RANK ANALYSIS IN PAIRED COMPARISON DESIGN

BY
S.C. GUPTA AND S.C. RAI
I.A.S. R.I. New Delhi-12
(Received : August, 1976)

1. INTRODUCTION

The analysis of experiments involving paired comparisons has received considerable attention in Psychological and Statistical methodologies. The method of paired comparison is described in detail by Titchner [10] in one of the earliest text books on psychological experiments. A number of other authors like Thurstone [9], Kendall and Smith [7], Guttman [6], Bradley and Terry [3], Bradley [1] [2], Gridgeman [4] and Rai [8] have contributed significantly to the development of the methodologies for paired comparisons.

In quality testing experiments, an observer examines the objects and arranges them in the order in which he judges them to possess the quality under consideration. More generalised method of ranking is prefered in quality testing experiments because if quality considered is not representable by a linear variable the ranking method gives incorrect results. The method of paired comparisons is useful in such situations. Experiments involving ranking within small groups of treatments or items seem particularly appropriate in sensory judgement investigations.

Here, a method of analysis for ranking in paired comparisons is developed. The method involves postulating a mathematical model involving treatment parameters, estimation of these parameters; the development of test procedures and investigations of the properties of the model and procedures developed.

2. MODEL FOR PAIRED COMPARISONS

Let us consider 't' treatments in an experiment involving paired comparisons. We postulate that these treatments have true ratings or preference $\pi_1, \ldots, \pi_t$ on a particular subjective continum throughout the experiment. The continuum is subject to the requirements that every $\pi_i \geqslant 0$ and

$$\sum_{i=1}^{t} \pi_i = 1.$$

Further we assume that when treatment i appears with treatment j in a block the probability that treatments i obtains the top rating (or a rank of 1) is taken to be

$$\pi_i^2 \Big/ \left(\pi_i^2 + \pi_j^2 \right).$$

Tied ranks are not permitted in the model. If r_{ijk} indicated the rank of the i-th treatment in the k-th repetition of the block in which treatment i appears with treatment j, then $r_{ijk} + r_{jik} = 3$.

3. The Likelihood Ratio Test and Estimation

By virtue of the fact that the treatment ratings corresponding to any two treatments examined by the observer should be independent of each other, we can assume probability independence of the treatment comparisons. The probability of the observed ranking in the k-th repetition for the block in which treatment i and j are compared is given by

$$\left(\frac{\pi_i^2}{\pi_i^2 + \pi_j^2} \right)^{2-r_{ijk}} \left(\frac{\pi_j^2}{\pi_i^2 + \pi_j^2} \right)^{2-r_{jik}}$$

$$= \frac{\pi_i^{4-2r_{ijk}} \pi_j^{4-2r_{jik}}}{\left(\pi_i^2 + \pi_j^2 \right)}$$

For, if treatment i obtains top ranking,

$$r_{ijk} = 1$$

and $$r_{jik} = 2$$

thus the above expression reduces to

$$\pi_i^2 \Big/ \left(\pi_i^2 + \pi_j^2 \right).$$

Alternatively, if

$$r_{ijk} = 2$$

and $$r_{jik} = 1$$

above expression reduces to

$$\pi_j^2 \Big/ \left(\pi_i^2 + \pi_j^2 \right).$$

When we multiply the appropriate expression for all comparisons within a repetition and for all n repetitions, the likelihood function is obtained as

$$L=\frac{\prod_{i} \pi_i^{4n(t-1)-2 \sum_{i\neq j}^{t} \sum_{k=1}^{n} r_{ijk}}}{\prod_{i<j} \left(\pi_i^2 + \pi_j^2 \right)^n}$$

Let the 't' treatments be grouped into m groups. Then a general class of tests of the null hypothesis

$$H_o : \pi_i = 1/t \qquad (i=1, 2, \ldots, t) \qquad \ldots(2)$$

against alternative hypothesis

$$H_a : \pi_i = \pi(h) \qquad (h=1, 2, \ldots, m) \qquad \ldots(3)$$

$$i = S_{h-1} \neq 1, \ldots, S_h$$

where $S_o = 0,\ S_m = t$

and $$\sum_{i} (S_h - S_{h-1})\, \pi\,(h) = 1,$$

are possible using likelihood ratio tests. This in other words means that tests of null hypothesis of identical treatment ratings may be performed against the alternative hypothesis that the treatments have identical ratings within a group of treatment where as m groups themselves may differ. Alternative hypothesis involving only a subset of parameters do not lead to parameter free tests.

Using Lagrangian multipliers, we maximize the logarithm of the likelihood function to obtain $p(h)$, the maximum likelihood estimate of $\pi(h)$, these estimates are obtained from the equations

$$\left[\left\{ 4n(t-1)\,(S_h - S_{h-1}) - 2 \sum_{i=S_{h-1}+1}^{Sh} \sum_{j}{}' \cdot \sum_{k} r_{ijk} - \tfrac{1}{2} n\,(S_h - S_{h-1}) \times (S_h - S_{h-1} - 1) \right\} \Big/ p\,(h) \right] - 2n\, p\,(h) \times (S_h - S_{h-1})$$

$$\sum_{f\neq h} (S_f - S_{f-1}) / \{ p^2(h) + p^2(f) \} = 0 \qquad (h=1, 2, \ldots, m) \qquad \ldots(4)$$

and $$\sum_{h} (S_h - S_{h-1})\; p(h) = 1 \qquad \ldots(5)$$

The general test statistic, a monotone function of the likelihood ratio is

$$B=n\sum_{h<f}(S_h-S_{h-1})\,(S_f-S_{f-1})\,\log\{p^2(h)+p^2(f)\}$$

$$-\sum\{4n(t-1)\,(S_h-S_{h-1})-\sum_{i=S_{h-1}-1}^{S_h}\;\sum_j{}'\;\sum_k r_{ijk}-\frac{n}{2}(S_h-S_{h-1})$$

$$\times(S_h-S_{h-1}-1)\}\log p(h). \qquad \ldots(6)$$

Where $\sum\limits_j{}'$ means that one value of j that appears in the argument of summation is omitted.

B is a function of the treatment sum of ranks.

Solving equation (4) and (5) we obtain the estimates of the true treatment ratings. Pairwise comparison of these estimates provides a quantitative measure of the ratings of a pair of items relative to the test attribute.

Now we consider here two special cases of the general alternative hypothesis given by (3) above.

Case (i) H_1 : *no* π_i is assumed equal to any π_j $(i\neq j)$. This is obtained if in the general hypothesis H_a, there is only one treatment in a group so that $m=t$. The equations (4) and (5) for this case are derived as follows:

$$\text{Log}\,L=\sum_i a_i\log p_i-n\sum_{i<j}\log(p_i^2+p_j^2)$$

$$\frac{\partial\log L}{\partial pi}=\frac{a_i}{p_i}-2n\sum_{j\neq i}p_i\,(p_i^2+p_j^2)^{-1}=0 \qquad \ldots(7)$$

and $$\sum_{i=1}^{t}p_i=1 \qquad \ldots(8)$$

where $$a_i=4n\,(t-1)-2\sum_j{}'\sum_k r_{ijk} \qquad \ldots(9)$$

The test statistic becomes

$$B_1=n\sum_{i<j}\log((p_i^2+p_j^2)-\sum_i a_i\log p_i\ldots\ldots \qquad \ldots(10)$$

The preparation of tables for the exact distribution of B_1 is discussed in the following section.

Case (ii) $H_2 : \pi_i = \pi \ (i=1, \ldots, S)$

$$\pi_i = \frac{1-S\pi}{t-S} (i=S+1,\ldots\ldots,t)$$

This is the reduction of the general hypothesis to the case in which, $m=2$.

For this case the maximum likelihood equations (4) and (5) may be solved simply and the test statistic B_2 written as an explicit function of treatment sum of ranks.

4. Tables for B_1 and B_2

All combinations of treatment sum of ranks can be generated for any given number of treatments and repetitions of the paired comparison design. We may obtain the probability of each such combinations under the null hypothesis of equality of true treatment ratings.

When three items are compared in a single repetition the possible sets of ranks sum are 2, 3, 4 and 3, 3, 3. Each of the six permutations of the elements of first set has a probability 1/8, while the probability of the second set is 2/8. The treatment sum of ranks for three treatments and two repetitions can be obtained by adding 2, 3, 4 and 3, 3, 3, in turn to corresponding elements in the sets of sum of ranks consisting of all permutations of 2, 3, 4 and to 3, 3, 3. All permutations of a given set of treatment sum of ranks are taken to be equivalent in the sets of sum of ranks so produced. The probability of a given permutation is obtained by multiplying the basic probabilities of combination and the permutation used to produce the given permutation. The probability of a given new combination of rank sums is obtained by adding the probabilities obtained for each permutation of the elements of the combination.

The procedure may be arranged systematically as shown in Table—1.

TABLE—1

The generation of treatment sum of ranks probabilities for three treatments and two repetitions

Probabilities	*Rank Sums*	*1/8*	*1/8*	*1/8*	*1/8*	*1/8*	*1/8*	*2/8*
		2,3,4	*2,4,3*	*3,2,4*	*3,4,2*	*4,2,3*	*4,3,2*	*3.3,3*
6/8	2,3,4	4,6,8	4,7,7	5,5,8	5,7,6	6,5,7	6,6,6	5,6,7
2/8	3,3,3	5,6,7	5,7,6	6,5.7	6,7,5	7,5,6	7,6,5	6,6,6

The combination 5, 6, 7, say, appears in its various permutations in nine places in the above table. In row 1, column 4, for example, 5, 6, 7 appears and its probability is 6/64 obtained by multiplying marginal probabilities of row and column. The probability of the combination 5, 6, 7 is then the sum of the nine individual probabilities and has the value 36/64. When three repetitions with three treatments are considered the generating rows at the top of the table unchanged, but the columns at the left above are replaced by the possible combinations of sum of ranks obtained for two repetitions with their corresponding probabilities. This procedure is continued for large numbers of treatments and repetitions.

When the sets of possible combinations of treatment sum of ranks are obtained with their probabilities of occurrance for each such set we substitute in equations (7), (8) and (9) to obtain estimates $P_1, P_2, \ldots P_t$. The solution of these equations is tedious, in some cases elementary methods are applicable in others it is necessary to use repeated approximations in an iterative procedure.

The equations can be most easily solved by taking the values of $P_1, P_2, \ldots, P_t$ in the initial trial in proportion to

$$(\Sigma r_2)\ldots\ldots(\Sigma r_t) : (\Sigma r_1)\,(\Sigma r_3)\ldots\ldots(\Sigma r_t) : \ldots\ldots$$
$$: (\Sigma r_1)\,(\Sigma r_2)\ldots\ldots(\Sigma r_{t-1}).$$

where $\Sigma r_1, \Sigma r_2, \ldots\ldots, \Sigma r_t$ are the sum of ranks for treatments T_1, T_2, ..., T_t respectively. The values are good first approximations in most cases. Using iterative procedure the final estimates of p_1, p_2, ..., p_t are obtained. If $p_i^{(0)}$ $(i=1,\ldots, t)$ denote the first approximation, then equation (7) gives the second approximations as

$$p_i^{(1)_2}=\frac{a_i}{2n}\left[\frac{1}{p_1^{(1)_2}+p_i^{(0)_2}}+\ldots\ldots+\frac{1}{p_{i-1}^{(1)_2}+p_i^{(0)_2}}\right.$$
$$\left.+\frac{1}{p_{i+1}^{(0)_2}+p_i^{(0)_2}}+\ldots\ldots+\frac{1}{p_t^{(0)_2}+p_i^{(0)_2}}\right]^{-1}$$

5. Combination of Experiments

Experiment may be performed in groups of repetitions of sizes

$$n_u \quad (u=1,\ldots g)$$

with

$$\sum_{u=1}^{g} n_u=n.$$

These groups may be judges in sensory experimentation, different localities, days etc. There are different methods for testing the significance depending upon the specification of the alternative hypothesis. We shall illustrate them as follows :

1. *Pooled Analysis*

If we assume in the alternative hypothesis that the true treatment ratings $\pi_1, \ldots, \pi_t$ exist for all groups of repetitions then no new analysis is required. Total treatment sums of ranks are obtained by adding the corresponding sums of ranks for each group. The experiment may be treated as though one group of n repetitions had been employed and tables of Appendix (A) may be used.

2. *Combined Analysis*

In some cases it is not realistic to assume that the same true ratings exist for all groups. Now let us specify an alternative hypothesis as follows:

(*a*) Within the uth of g groups, true ratings $\pi_{1u} \ldots\ldots, \pi_{tu}$;

$\sum_{i=1}^{t} \pi_{iu}=1$ exist and they may be changed from group to group.

(*b*) Group Experiments are independent in probability. If we define $B_1{}^u$ to be the likelihood ratio statistic corresponding to B_1 for the uth group $(u=1, 2, \ldots g)$ we can combine the group and perform an overall test of significance which depends on a statistic.

$$B_1{}^c = \sum_{u=1}^{g} B_1{}^u$$

The decision to pool or combine group results should be made from a prior knowledge of group behaviour. A measure of consistency of ranking from group to group is provided by the difference between the pooled value of B_1 and $B_1{}^c$. Small values of $B_1 - B_1{}^c$ will indicate good agreement in ranking from group to group while large values indicate discordant rankings.

If we set up the hypothesis

$$H_0 : \pi_{iu} = \pi_i \ (u=1, \ldots, g; \quad i=1, \ldots t)$$
$$H_1 : \pi_{iu} \qquad (u=1, \ldots g; \quad i=1, \ldots t)$$

unrestricted by groups

then $$-2 \log_e \lambda = 2\left(B_1 - B_1^c \right) \log_e 10 \qquad \ldots(11)$$

where λ is the likelihood ratio statistic for comparison of H_0 and H_1, $B_1 - B_1^c$ is then a monotone function of likelihood ratio statistic.

LARGE SAMPLE DISTRIBUTIONS

If λ is the likelihood ratio, it is known that $-2 \log_e \lambda$ is distributed as x^2 under very general conditions. This result can be applied in the special cases considered above.

In the first special test :

$$-2 \log_e \lambda = n\, t\, (t-1) \log_e (2) - 2 B_1 \log_e 10$$

is distributed in the limit as x^2 with $(t-1)$ degrees of freedom.

6. THE EXPERIMENTAL PROCEDURE AND ANALYSIS ILLUSTRATED

We shall use the data from the taste testing experiment on the flavour characterstic of pork roasts as given as an example by Bradley and Terry (1952). We shall redescribe the experiment in detail.

In a taste-testing experiment, pork roasts were compared by ranking in pairs on their flavour characterstics. The roasts were obtained from three groups of hogs which had been fattended on three different rations, Corn (Maize), corn plus a peanut supplement, and corn plus a large peanut supplement. The object was to determine whether the addition of peanuts to the diet was recognizable in the fresh-pork roasts or not. Expert judges were asked to rank pairs on the basis of flavour attributable to the peanut diet. It is useful to show a systematic listing of the steps involved with reference to the results of two of the several judges used in the experiment described above. Each judge performed five repetitions of the paired design ($t=3$, $n=5$)

PROCEDURE

Step 1. (experimental) : A competent panel of judges was selected and so instructed that they all had experience with the experimental material.

Step 2 : Six small containers were coded for each judge and for each repetition. Two samples from roasts from each of the three treatment groups of animals were placed in the containers and the three requisite pairs formed. Code numbers were recorded and the pairs presented to the judge in a random order together with score cards.

Step 3 : For each pair a judge tested each sample and recorded the value 1 for the sample preferred and 2 for the other sample.

Step 4 : (Analysis). The experimenter collected and decoded the data for each judge and recorded the results as follows, C denotes the corn ration, C_p the corn plus peanut supplement ration and cp the corn plus large peanut supplement ration. The treatment sum of ranks,

Σr_i for c, C_p, cp are respectively 19, 13, 13 and 13, 15, 17 for the two judges.

TABLE 2

Ranking for Two Judges In the Pork Experiment

Repetition	*1*			*2*			*3*			*4*			*5*		
	c	C_p	cp	c	C_p	cp	c	C_p	cp	c	C_p	cp	c	C_p	cp
Pair								*Judge 1*							
c, c_p	2	1	—	2	1	—	2	1	—	2	1	—	2	1	—
c, Cp	2	—	1	1	—	2	2	—	1	2	—	1	2	—	1
c_p, Cp	—	2	1	—	1	2	—	2	1	—	1	2	—	2	1
								Judge 2							
c, c_p	2	1	—	2	1	—	1	2	—	1	2	—	1	2	—
c, Cp	1	—	2	1	—	2	1	—	2	1	—	2	2	—	1
c_p, Cp	—	2	1	—	1	2	—	2	1	—	2	1	—	1	2

Step 5 : For judge 1,

$$P_C=0.14$$
$$p_{C_p}=0.43$$
$$p_{C_P}=0.43$$
$$B_1=2.917$$

The significance level is 0.057. For judge 2

$$p_C=0.43$$
$$p_{C_p}=0.33$$
$$p_{C_P}=0.24$$
$$B_1=0.34.$$

The significance level is 0.404.

Step 6 : The pooled sums of ranks over two judges are 32, 28,30. The table of Appendix, Gupta [5] for

$$n=10$$

gives

$$p_C=0.29$$
$$p_{C_p}=0.33$$
$$p_{C_P}=0.33$$
$$B_1=8.797$$

and the significance level is 0.630. Since it seems extremely unlikely on the basis of this method that treatment differences are present, it is not meaningful here to compare treatments by use of their estimated ratings.

Summary

A method of analysis of experiments has been developed involving paired comparisons which can easily be extended to the case in which only a fraction of the pairs are retained. This model for paired comparisons permits test of hypothesis of a general class and the estimation of treatment ratings or preferences. It is simple and easy to interpret and apply. The method of maximum likelihood is employed and the tests depend upon the likelihood ratio statistic. Two special tests are considered to test the null hypothesis that true treatment ratings are equal. The alternative hypothesis (*i*) makes no assumptions of equality of treatment ratings and (*ii*) makes the assumption that there are only two groups of treatments where-in within group treatments do not differ in ratings but the groups themselves may have different ratings. The methods of pooling and of combining the results of several judges are given which permit an over-all test of significance. The utility and application of the model are explained by a numerical example.

References

[1] Bradley, R.A. (1953) : Some Statistical methods in taste-testing and quality evaluation, Biometrics 9, 22-35.

[2] Bradley, R.A. (1954) : Rank analysis of incomplete block designs II Additional tables for the method of Paired comparisons, Biometrika 41, 502-37.

[3] Bradley, R.A. and Terry M.E. (1952) : Rank analysis of incomplete block designs I—The method of paired comparisons *Biometrika*, 39 ; 324-45,

[4] Gridgeman, N.T. (1955) : The Bradley-Terry probability model on preference testing, Biometrics 11, 335-343.

[5] Gupta, S.C. (1974) : Planning and Analysis of experiments involving rankings Unpublished thesis of M.Sc. in Ag. Statistics of Indian Agricultural Research Institute, New Delhi.

[6] Guttman, L. (1946) : An approach for quantifying paired comparisons and rank orders. Annals of Math. Stat. 17, 145.

[7] Kendall, M.G., (1939) Smith, B.B. : On the method of paired comparisons. Biometrika 31, 324-345.

[8] Rai, S.C. (1971) : On a model for rank analysis in Fractional paired comparisons. 'J.I.S.A.S.', 23, 61-68.

[9] Thurstone, L.L. (1927) : Psychophysical Analysis Amer. J. Psycho. 368.

[10] Titchner, E.B. (1901) : Experimental Psychology I Macmillan & Co. Ltd., London.

APPENDIX

Table for Rank Analysis of Incomplete Block Design

($t=3$, $b=3$, $r=2$, $k=2$, $\lambda=1$) for $n=10$

The following table gives the values of the likelihood ratio statistic, B_1 and the likelihood estimate of the true treatment ratings $p_1, p_2, \ldots, p_t$ together with probabilities P that B_1 will not be exceeded if the null hypothesis is true, $\Sigma\, r_i$ is the sum of ranks for treatment i.

Σr_1	Σr_2	Σr_3	P_1	P_2	P_3	B_1	P
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
20	30	40	.89332	.10667	.00000	.0614	.0000
20	31	39	.87696	.09227	.03075	1.4649	.0000
21	29	40	.74341	.25658	.00000	1.4127	.0000
20	32	38	.87186	.08542	.04270	2.2251	.0000
22	28	40	.66235	.33764	.00000	2.1737	.0000
20	33	37	.86959	.07881	.05159	2.7037	.0000
23	27	40	.60106	.39893	.00000	2.6532	.0000
21	30	39	.70047	.22780	.07171	2.8718	.0000
20	34	36	.86789	.07272	.05937	2.9734	.0000
24	26	40	.54800	.45190	.00000	2.9230	.0000
20	35	35	.86736	.06631	.06631	3.0609	.0000
25	25	40	.49999	.49999	.00000	3.0102	.0000
21	31	38	.68993	.21095	.09910	3.6845	.0000
22	29	39	.61748	.29506	.08745	3.6842	.0000
21	32	37	.68330	.19659	.12009	4.2210	.0000
23	28	39	.55868	.34537	.09594	4.2204	.0000
22	30	38	.60744	.27233	.12021	4.5495	.0000
21	33	36	.68020	.18223	.13786	4.5546	.0001
24	27	39	.51041	.38894	.10063	4.5539	.0001
21	34	35	.67879	.16795	.15325	4.7155	.0001
25	26	39	.46690	.43037	.10273	4.7149	.0001
22	31	37	.60220	.25305	.14473	5.1411	.0002
23	29	38	.55026	.31807	.13170	5.1410	.0002

(*table contd.*)

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
22	32	36	.59823	.23595	.16581	5.5342	.0005
24	28	38	.50419	.35742	.13837	5.5339	.0005
22	33	35	.59669	.21888	.18442	5.7603	0.009
25	27	38	.46386	.39424	.14188	5.7601	.0009
23	30	37	.54607	.29551	.15841	5.7881	.0012
22	34	34	.59621	.20189	.20189	5.8342	.0014
26	26	38	.42844	.42844	.14310	5.8339	.0014
23	31	36	.54273	.27601	.18125	6.2383	.0025
24	29	37	.50143	.33196	.16660	6.2382	.0025
23	32	35	.54145	.25708	.20145	6.5250	.0046
25	28	37	.46268	.36593	.17138	6.5249	.0046
23	33	34	.54087	.23873	.22038	6.6647	.0047
26	27	27	.42799	.39856	.17343	6.6647	.0074
24	30	36	.49998	.30964	.19036	6.7451	.0090
24	31	35	.49920	.28916	.21162	7.0898	.0157
25	29	36	.46234	.34142	.19622	7.0900	.0157
24	32	34	.49880	.26970	.23148	7.2907	.0261
26	28	36	.42960	.37123	.19016	7.2907	.0261
24	33	33	.49768	.25115	.25115	7.3569	.0320
27	27	36	.39970	.39970	.20059	7.3567	.0320
25	30	35	.46230	.31925	.21843	7.4922	.0399
25	31	34	.46235	.29853	.23911	7.7516	.0674
26	29	35	.43039	.34742	.22222	7.7516	.0674
25	32	33	.46238	.27862	.25898	7.8789	.1035
27	28	35	.40128	.37479	.22392	7.8788	.1035
26	30	34	.43094	.32541	.24364	8.0689	.1306
26	31	33	.43131	.30454	.26413	8.2549	.2112
27	29	34	.40232	.35157	.24610	8.2549	.2112
26	32	32	.43144	.28437	.28427	8.3162	.2571
28	28	34	.37627	.37627	.24744	8.3162	.2571
27	30	33	.40406	.32901	.26691	8.4988	.3250
27	31	32	.40450	.30802	.28747	8.6189	.5009
28	29	33	.37795	.35363	.26840	8.6189	.5009
28	30	32	.37873	.33178	.28952	8.7971	.6229
28	31	31	.37899	.31050	.31050	8.8559	.7762
29	29	32	.35465	.35465	.29068	8.8165	.7752
29	30	31	.35502	.33309	.31187	8.9728	.9644
30	30	30	.33333	.33333	.33333	9.0308	1.0000