DOCUMENT RESUME

ED 347 191 TM 018 671

AUTHOR Linacre, John M.
TITLE Rank-Order and Paired Comparisons as the Basis for Measurement.
PUB DATE Apr 92
NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 20-24, 1992).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Attitude Measures; Case Studies; *Comparative Analysis; *Consumer Economics; *Evaluators; Forced Choice Technique; Mathematical Models; *Rating Scales; Test Construction
IDENTIFIERS Interval Scales; *Paired Comparisons; Preference Data; *Rank Order; Rasch Model; Taste Preference

ABSTRACT

Three case studies are presented demonstrating the application of straight-forward Rasch techniques to rank order data. Paired comparisons are the simplest form of rank ordering. A consumer preference test with 56 pairs of cups of coffee tasted by each of 26 consumers illustrates analysis of these rankings. When subjects are allowed the option of "no difference," an approach analogous to a rating scale is used. Data from a study by A. Springall (1973) with about 28 assessors judging the flavor strength of a product illustrate analysis of the situation in which ties are allowed. A convenient method of constructing measures from rank orders is to regard the rankings as ordered categories on a rating scale. Data from D. E. Critchlow (1985) illustrate partial rankings of three top choices of crackers by 22 small boys and 16 mothers. These approaches demonstrate methods of producing measures from rankings by judges. Four figures and six tables present details of the analyses. A nine-item list of references and three appendixes of preference data are included. (SLD)

Rank-Order and Paired Comparisons as the Basis for Measurement

by

John M. Linacre

MESA Psychometric Laboratory
University of Chicago
5835 S. Kimbark Ave.
Chicago IL 60637

Presented at the
American Educational Research Association
Annual Meeting

April 1992

Introduction

Rank order data have features which sometimes make them more useful than data based on pre-defined rating scales (Linacre 1990), but the utility of such data has been severely restricted by the lack of convenient and informative analytical techniques. A ranked comparison, whether of a pair of objects or of larger groupings, produces ordinal counts not interval measures. Consequently, to analyze rankings using techniques designed for interval data necessarily distorts and confuses their meaning. Though there are analytical techniques intended precisely for these types of data (Bradley & Terry 1952, David 1988, Critchlow 1985), they tend to be inaccessible and demanding on the analyst.

In this paper, three case studies are presented, demonstrating the application of straight-forward Rasch techniques to these types of data.

Paired Comparisons: Forced Choice

Paired comparisons are the simplest form of rank ordering. Bradley & El-Helbawy (1976) present the results of a consumer preference test conducted by General Foods Corporation. Brew strength, roast color and coffee brand were each tested at two levels, resulting in 8 different coffee treatments. Each treatment was paired with the others for a paired test, resulting in 8x7=56 pairs of cups of coffee. In the test itself, 56 pairs of cups of coffee were tasted by each of 26 consumers.

Preferred Treatment	Treatment not preferred SDY	SDX	SLY	SLX	WDY	WDX	WLY	WLX
SDY	-	15	15	16	19	14	19	16
SDX	11	-	10	15	15	14	15	12
SLY	11	16	-	15	15	14	18	15
SLX	10	11	11	-	14	11	15	13
WDY	7	11	11	12	-	9	14	13
WDX	12	12	12	15	17	-	16	18
WLY	7	11	8	11	12	10	-	12
WLX	10	14	11	13	13	8	14	-

Table 1. Coffee Preferences of 26 consumers.

The preference data are presented in Table 1. Here are shown counts of the number of consumers who preferred the row treatment to the column treatment. Each treatment is conceptualized to represent the additive effect of three facets: brew strength (S or W), roast color (D or L) and brand (X or Y). Thus each facet contains two elements. The elements were not identified in the published data, but have been assigned convenient labels for use here (e.g. S = Strong). The paired comparison is specified by opposing the measures of the three facet elements of the second treatment against those of the first.

Measurement Model

$$\log\left(\frac{P_{brtb'r't'}}{1 - P_{brtb'r't'}}\right) = (B_b+R_r+T_t) - (B_{b'}+R_{r'}+T_{t'})$$

where $P_{brtb'r't'}$ is the probability that combination brt is preferred to b'r't'

B_b is the brew strength of element b
R_r is the roast color intensity of element r
T_t is the brand type measure of brand t
$B_{b'}$ is the brew strength of element b'
$R_{r'}$ is the roast color intensity of element r'
$T_{t'}$ is the brand type measure of brand t'

Appendix 1 contains the *Facets* (Linacre 1988) specifications for this analysis. There are 3 facets: Brew, Roast and Brand. The ordering of the facets in the data records is specified in the "Entered=" statement. Each data line (following "Data=") consists of element numbers for the first treatment in order by facet, followed by the element numbers for the second treatment in the same order. Finally, the number of times the first treatment is preferred over the second is recorded. Thus "1,2,1,2,1,2,14" means that Treatment 1 is "1,2,1", i.e, facet 1 element 1, facet 2 element 2, facet 3 element 1, which is Strong, Light, Brand Y (SLY). Similarly, Treatment 2 is Weak, Dark, Brand X (WDX). Finally SLY is preferred to WDX fourteen times. To assist with interpretation, greater preference (i.e. greater scores) correspond to more positive measures in all three facets. This is the meaning of "Positive=1,2,3". The frame of reference is established by anchoring the measures of the least preferred elements at zero. Consequently, all three facets are non-centered.

The "Models=" statement is "Models=?,?,?,-?,-?,-?,B26". This specifies that, from the sum of the measures corresponding to the first set of three elements, the sum of the measures of the second set of three elements is to be subtracted. From the resulting logit value the number of expected successes in 26 binomial trials (B26) is to be estimated.

The "Labels=" section identifies the names of the facets and the elements within the facets.

The results of the *Facets* analysis are presented graphically in Figure 1. The measures themselves are presented in Table 2. The Brew and Roast elements are noticeably different. The Brands are almost indistinguishable. Measures preceded by "A" are preset to establish the frame of reference. The count is that of the number of cells in which each element is contrasted with the other element in the same facet. The Observed Average, e.g. 14.9, is the average number of times a treatment containing that element, e.g. "Strong" is preferred over a treatment containing the other element, e.g. "Weak" in the same facet.

Score	Count	Observed Average		Calib Logit	Model Error	Infit MnSq	Infit Std	Outfit MnSq	Outfit Std	N Element
Brew Strength										
239	16	14.9		0.30	0.10	0.7	-1	0.7	-1	1 Strong
177	16	11.1	A	0.00	0.10	0.7	-1	0.7	-1	2 Weak
Roast Color										
229	16	14.3		0.20	0.10	0.7	0	0.8	0	1 Dark
187	16	11.7	A	0.00	0.10	0.7	0	0.8	0	2 Light
Coffee Brand										
210	16	13.1		0.09	0.10	0.5	-1	0.5	-1	1 Brand Y
206	16	12.9		0.07	0.10	0.5	-1	0.5	-1	2 Brand X

Table 2. Measures from Coffee Preference paired comparisons.

```
Logit| Brew     | Roast  | Brand
------------------------------------------
 .3  | Strong   |        |
 .2  |          | Dark   |
 .1  |          |        | Brand X   Brand Y
  0  * Weak     * Light  *
```

Figure 1. Depiction of measures from Coffee Preference data.

Paired Comparisons: Ties Allowed

Allowing subjects the option of "no difference" complicates the analysis of paired comparisons (Davidson 1970). The approach used here is analogous to a rating scale. When treatments A and B are compared, a preference of A is rated with 2, a preference of B with 0, and a tie is rated 1. Springall (1973) presents such data. 28 Assessors were asked to state which treatment of a pair had the greater flavor strength. Three flavor concentrations were crossed with three gel concentrations giving 9 different treatments. There were thus 9x8 pairings. The data is shown in Table 3. The numbers give the count of assessors who stated that the row treatment the stronger flavor. The numbers after the comma are counts of those who perceived "no difference".

Stronger Flavor	Weaker Flavor								
	WL	IL	SL	WM	IM	SM	WH	IH	SH
WL	-	2,7	0,1	5,10	2,7	0,2	12,9	13,5	10,3
IL	16	-	6,5	17,7	9,8	8,6	24,0	16,5	15,7
SL	21	11	-	22,4	14,5	11,7	24,3	22,3	17,6
WM	10	2	2	-	3,9	2,4	12,5	10,6	8,11
IM	15	6	4	14	-	5,9	20,2	15,5	13,9
SM	22	12	4	18	11	-	27,0	19,5	18,2
WH	3	2	1	6	0	1	-	2,8	2,5
IH	9	3	0	6	2	1	13	-	6,8
SH	12	2	2	9	5	2	21	8	-

Table 3. Flavor strength comparisons by about 28 assessors.

Measurement Model

$$\log\left(\frac{P_{sgs'g'j}}{P_{sgs'g'j-1}}\right) = (S_s+G_g) - (S'_s+G'_g) - F_j$$

where $P_{sgs'g'j}$ is the probability that gel sg is rated relative to gel s'g' in category j
S_s is the flavor concentration strength element s
G_g is the gel concentration of element g
$S_{s'}$ is the flavor concentration strength of element s'
$G_{g'}$ is the gel concentration of element g'
F_j is the additional strength required to be rated in category j, j=0,2

Appendix 2 contains the *Facets* specifications for this analysis. The chief additional feature is that the observation model is no longer binomial trials, but categories of the three category rating scale. The frame of reference is established by anchoring the lowest gel concentration at 0 logits. The results of this analysis are shown in Figure 2 and Table 4.

Score	Observed Count	Observed Average	Calib Logit	Model Error	Infit MnSq	Infit Std	Outfit MnSq	Outfit Std	N Element
Flavor Concentration									
208	456	0.5	-0.94	0.07	1.0	0	1.0	0	1 W 0.6
492	440	1.1	-0.09	0.07	1.0	0	1.0	0	2 I 4.8
650	454	1.4	0.31	0.07	1.0	0	1.0	0	3 S 9.0
Gel Concentration									
579	449	1.3 A	0.00	0.07	1.0	0	1.1	0	1 L 0.0
539	440	1.2	-0.05	0.07	1.0	0	1.0	0	2 M 2.4
218	447	0.5	-1.04	0.07	1.1	1	1.1	1	3 H 4.8

Table 4. Measures obtained from Coffee Preference Data

```
Logit|Flavor|Gel
------------------
   .3 | S 9.0|
      |      |
      |      |
    0 *      * L 0.0
  -.1 | I 4.8| M 2.4
      |      |
      |      |
      |      |
      |      |
      |      |
      |      |
  -.9 | W 0.6|
 -1.0 +      + H 4.8
```

Figure 2. Depiction of Measures from Flavor Strength Comparisons

Rank ordering

A convenient method of constructing measures from rank orders is to regard the rankings as ordered categories on a rating scale. The scale definition is established spontaneously by each judge. With this approach, tied and partial rankings present no unusual difficulties. Critchlow (1985 p.119) presents the partial rankings of five types of crackers by 22 small boys and also by 16 mothers. He reports rankings of only their top 3 choices. The aim of the analysis is to compare how the boys ranked the crackers with how the mothers did.

Measurement Model

$$\log\left(\frac{P_{cj}}{P_{cj-1}}\right) = C_c - F_j$$

where P_{cj} is the probability that cracker c is ranked in category j
C_c is desirability of Cracker c
F_j is the additional desirability required to be ranked in category j, j=0,3
Note: ranks are converted into rating as follows:

Rank	Rating
1	3
2	2
3	1
Unranked (4 and 5)	0

The data is presented in Table 5. The specifications for a BIGSTEPS (Wright et al. 1992) analysis of this data are shown in Appendix 3. Each set of three letters in the data corresponds to one ranking. Three crackers in each ranking are assigned their rank order number. The unranked, but less preferred crackers, are given the joint rank of 4. The results of the two BIGSTEPS analyses, one for the boys and the other for the mothers, are shown in Table 6.

Boys' 22 Partial Rankings	Mothers' 16 Partial Rankings
ACS GCA ACG CAG CGA ARC CSA SCR AGC ARG AGC ACS GRA CGA ACS CGS ARC ACG RAC AGC ACG CAG	CRA SRG CSA CSA SRA SCR SCG GAR SAR CSA RSC RAG SCG SAR GAS SCA
Partial rankings are in the form: first, second, third choice. A=animal crackers, C=cheese crackers, G=graham crackers, R=Ritz crackers, S=saltines.	

Table 5. Partial ranking data of cracker preference.

NAME	Boys' 22 Rankings SCORE	MEASURE	ERROR	MNSQ	Mother's 16 Rankings SCORE	MEASURE	ERROR	MNSQ
Animal Crackers	41	-.54	.22	1.17	48	.41	.23	.53
Cheese Crackers	48	-.24	.20	.75	43	.16	.22	1.20
Graham Crackers	64	.36	.20	.86	54	.78	.27	1.18
Ritz Crackers	76	.95	.26	1.30	48	.41	.23	.95
Saltines	79	1.18	.30	1.17	31	-.43	.24	1.19

Table 6. Measures from Boys and Mothers partial rankings.

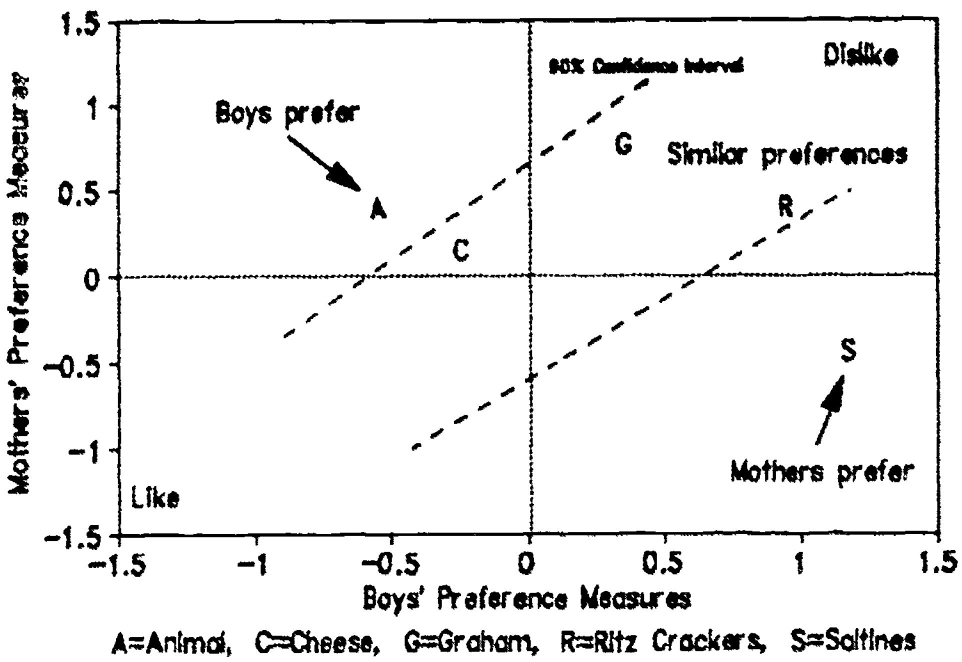

Figure 3. Comparison of Boys and Mothers crackers preference measures.

The measurement analysis has provided us with two frames of reference: that of the boys and that of the mothers. They are compared in Figure 3. We can now determine how well each mother's ranking fits in the boys' frame of reference and how well each boy's ranking fits in the mothers' frame of reference. Two further analyses were performed in which all the boys' and mothers' rankings were included in both analyses. In the first analysis, the calibrations for the crackers were anchored at the values obtained from the earlier boys' analysis. In the second analysis, the calibrations were anchored at the values obtained from the earlier mothers' analyses. Thus two fit statistics were obtained for each ranking - one in the mother's frame of reference, one in the boy's frame of reference. The fit statistic used is a mean-square variance ratio statistic with expectation 1, minimum value 0, and infinite maximum value.

Figure 4 is a cross-plot of the two values of the ratio-scale fit statistic obtained for each ranking. As can be seen, it is clear that most rankings fit with their own frame of reference, but misfit the other frame. There are exceptions, those marked by arrows and those in the top right quadrant, which could provoke further investigation.

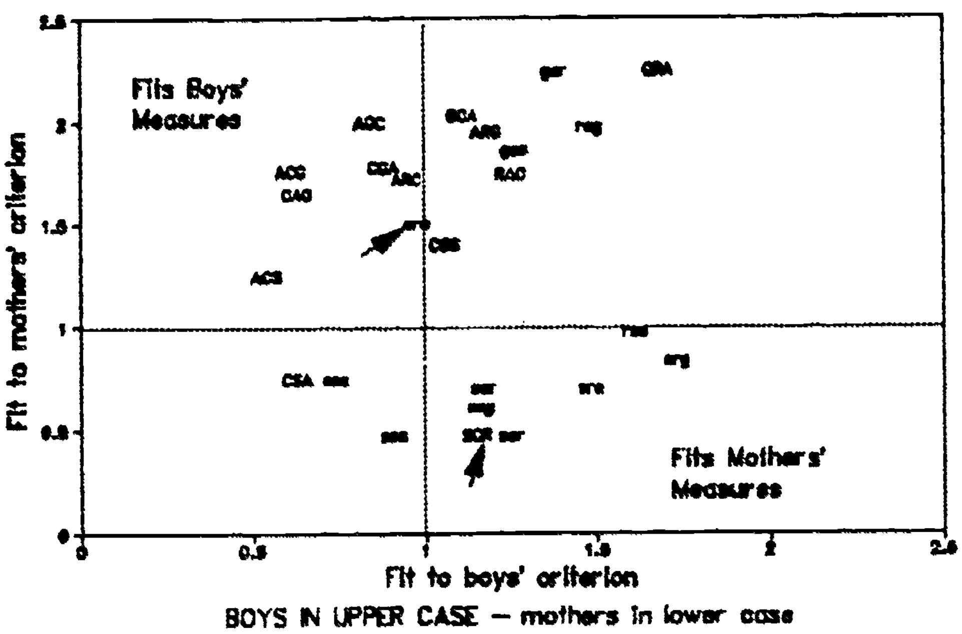

Figure 4. Diagnostic plot of Boys and Mothers rankings of crackers.

Conclusion

In some circumstances, judge-assigned rankings have distinct advantages over judge-assigned ratings. This paper demonstrates several methods of producing measures from rankings, and also illustrates their value in enhancing our understanding of the latent variables, and also identifying features of the data which depart from the consensus.

References:

Bradley R.A., El-Helbawy A.T. 1976. Treatment contrasts in paired comparisons. Biometrika 63, 2, 255-262.

Bradley R.A., Terry M.E. 1952. The method of paired comparisons. Biometrika 39, 324-345

Critchlow D.E. 1985. Metric methods for analyzing partially ranked data. New York: Springer.

David H.A. 1988. The method of paired comparisons. London: Charles Griffin.

Davidson R.R. 1970. On extending the Bradley-Terry model to accommodate ties in paired comparison experiments. J. Amer. Statist. Assoc. 65, 317-328.

Linacre J.M. 1988. Facets computer program for Rasch analysis. Chicago: MESA Press.

Linacre J.M. 1990. Rank ordering or Judge-awarded rankings. Paper presented at AERA, Boston, Mass. April. (ERIC ED 318 802)

Springall A. 1973. Response surface fitting using a generalization of the Bradley-Terry paired comparison model. Applied Statistics 22:1 59-68.

Wright B.D. et al. 1992. BIGSTEPS program for Rasch analysis. Chicago: MESA Press.

Appendix 1. Coffee Preference Data: *Facets* specifications.

```
Title=Preference data in coffee testing
Facets=3
Entered=1,2,3,1,2,3       ; 3 elements in each of two treatments
Positive=1,2,3
Non-centered=1,2,3

Models=?,?,?,-?,-?,-?,B26          ;  paired  comparisons:  first
treatment against second

Labels=
 1,Brew Strength,A
  1,Strong
  2,Weak,0
 *
 2,Roast Color,A
  1,Dark
  2,Light,0
 *
 3,Coffee Brand
  1,Brand Y
  2,Brand X
*

data=
1,1,1,1,1,2,15      (SDY vs. SDX: 15 out of 26 preferred SDY)
1,1,1,1,2,1,15
1,1,1,1,2,2,16
1,1,1,2,1,1,19
1,1,1,2,1,2,14
1,1,1,2,2,1,19
1,1,1,2,2,2,16
1,1,2,1,2,1,10
1,1,2,1,2,2,15
1,1,2,2,1,1,15
1,1,2,2,1,2,14
1,1,2,2,2,1,15
1,1,2,2,2,2,12
1,2,1,1,2,2,15
1,2,1,2,1,1,15
1,2,1,2,1,2,14
1,2,1,2,2,1,18
1,2,1,2,2,2,15
1,2,2,2,1,1,14
1,2,2,2,1,2,11
1,2,2,2,2,1,15
1,2,2,2,2,2,13
2,1,1,2,1,2,9
2,1,1,2,2,1,14
2,1,1,2,2,2,13
2,1,2,2,2,1,16
2,1,2,2,2,2,18
2,2,1,2,2,2,12
```

Appendix 2. Flavor Strength Data: *Facets* specifications.

```
Title=Flavor Strength using Rating Scale
Facets=2
Positive=1,2
Non-center=1,2
Entered=1,2,1,2
Models=?,?,-?,-?,R2

Labels=
1=Flavor
1=W 0.6
2=I 4.8
3=S 9.0
*
2=Gel,A
1=L 0.0,0
2=M 2.4
3=H 4.8
*

data=
2,1,1,1,0
2,1,1,1,0
2,1,1,1,1
2,1,1,1,1
2,1,1,1,1
2,1,1,1,1
2,1,1,1,1
2,1,1,1,1
2,1,1,1,1
2,1,1,1,2
2,1,1,1,2
2,1,1,1,2
2,1,1,1,2
2,1,1,1,2
2,1,1,1,2
2,1,1,1,2
2,1,1,1,2
2,1,1,1,2
2,1,1,1,2
2,1,1,1,2
2,1,1,1,2
2,1,1,1,2
2,1,1,1,2
2,1,1,1,2
2,1,1,1,2
3,1,1,1,1
|   (More data follows)
```

Appendix 3. Partial rankings of Crackers: BIGSTEPS specifications.

```
&INST
 TITLE = "Partially Ranked Data by Rating Scale"
 CODES = 1234
 ITEM1 = 17
 NI    = 39
 IDELQU=Y
 ; For Boys' - delete items 24-39
 ; For Mothers' - delete items 1-22
&END
ACS
GCA
ACG
CAG
CGA
ARC
CSA
SCR
AGC
ARG
AGC
ACS
GRA
CGA
ACS
CGS
ARC
ACG
RAC
AGC
ACG
CAG
(Blank)
cra
srg
csa
csa
sra
scr
scg
gar
sar
csa
rsc
rag
scg
sar
gas
sca
END NAMES
Animal Crackers 1312313411113314112112 3433344223424223
Cheese Crackers 2221131234324121323321 1411422441342442
Graham Crackers 4133244423241242434233 4344443144433414
Ritz Crackers   4444424342442444241444 2244234334114344
Saltines        3444442144434433444444 4122111412241131
```