An Analysis of Rank Ordered Data

Krishna P Paudel, Louisiana State University

Biswo N Poudel, University of California, Berkeley

Michael A. Dunn, Louisiana State University

Mahesh Pandit, Louisiana State University

Corresponding Author:

Krishna P Paudel

Department of Agricultural Economics and Agribusiness

Louisiana State University Agricultural Center

Baton Rouge, LA 70803

Phone: 225 578-7363

Fax: 225 578-2716

Email: kpaudel@agcenter.lsu.edu

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2009 AAEA & ACCI Joint Annual Meeting, Milwaukee, WI, July 26-28, 2009.

An Analysis of Rank Ordered Data

Abstract

Many methods are available to analyze rank ordered data. We used a spectral density method to analyze Formosan subterranean termite control options ranked by Louisiana homeowners. Respondents are asked to rank termite control options from the most preferred to the least preferred option. Spectral analysis results indicated that the most preferred termite control choice is a relatively cheap ($0.13 per square foot) option of liquid treatment.

Keywords: FST, rank ordered data, spectral analysis

An Analysis of Rank Ordered Data

The categories of ordinal variables cannot be measured in ratio or interval scale because the social distance cannot be measured quantitatively. In order to do a valid analysis, this type of data is ranked. For example, in research, opinion is often ranked as strongly agree, agree, disagree, and strongly disagree, but the distance between any two categories is not measurable in quantity. Rank order data are often coded as consecutive integer from 1 to n category. To illustrate, a group of homeowners may rank invasive insect control options as first, second, third, fourth preference and so on. Further, the ordinal data can be in a different form; therefore, the first thing to do rank order data analysis is to classify the observed rank data in a particular pattern that meets our objectives.

When an individual ranks all of choices according to his preference, it is difficult to analyze such preference data. There are several approaches to analyzing such kind or preference rank data, but one of the best ways is to fit model that represent measure of the interest. In particular, the preference rank data is a permutation group. A permutation means any one of the arrangement of a given set. A precise definition of permutation based on the notion of function: A permutation of nonempty set G is one-one function onto itself. So the objective of this paper is to use a permutation group spectral analysis to find the most preferred FST control option in Louisiana.

Literature Review

There exist several approaches to analyze rank order data. Few examples includes nonparametric analysis of unbalanced paired- comparison or ranked data (Andrew and David, 1990). Andrews and David compared simple nonparametric method of analyzing unbalanced ranked data to an existing method of rank analysis for unbalanced data. Haunsperger (2003) states that the Kruskal-Wallis nonparametric statistical test on n samples can be used to rank-order a list of alternatives and it is subjects to such a Simpson –kike paradox of aggregation. In addition, Krusakal-Wallis samples contain two or more data sets in which each may individually support a certain order. Another way to study the rank preference data is a Bayesian investigation for rank ordered multinomial logit models (Koop and Poirier, 1994). It is also used as a test for independence of irrelevant substitute hypothesis. In case of multinomial logit models, Luce and Suppes (1965) derive the probability associated with a particular ranking of alternatives. Koop and Poirier (1994) applied this idea to investigate voter preferences in 1988 Canadian Federal Election. An alternative way of investigating preference of rank data is completely randomized factorial designs (James *et al.,* 1976). This procedure is an extension of the Kruskal-Wallis rank test that allows for the calculation of interaction effects and linear contrasts. Paudel et al. (2007) applied exploded logit and ordered probit models to identify the most preferred Formosan termite control method in Louisiana. Thompson (1993) applies a generalized permutation polytopes and exploratory graphical method for ranked data. The author presents an exploratory graphical method to display frequency distribution for fully and partially ranked data. Shirley (1981) demonstrated a standard analysis of covariance computer program to analyze rank order data. Wallis (1939) used correlation ratio of ranked data. Diaconis (1989) has focused on the spectral analysis method in order to find the preference from a ranked ordered data.

Data

Data for the study come from a survey of homeowners regarding the preference of Formosan subterranean termites (FST) control options in Louisiana. FST are invasive species of termites which is currently present in more than 13 states in the U.S. The damage is so severe that if not controlled infested houses become uninhabitable. Most damage by the FST is caused in Louisiana where the damage cost reaches close to a billion a year. Four FST control options are provided for each individual homeowner to rank from the most preferred choice to the least preferred choice. The FST control choices provided are i. No control – cost zero, ii. Liquid treatment: cost $0.13 per square foot, iii. Bait treatment: $0.43 per square foot, iv. Liquid + bait treatment: $0.56 per square foot. Individuals ranked these options as first, second, third and fourth preferred option to control termites.

Model

We applied a spectral density method to the data from survey in which homeowner expressed their preferences for various Formosan subterranean termite (*Coptotermes formosanus Shiraki*) control methods. By this method a variety of inferential methods are considered, and spectral ideas are then extended to general homogenous spaces.

Spectral Analysis: Choice of controlling the FST sometimes is based on the rank preference of people. For example, considering the termite control option, (which is described below in details) is a permutation group, Let π(i) be the ranked given to i^{th} control method. Then the collection of such ranking makes up a data set. First order spectral analysis is the linear combination of the number of times item i is ranked in position j. And, the second order spectral analysis is the number of times items i and i' are ranked in the positions j and j'. Let S4 denote the symmetric group on 4^{th} letters. A function of f on Sn, with f(π) being the number of rankers choosing ranking preference π forms a data set. Here data can be considered as a permutation group with decomposition of space of all function into invariant orthogonal subspaces. By group theory, we can write the general model as

$$f(\pi) = \sum_{p} f p(\pi)$$

where p indexes the various subspaces and f_p denotes the projections. In general Let X be the finite set and G be a finite permutation group operating transitively on X. Then, L(x) is the space of all function on X with values in R. Then, L(x) decomposes into a direct sum of invariant irreducible subspace, as follows, In particular for S4 decomposes in five irreducible forms

$L(x) = V1 \oplus V2 \oplus V3 \oplus V4 \oplus V5$

If f(x) be a data set or the number of times x appears in the sample, spectral analysis is the projection of f onto the invariant subspaces and the approximation of f by as many pieces as required to give a reasonable fit. And the projection is given by discrete Fourier transformation

$f(x)=\frac{1}{n}\sum_{j=0}^{n-1} f(j)e^{-\frac{2\pi i j x}{n}}$ $f(x)=\frac{1}{n}\sum_{j=0}^{n-1} f(j)e^{-\frac{2\pi i j x}{n}}$

First order spectral analysis

Survey results indicated that a total of 972 ranked results were obtained, out of these respondents only 747 were completely ranked and others are partially ranked in the four alternative methods. The complete preference ranked by respondents is shown in Table 1. The complete ranked data is analyzed first and then, partially ranked data is analyzed later. The entrees of columns of Table 1 shows the control method ranked in the given permutation. Thus, 1234 respondents ranked No control method in first preference, liquid treatment in second, bait treatment in third and Liquid+ bait treatment in fourth preference.

By observing the 24 numbers in table 1—some of the counts are much larger than others. Table 2 show the percentage of respondents ranking preference i in position j. It is clearly seen that liquid treatment method is most preferred control method, which is being ranked first by 52.2 percentage of the respondents. Similarly, bait treatment, which is rank by 55.7 percentages of respondents, is taken as second most alternative control method.

Higher order analysis

The data vector can be considered in a function of f(π), where π is the permutation group and f(π) is number of respondents choosing the preference π. Thus f(1234)=123. Let M be the space of all real valued function of symmetric group S4. This is vector space under addition of functions.

The usual inner product on M is defined by

$$(f1/f2)=\sum_{\pi} f1(\pi)\, f2(\pi) \quad .$$

The space M decomposes uniquely into the direct sum of five subspaces. These are shown with their dimension in Table 3. The space V1 is the set of constant functions. This has one dimension. The space V2 will be called the space of first order functions. A function $\pi \rightarrow \delta i\, \pi(j)$, which is 1 if $\pi(j) = I$ and 0 otherwise, which only depends on value of one coordinate. Then, the first order general function is in the following form:

$$\Sigma_{(i,j)} [a_{ij}] \; (\delta_i\, \pi(j))$$

Such that $\sum_{i,j} a_{ij} = 0$

V3 is the space of second order function. In this space, as first order, a typical element:

$$\Sigma_{(i,i',j,j')} [a_{ii',jj'}] \; (\delta_{\{i,i'\}} \{\pi(j), \pi(j')\})$$

Here, $a_{ii',jj'}$ are chosen so that V3 is orthogonal to V1 $\oplus$ V2.

Table 4 gives the first order summary of Table 2 on the basis of number, which has entry i,j the number of respondents i in the j^{th} preference minus the sample size over 4, so the both rows and columns sum is zero. Where, the entries are rounded to integer.

The largest number 213, in first ranking indicates that liquid control receive most first place preference. The largest number in second ranking indicates that bait is second alternative to termite control. The largest positive number in third column 173 means that liquid and bait treatment is preferred in third alternative method in for termite control. Similarly, the largest number 337 in fourth column for No control treatment indicates that, each respondents wants a kind of treatments to the termite control. In addition, no one wants to leave the invasive termite without any treatment.

Conclusions

We presented first order spectral analysis methods of analyzing rank order data based on 747 completed ranked data. Results of spectral analysis indicate that second preference or liquid treatment effect is chosen by respondents as the most desired control method for termite control in Louisiana.

References

Andrew, D. M. and H. A. David. Nonparametric analysis of unbalanced paired-comparison or ranked data. *Journal of the American Statistical Association* 85(1990):1140-1146.

Breslaw, J.A. Multinomial probit estimation without nuisance parameters. *The Econometrics Journal*, 5(2002):417-434.

Diaconis, P. A generalization of spectral analysis with application to ranked data. *The annals of statistics* 17(1989):949-979.

Diaconis, P. (1988) Group Representation in Probability and Statistics. Lecture Notes--Monograph Series, Volume 11, Hayward, CA: Institute of Mathematical Statistics, 1988.

Koop, G. and D.J. Poirier. Rank-ordered logit models: an empirical analysis of Ontario voter preferences. Journal of Applied Economietrics 9(1994):369-388.

Luce, R.D. and P. Suppes. Preference, utility and subjective probability. Handbook of Mathematical Psychology, John Wiley and Sons, 1965.

Paudel, K.P., M.A. Dunn, D. Bhandari, R.P. Vlosky, and K.M. Guidry. Alternative methods to analyze the rank ordered data: a case of invasive species control. *Natural Resource Modeling* 20(2007):451-471.

Shirley, E.A. A Distribution-Free Method for Analysis of Covariance Based on Ranked Data. *Applied Statistics* 30(1981):158-162.

Scheirer C.J., Ray W.S., and N. Hare. The analysis of ranked data derived from completely randomized factorial designs. *Biometrics* 32(1976):429-434.

Thompson, G.L. Generalized Permutation Polytopes and Exploratory Graphical Method for Ranked Data. *The Annals of Statistics* 21(1993):1401-1430.

Wallis, W.A. Correlation Ratio for Ranked Data. *Journal of the American Statistical Association* 34(1939):533-538.

Table 1. Preference ranking on FST control options

S. No (π)	Ranking				Number
	First Choice	Second Choice	Third Choice	Fourth Choice	
1	1	2	3	4	123
2	1	2	4	3	1
3	1	3	2	4	6
4	1	3	4	2	4
5	1	4	2	3	1
6	1	4	3	2	7
7	2	1	3	4	55
8	2	1	4	3	1
9	2	3	1	4	15
10	2	3	4	1	305
11	2	4	3	1	24
12	2	4	1	3	0
13	3	1	2	4	1
14	3	1	4	2	1
15	3	2	1	4	2
16	3	2	4	1	48
17	3	4	1	2	2
18	3	4	2	1	39
19	4	1	2	3	2
20	4	1	3	2	0
21	4	2	1	3	2
22	4	2	3	1	20
23	4	3	1	2	0
24	4	3	2	1	88

Table 2. Percentage of respondents ranking preference i in position j

	Rank			
Method	1	2	3	4
No control	22.8	7.98	2.8	70
Liquid	52.2	26.5	18.3	1.9
Bait	12.6	55.7	30.7	0.9
liquid +bait	12.5	9.8	48.2	27

Table 3. Decomposition of the regular representation

$M^{11111} = S^{4} \oplus S^{3,1} \oplus S^{2,2} \oplus S^{2,1,1} \oplus S^{1,1,1,1}$

M	=	V1	$\oplus$	V2	$\oplus$	V3	$\oplus$	V4	$\oplus$	V5
Dim 24		1		9		4		9		1

Table 4. First order effects

	Rank			
Method	1	2	3	4
No control	-45	-127	-166	337
Liquid	213	9	-50	-173
Bait	-94	231	42	-180
Liquid + Bait	-75	-114	173	15