonly used rule of thumb (Agresti, 2013) would suggest good fit of the model to the data. If we assume that the deviance follows the Chi-square distribution, the *p*-value for the goodness of fit test was 0.99, also indicating that the PLM fits the data well. Again, use of the deviance statistic in this way has been suggested as possible based on the `PlackettLuce` users-manual (Turner, et al., 2021b). However, further work in this regard would seem to be warranted, given that there has not been a great deal of empirical evaluation as to its performance as a measure of fit..

Plackett-Luce model with covariates and Plackett-Luce trees

In this example, it is of interest to assess whether there are relationships between specific characteristics of the participants, level of education and years of experience, and the worth of each item. This analysis can be conducted using the PLMC, which is described above and is fit to the data using the `PlackettLuce` function. The models for each covariate were fit using R and the resulting model coefficients and standard errors appear in Table 9. Statistical significance for each coefficient can be inferred using the ratio of the model parameter estimate and its associated standard error. The null hypothesis being tested by this statistic

Table 7. Plackett-Luce model parameter estimates for job characteristics data with mean as the reference

Item	Worth	Standard error	*Z*	*p*
Contract	-0.10	0.18	-0.53	0.60
Salary	1.47	0.20	7.24	<0.001
Health care	0.65	0.18	3.65	<0.001
Workload	-0.58	0.18	-3.18	0.002
Chair support	-0.19	0.18	-1.06	0.29
Travel budget	-1.27	0.21	-6.08	<0.001

Table 8. Probabilities that job quality items received the highest rank

Item	Probability of being highest ranked
Contract	0.10
Salary	0.49
Health care	0.22
Workload	0.06
Chair support	0.09
Travel budget	0.03

Table 9. Coefficients and standard errors for experience and degree with item worth

Item	Experience coefficient	Experience standard error	Degree coefficient	Degree standard error
Contract	1.54	1.18	0.90*	0.45
Salary	-0.14	0.29	-0.04	0.60
Health care	0.40	0.32	-0.09	0.68
Workload	0.31	0.30	-0.52	0.64
Chair support	-0.04	0.30	-0.01	0.59
Travel budget	-0.12	0.31	0.33	0.60

is that there is not a relationship between the covariate and the item worth, with values greater than 2 leading to rejection of the null. Based on the results in Table 9, the relationship between degree and contract length was positive and statistically significant. Therefore, we conclude that participants with more advanced degrees tended to give contract length higher ranks. None of the other coefficients were statistically significant.

The fit of the PLM and PLMC can be compared in order to determine whether inclusion of the model covariates provides better fit to the data than does excluding them. As noted in the descriptions of the PLM and PLMC, this comparison can be made using a Chi-square difference test and the AIC statistic. The Chi-square test statistic for this example is calculated using deviance values and degrees of freedom obtained from the R output as:

$$\chi^2_\Delta = \chi^2_{\text{PLM}} - \chi^2_{\text{PLMC}} = 447.69 - 441.43 = 6.26$$

The degrees of freedom for χ^2_Δ is calculated as:

$$df_\Delta = df_{\text{PLM}} - df_{\text{PLMC}} = 610 - 600 = 10.$$

We can then use R to obtain the *p*-value for χ^2_Δ with df_Δ.

```
pchisq(q=6.26,df=10,lower.tail=FALSE)
```

The resulting p-value is 0.79, which is larger than our α of 0.05 meaning that we do not reject the null hypothesis that the models yield the same fit to the data. In other words, including the experience variable as a predictor of the ranks does not improve the fit of the model to the data. In addition, as noted above, we can compare the fit of the models using the AIC where the model with the smaller value is taken to yield the best fit after accounting for model complexity. The AIC for the PLM without covariates was 457.69, whereas the AIC for the PLMC with experience as the covariate was 461.43. Thus, we would conclude that after we account for model complexity, the PLM with no covariates yielded better fit to the data.

We can go through the same steps for the PLMC with highest degree as the covariate. The χ^2_Δ and df_Δ are calculated below.

$$\chi^2_\Delta = 447.69 - 426.93 = 20.56$$

$$df_\Delta = 610 - 600 = 10.$$

The *p*-value for this test statistic is 0.02, based on the following R command.

```
pchisq(q=20.56,df=10,lower.tail=FALSE)
```

The AIC for the model with grad degree was 446.93, which was smaller than the AIC for the model without covariates (457.69). Taken together, these results indicate that the PLMC with highest degree yielded better fit to the data than did the PLM with no covariates. This finding confirms the statistical significance of the relationship between degree and contract length, which was described above. Finally, in order to further investigate the relationships between participant covariates and item worth, a PLT was fit to the data using the `pltree` function from the `PlackettLuce` R library. As described above, the PLT is particularly effective for exploring interactions of the covariates with regard to the item worth parameters. For this example, the PLT model did not find any statistically significant splits with regard to either of the covariates. Therefore, the resulting tree was simply a single node including all of the participants. The worth estimates yielded by the tree were very close to those provided by the PLM as displayed in Table 4.

Synthesis of results

Now that the results from the various analyses have been described, it is important to synthesize them in order to obtain a more complete picture of the rankings considered in this study. Based upon both the raw sample means, the centrality of its position in the UMDS plot, and the PLM worth estimates, it is clear that respondents valued the salary paid by their employer most highly, followed by the health care insurance coverage that they received. They ranked the travel budget as being least important. In addition, the hypothesis tests associated with the PLM revealed that salary was the single most important job quality of those included in this study. In sum, respondents valued salary as the most important job quality, followed by health care coverage, and they valued travel budget least among the traits that they ranked.

The results of the UMDS revealed that respondents who valued salary highly also tended to value health care highly as well. In other words, the two job qualities that were most highly ranked individually were also ranked highly by the same participants. In addition, the UMDS results revealed that rankings of contract terms, chair support, and workload were loosely associated with one another such that higher ranks for one were associated with higher ranks for the others. In contrast, individuals who ranked travel budget more highly tended to give lower ranks to contract terms, chair support, and workload. With respect to qualities of the respondents themselves, the results presented above showed that individuals with a higher terminal degree were more likely to give higher ranks to the terms of the contract. Otherwise, none of the demographic information associated with the respondents was related to their ranking behavior.

Taken together, we can see that the respondents tended to value salary and health care coverage the most, that rankings on these two job qualities were positively correlated with one another, and that between the two salary was significantly more important to the respondents than was health care. In addition, these were the only two job qualities that were likely to be ranked first by most respondents. The rankings of other aspects of the job, including contract terms, chair support, and workload were positively associated with one another, though not as strongly as was the case for salary and health care. Contract term rankings were also positively related to level of the terminal degree of the study respondent. The scores given to travel budget were not related to rankings given to any other job quality, and indeed the travel budget was viewed as the least important from among those included in this study.

Conclusion

The goal of this manuscript was to describe a strategy for analyzing ranking data, and to demonstrate the utilization of that strategy with an existing dataset. Ranking data presents special challenges to researchers, not least because the scores provided by members of the sample are partially deterministic. In other words, when an individual is asked to rank a set of 6 items from most to least favorable, the rankings of the first

5 items will by necessity determine the rank of the 6th item. In addition, ranking data is generally of interest *en toto*, rather than each ranked item being an independent entity. The fact that the permutations from which the rankings emerge do not have a natural linear ordering also makes use of standard statistical methods less than optimal. As we saw in the extended example presented above, the primary research interest was in how the full set of items was ranked, as opposed to the ranking for a single item. Furthermore, when covariates were included in the analysis, we were interested in how they were related to full pattern of rankings rather than the rank given to a single item in the set. For these reasons, specific methods for dealing with ranking data are necessary, with standard models and approaches being too limited when it comes to understanding the full pattern of ranked scores.

Although the models for ranking data may be unique, the overall strategy for examining ranks is relatively similar to that used with other types of data. xFor example, we will generally want to begin our analysis with an examination of descriptive statistics. In this context, description of the sample involves calculating the mean and standard deviation of the rank for each item. In addition, it is important to present both pairwise and marginal rank frequencies as a way of fully exploring the patterns that members of the sample valued the various items. We saw, for example, that whereas salary and health care were clearly the two most important job qualities for the non-tenure line faculty, the third and fourth most important items were less clear. Contract had a slightly higher sample mean than did chair support, and work load was more frequently ranked third than either of these other two items. On the other hand, chair support had nearly equal numbers of respondents ranking it third, fourth, and fifth. Thus, it is difficult to say with much certainty what the third most important job quality is, for example. The descriptive information provides useful insights into this issue.

The strategy for analyzing ranking data also included models for examining relationships among the rankings and between the rankings and covariates associated with the raters themselves. MDS is a powerful tool for investigating how item rankings are related to one another, and how individual respondents cluster with respect to the ranked items. In this case, we saw that salary and health care were consistently the two most highly ranked items, given that they appeared close together in the middle of the participant cluster. In contrast, the travel budget lay furthest from the participant cloud and from the other ranking items. Together, these results reflect the consistent lack of importance with which the participants rated this element.

The relative importance of the ranking items can be further explored using the PLM. This approach provides information about the worth placed on the items by the study participants, as well as whether these worth values differ from one another in the population. A major advantage of the PLM is its ability to incorporate both item and person covariates for the rankings. In the contract faculty example, two participant level covariates were included in the model in order to ascertain whether they are associated with the individual worth assigned to each of the items. From these analyses, we saw that the highest degree attained by the respondent was associated with the worth assigned to contract length, such that those with a higher terminal degree gave this item a higher rank.

The statistical tools necessary to analyze ranking data are available in the R software environment. As demonstrated here, they can be applied in a relatively straightforward manner and the results integrated so as to provide a full picture of the ranking patterns and the covariates that are associated with them. Therefore, researchers who are faced with this type of data have a variety of options available to them for gaining deeper insights into the rankings provided by a sample than could be obtained through more traditional statistical tools that treat items in isolation. It is hoped that the current manuscript and the accompanying R code will prove to be helpful for researchers who work with ranking data.

References

Acuna-Soto, C., Liern, V., & Perez-Gladish, B. (2021). Normalization in TOPSIS-based approaches with data of different nature: Application to the ranking of mathematical videos. *Annals of Operations Research, 296*, 541-569.

Adlakha, K. & Sharma, S. (2020). Brand positioning using Multidimensional Scaling technique: An application to herbal healthcare brands in Indian market. *Vision: The Journal of Business Perspective, 24*, 345-355.

Agresti, A. (2013). *Categorical Data Analysis*. Hoboken, NJ: John Wiley & Sons, Inc.

Alvo, M. & Cabilio, P. (1995). Rank correlation methods for missing data. *Canadian Journal of Statistics, 23*, 345-358.

Armstrong, II D.A. Bakker, R., Carroll, R., Hare, C., Poole, K.T., & Rosenthal, H. (2014). *Analyzing spatial models of choice and judgment with R*. Boca Raton, FL: CRC Press, Taylor & Francis Group.

Askell-Williams, H. Lawson, M.J. (2002). Mapping students' perceptions of interesting class lessons. *Social Psychology of Education, 5*, 127-147.

Bothung, C., Fischer, K., Schiffer, H., Springer, I. & Wolfart, S. (2015). Upper canine inclination influences the aesthetics of a smile, *Journal of Oral Rehabilitation. 42*, 144–152.

Craig, B.M., Busschblack, J.J.V., & Salomon, J.A. (2009). Modeling ranking, time trade-off, and visual analog scale values for eq-5d health states: A review and comparison of methods. *Medical Care, 47(6),* 634-641.

Firth, D. (2020). Qvcalc: Quasi Variances for factor effects in statistical models.. R software library.

Fligner, M.A. & Verducci, J.S. (1986). Distance based ranking models. *Journal of the Royal Statistical Society, Series B. 48*(3), 359-369.

Gormley, I.C. & Murphy, T.B. (2008). Exploring voting blocs within the Irish electorate: A mixture modeling approach. *Journal of the American Statistical Association, 103,* 1014-1027.

Hackert, M.Q.N., Brouwer, W.B.F., Hoefman, R.J., & van Exel, J. (2019). Views of older people in the Netherlands on wellbeing: A Q-methodology study. *Social Science & Medicine, 240*, 1-9.

Hamming, R. W. (April 1950). Error detecting and error correcting codes. *The Bell System Technical Journal. 29 (2)*: 147–160.

Harakawa, R. (2021). Ranking of importance measures of tweet communities: Application to keyword extraction from COVID-19 tweets in Japan. *IEEE Transactions on Computational Social Systems, 8(4)*, 1029-1040.

Hatzinger, R. & Maier, M.J. (2017). prefmod: Utilities to fit paired comparison models for preferences. R software library.

Jeong, B-G & Kwon, H-K. (2016). buApplication of Multidimensional Scaling (MDS) structural analysis: A case study of dog food products. *Journal of Science and Technology, 9*(44).

Kamishima, T. & Akaho, S. (2006). Efficient clustering for orders. In *Proceedings of the 2nd International Workshop on Mining Complex Data*, Hong Kong, China (pp. 274-278).

Kidwell, P., Lehman, G., & Cleveland, W.S. (2008). Visualizing incomplete and partially ranked data. *IEEE Transactions on Visualization and Computer Graphics, 14(6),* 1356-1363.

Krabbe, P.F.M., Salomon, J.A., Murray, C.J.L. (2007). Quantification of health states with rank-based nonmetric multidimensional scaling. *Medical Decision Making, 27(4),* 395-405.

Lee, P.H. & Yu, P.L.H. (2015). pmr: Probability models for ranking data. R software library.

Lohr, C.A., Cox, L.J., & Lepczyk, C.A. (2012). Patterns of hypothetical wildlife management priorities as generated by consensus convergence models with ordinal ranked data. *Journal of Environmental Management, 113*, 237-243.

Luce, R.D. (1959). *Individual choice behavior: A theoretical analysis*. Wiley, New York.

Mair, P., de Leeuw, J., Groenen, P.J.F., & Borg, I. (2021). smacof: Multidimensional scaling. R software library.

Moors, G. & Vermunt, J. (2007). Heterogeneity in post-materialists value priorities: Evidence from a latent class discrete choice approach. *European Sociological Review, 23*, 631-648.

Mollica, C. & Tardella, L. (2014). Epitope profiling via mixture modeling of ranked data. *Statistics in Medicine, 33(21)*, 3738-3758.

Peiro-Palomino, J. & Picazo-Tadero, A.J. (2018). OECD: One or many? Ranking countries with a composite well-being indicator. *Social Indicators Research, 139*, 847-869.

Plackett, R.L. (1975). The analysis of permutations. *Applied Statistics, 24(2),* 193-202.

Qualtrics. (2020). Qualtrics software. Provo, UT: USA.

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Regenwetter, M., Ho, M.H.R., & Tsetlin, I. (2007). Sophisticated approval voting, ignorance priors, and plurality heuristics: A behavioral social choice analysis in a Thurstonian framework. *Psychological Review, 114(4),* 994-1014.

Strobl, C., Wickelmaier F., & Zeileis, A. (2011) Accounting for individual differences in Bradley-Terry models by means of recursive partitioning. *Journal of Educational and Behavioral Statistics, 36(2),* 135–153.

Turner, H., Kosmidis, I., Firth, D., & van Etten, J. (2021a). Modelling rankings in R: The PlackettLuce Package. *Computational Statistics, 35,* 1027-1057.

Turner, H., Kosmidis, I., Firth, D., & van Etten, J. (2021b). PlackettLuce. R software library.

Citation:

Finch, H.. (2022). An Introduction to the Analysis of Ranked Response Data. *Practical Assessment, Research & Evaluation*, 27(7). Available online: https://scholarworks.umass.edu/pare/vol27/iss1/7/

Corresponding Author:

Holmes Finch
Ball State University

Email: whfinch [at] bsu.edu

Appendix

```
library(readxl)
library(pmr)
library(PlackettLuce)
library(prefmod)
library(qvcalc)
library(smacof)

#READ AND PREPARE THE DATA#
#Experience: 1=0-5, 2-6-10, 3=11-15, 4-16-20, 5=21+
#Degree: 1=BA/BS, 2=MA/MS, 3=Specialist, 4=PhD

faculty.rankings<-data.frame(faculty.survey[,1:6])
faculty.rankings.agg<-rankagg(faculty.rankings)

#DESCRIPTION OF THE SAMPLE#
faculty.desc<-destat(faculty.rankings.agg)
faculty.desc #DESCRIPTIVES
sd(faculty.survey$contract)
sd(faculty.survey$salary)
sd(faculty.survey$health_care)
sd(faculty.survey$workload)
sd(faculty.survey$chair_support)
sd(faculty.survey$travel_budget)

#SMACOFF#
faculty.smacof = smacofRect(faculty.survey[,1:6], itmax=1000)
plot(faculty.smacof, joint=TRUE, plot.type="confplot", what="both")
plot(faculty.smacof, plot.type = "Shepard")

#MUTLIDIMENSIONAL PREFERENCE#
mdpref(faculty.rankings.agg, rank.vector=T) #2 dimensions
mdpref(faculty.rankings.agg, rank.vector=T, ndim=3) #3 dimensions

#TEST FOR RANDOM MEAN#
null_mean<-rep(3.5,6)
A<-((12*41)/(6*(6+1)))
chi<-A*sum((faculty.desc$mean.rank-null_mean)^2)
chi
dchisq(chi,5)

#COMPARE RANKINGS ACROSS GROUPS#
bachelors.rankings<-faculty.survey[ which(faculty.survey$degree==1),1:6]
graduates.rankings<-faculty.survey[ which(faculty.survey$degree>1),1:6]
```

```
bachelors.rankings.agg<-rankagg(bachelors.rankings)
graduates.rankings.agg<-rankagg(graduates.rankings)

bach.ranks<-destat(bachelors.rankings.agg)
grad.ranks<-destat(graduates.rankings.agg)
chisq.test(cbind(as.vector(bach.ranks$mar), as.vector(grad.ranks$mar)))
fisher.test(cbind(as.vector(bach.ranks$mar), as.vector(grad.ranks$mar)))

t.test(bachelors.rankings[,1],graduates.rankings[,1])
t.test(bachelors.rankings[,2],graduates.rankings[,2])

#PHI COMPONENT AND WEIGHTED DISTANCE BASED MODEL#
faculty.phicom<-phicom(faculty.rankings.agg)
faculty.wdbm<-wdbm(faculty.rankings.agg, dtype="foot")
faculty.phicom@min
faculty.wdbm@min
faculty.phicom@coef
faculty.wdbm@coef

#PlackettLuce analysis#
faculty.rankings2<-as.rankings(faculty.rankings)
faculty.mod_mle <- PlackettLuce(faculty.rankings2, npseudo=0)
coef(faculty.mod_mle)
coef(faculty.mod_mle, log=FALSE)
summary(faculty.mod_mle)  #CATEOGRY 1 WORTH IS THE REFERENCE
summary(faculty.mod_mle, ref=NULL)  #MEAN WORTH IS THE REFERENCE

faculty.mod_mle.itempars<-itempar(faculty.mod_mle, vcov=TRUE)

#QUASI STANDARD ERRORS#
faculty.qv<-qvcalc(faculty.mod_mle) #QUASI STANDARD ERRORS
summary(faculty.qv)
plot(faculty.qv, xlab="Job qualities", ylab="log of worth", main="Log worth of
job qualities for contract faculty")


#ITEM PROBABILITIES FOR TOP RANK#
faculty.itempars<-itempar(faculty.mod_mle, ref= 1, log = TRUE, vcov=TRUE)
attributes(faculty.itempars)

itempar(faculty.mod_mle, ref=1:6)
faculty.itempars.probabilities<-itempar(faculty.mod_mle, ref= 1:6)
attributes(faculty.itempars.probabilities)
faculty.itempars.probabilities
```

```
   #PLACKETTLUCE MODEL WITH COVARIATES#
   faculty.survey$grad<-ifelse(faculty.survey$degree>1,1,0)
   summary(rol(faculty.rankings2,faculty.survey$experience))
   summary(rol(faculty.rankings2,faculty.survey$grad))

   #PLACKETTLUCE TREE#
   faculty.n<-nrow(faculty.survey)
   faculty.g<-group(faculty.rankings2, index = rep(seq_len(faculty.n), 1))
   faculty.tree  <-  pltree(faculty.g  ~  grad+experience,  data  =  faculty.survey,
minsize = 2, maxdepth = 3)
   faculty.tree
   plot(faculty.tree)
```