Reprinted from IFPress (2004) 7(4)2,3

From Ranks to Intensities

Daniel M. Ennis

Background: In a classic voting problem, 24 people are asked to choose among three restaurants – one specializing in steak, one in Thai food, and one vegetarian. Nine vote for steak, eight for Thai and seven for vegetarian. The vegetarians would much prefer Thai to steak and the Thai choosers would prefer vegetarian to steak. Since the steak lovers won, they all go out for steak and most of them are dissatisfied. A ranking approach in which some account is taken of second and third choices would lead to a fairer outcome. Voting for candidates in the U.S. suffers from the same shortcomings and some other countries attempt to address these problems using various methods that incorporate ranking-type features.

Ratings are commonly used to determine the intensity of an item on some characteristic, such as liking or sweetness. Attempts to directly measure intensities using rating scales can be problematic if ratings data exhibit end effects or non-interval scale properties, as might arise when categories are labeled with words. In addition, outlier ratings can lead to unrepresentative results. A model for rating data that accounts for scale usage was discussed in an earlier newsletter article and is available as part of the IFPrograms software[1]. Deriving intensity information from ranked data may sometimes be a useful alternative to ratings. When ranking data are obtained, how do we construct an intensity scale on which to place the ranked objects taking into account the fact that the ranking process itself might introduce dependencies into the ranking decision? The purpose of this report is to discuss an approach to converting ranks to intensities so that the degree of difference among the ranked objects can be assessed.

Scenario: You are interested in the relationship between the age of whole milk and its freshness to consumers assessed by odor evaluation. In a large consumer study, 990 consumers who regularly consume whole milk evaluate 10 milk samples aged for varying periods of time at 40°F. Data collection is made manageable by arranging the 10 samples into incomplete blocks of 5 samples per block. Each consumer corresponds to a single block and the samples are presented within a block in random order. There are 18 blocks in the basic plan and 55 repetitions of the plan. The incomplete block design is shown as Table 1.

Block (Consumer)	Samples				
1	1	2	3	4	5
2	1	2	3	6	7
3	1	2	4	6	9
4	1	2	5	7	8
5	1	3	6	8	9
6	1	3	7	8	10
7	1	4	5	7	10
8	1	4	8	9	10
9	1	5	7	9	10
10	2	3	4	8	10
11	2	3	5	9	10
12	2	4	7	8	9
13	2	5	6	8	10
14	2	6	7	9	10
15	3	4	6	7	10
16	3	4	5	7	9
17	3	5	6	8	9
18	4	5	6	7	8

Table 1. Incomplete block design for the milk freshness study. The basic plan of 18 blocks (consumers) is repeated 55 times and the samples are presented in random order to be ranked by each consumer

Each consumer ranks the 5 presented samples in order from the least fresh to the most fresh, evaluating the samples based on odor. Your goal is to determine the relative location of the 10 products on a freshness scale obtained by modeling the ranked data.

Basis for Ranking: Individual consumers differ in their perceptions of any product and their perceptions of a given product vary from time to time within a consumer. We can think of the 10 milk samples as 10 distributions of freshness intensity of equal variance to capture the ideas about variability just described. Our goal is to locate the positions of the distribution means using individual ranking data. When a consumer orders the 5 samples on freshness, we assume that at the moment of ranking, the perceived freshness intensities are such that the highest ranked sample has the greatest intensity followed by the next highest sample and so forth. There are a large number of possible orders of the 5 objects (5! or 120) and some are more likely to occur than others for a particular set of intensities. Models for full and partial ranking data based on the idea that object perceptions follow a normal distribution have been published[2,3]. The models are complex and computer intensive, and in certain cases they can be inaccurate. Even if the objects are assumed to be perceived independently, the resulting model is computationally complex. However, there is another approach that leads to great simplicity in modeling ranked data that will now be described.

Rank-Induced Dependencies: The ranking process is likely to induce correlations among adjacent pairs of ranked samples. This type of effect is not included in the ranking models just described[2,3]. Assume that the most highly ranked object is most highly correlated with the next highest ranked object, and that the correlation coefficients decrease successively until the most highly ranked object is least correlated to the least highly ranked object as shown in Figure 1. This induction of dependency among the ranked objects would seem to introduce tremendous complexity to an already complex problem as already described, especially if one is free to choose any correlation coefficients within the constraint given. Remarkably, the opposite turns out to be the case. By

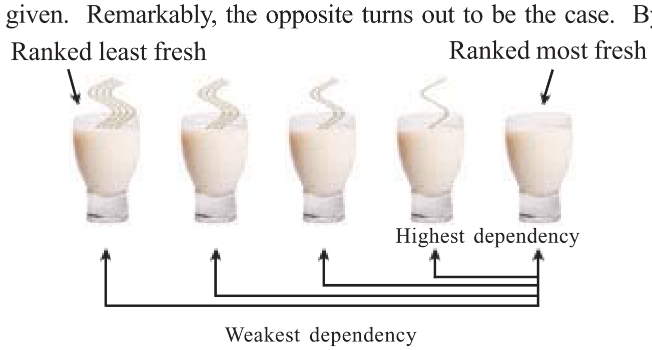

Figure 1. The correlation or dependency among samples is greatest for the two freshest samples and least for the freshest and least fresh ranked samples

assuming rank-induced dependencies, the ranking model simplifies to a product of univariate normal distribution functions irrespective of the number of objects ranked[4]. In addition to providing a rapid and highly accurate method for estimating the location of the intensity means of the products on the dimension used to produce the ranks, the model also provides information on the rank-induced dependencies. In other words, it separately accounts for the underlying perceptual information about the ranked objects and also accounts for effects associated with the ranking process itself.

From Ranks to Intensities: As shown in Table 1, there are 18 blocks, each one with a particular set of 5 of the 10 samples and the plan is repeated 55 times. This means that 55 consumers are randomly assigned to each of the 18 blocks leading to a total of 990 consumer evaluations. In the case of each block, or set of 5 samples, the consumer ranks the objects with respect to freshness. Counts for each of the rank orders are obtained as shown in Table 2 and the data are fit to the new ranking model using the method of maximum likelihood. The purpose of the analysis is to find the mean freshness value for each sample. There is a univariate normal distribution corresponding to each sample. The mean of each distribution is the mean freshness of each sample in perceptual standard deviation units and differences among the samples correspond to ι values. These are the familiar units of measurement that we have derived from other methods, such as difference testing methods and ratings[5,6,7]. Figure 2 shows the relationship between the age of the sample and the degree of freshness. Freshness generally decreases with sample age, but it can be seen that product freshness remains constant until after the fourth time point. After that time, the decrease is rapid until the tenth time point.

Conclusion: Ranks provide a convenient and useful method for determining degrees of differences among ranked objects on the sensory or hedonic basis for producing the ranks. Unlike ratings, ranks do not require strong assumptions about consistent category usage and understanding of category labels. In addition,

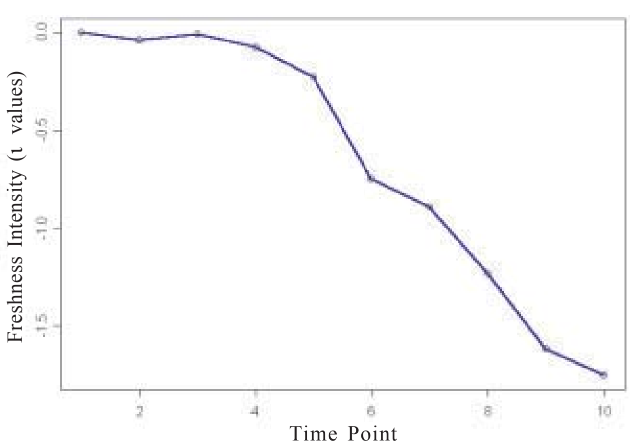

Figure 2. **Derived freshness intensity plotted against storage time at 40°F**

Least Fresh				Most Fresh	Count
1	2	3	4	5	2
1	2	3	5	4	1
1	3	2	4	5	1
...	...	...	...	...	...
9	8	5	6	3	7
9	8	6	1	3	13
10	9	6	7	2	11
10	9	7	1	5	10
10	9	7	2	6	2
10	9	7	5	1	8
10	9	7	6	2	14
10	9	8	1	4	9
10	9	8	4	1	10

Table 2. **A subset of sets of 5 samples ranked from least fresh to most fresh and the frequency with which that rank outcome occurred (count)**

ranks do not require the extensive testing that would be required if all possible pairs of objects were compared in a paired comparison design. A major difficulty with deriving perceptual intensities from ranks has been the complexity of the models that result from modeling the ranking process. However, a new model that introduces the idea of rank-induced dependencies resolves the complexity of the problem and at the same time provides additional information on dependencies created by the ranking process. From this analysis, the results of other choice procedures can be predicted such as first choice, 2-alternative forced choice, and the triangular method. Using this model it can be shown how the freshness of a food product changes over time so that a better understanding of the consumer perception of product quality can be achieved.

References:

1. Ennis, D.M. (1999). Thurstonian models for intensity ratings. *IFPress*, **2(3)**, 2-3.

2. Böckenholt, U. (1992). Thurstonian representation for partial ranking data. *British Journal of Mathematical and Statistical Psychology*, **45**, 31-49.

3. Yao, G. and Böckenholt, U. (1999). Bayesian estimation of Thurstonian ranking models based on the Gibbs sampler. *British Journal of Mathematical and Statistical Psychology*, **52**, 79-92.

4. Ennis, D.M. (2004). A Thurstonian model for ranks based on rank-induced dependencies. *Submitted.*

5. Ennis, D.M. (1993). The power of sensory discrimination methods. *Journal of Sensory Studies*, **8**, 353-370.

6. Ennis, D.M. (1998). Thurstonian scaling for difference tests. *IFPress*, **7(3)**, 2-3.

7. Dorfman, D.D. and Alf, E., Jr. (1969). Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals - Rating method data. *Journal of Mathematical Psychology*, **6**, 487-496.