Nonparametric Tests for Ordering in Completely Randomized and Randomized Block Mixed Designs

Rhonda Magel* and Alfred Ndungu

Department of Statistics, North Dakota State University Fargo, ND, USA

Abstract

Two nonparametric tests are proposed in testing for k nondecreasing treatment effects under a mixed design consisting of a completely randomized portion and a randomized block portion. The randomized block portion may contain missing observations for some of the treatments within a block. A simulation study is conducted to estimate the powers of the two tests for 3, 4, and 5 treatments under a variety of different treatment effects and underlying distributions. Powers are estimated when the randomized block portion is 1/4, 1/3, 1/2, 1.0, 1.5, 2.0, 3.0 and 4.0 times the completely randomized portion. A recommendation as to which test has higher powers is given.

Keywords: Completely randomized design; Randomized block design; Missing observations; Ordered alternative

Introduction

There are times when we may wish to compare the effects of k treatments without having to make assumptions as to the underlying distributions. Initially, we might plan on using a randomized complete block design or a repeated measures design. Blocks may be formed when not all the experimental units are homogenous. A block would consist of a collection of experimental units that are homogenous. In a complete block design there will be enough experimental units within each block so that every treatment can be randomly applied to an experimental unit within the block. It is possible that missing observations for one or more of the treatments within a block may occur. In some cases, the researcher may wish to use a completely randomized design in order to collect additional information about the treatments. In this case, treatments would be randomly assigned to the experimental units and no blocking would occur. This would be in addition to the randomized block portion of the design.

As an example as to when we might want to use this type of mixed design, a business may have several thousand employees and wants to reduce the annual cost of health care for the employees. The business attempts to do this through educating workers as to good nutrition and fitness, changing the type of food available for purchase by the workers while they are at work, and by making exercise programs more readily available. This is because there are certain factors that the business believes are related to health. These factors are cholesterol levels, blood pressure, BMI, the amount of sleep a person gets, and the amount and types of exercise, among others. The business asks a sample of its employees if they would volunteer to go in at certain times to get these factors measured. Each employee in this sample who volunteers is given a health number for each time period that they go in to get these factors measured. This health number is based on the values of the observed factors that have been mentioned for the employee, with a higher health number indicating better health. It is possible that an employee who was asked to volunteer may skip a time period or multiple time periods in getting these factors measured. Therefore, there could be missing observations within a block, or in this case, an individual. In testing whether or not this program is effective, the business may also opt to include a completely randomized portion of the design in which additional random samples of employees are asked to participate during one time period only. This would provide additional information regarding the treatment effects. The business would like to test the following set of hypotheses:

$$H_0: \tau_1 = \tau_2 = \ldots = \tau_k \quad \text{versus} \tag{1}$$

$$H_a: \tau_1 \leq \tau_2 \leq \ldots \leq \tau_k \quad \text{(at least one inequality is strict)}$$

with τ_i indicating the treatment effect for the ith time period and k indicating the number of treatments. In this case, the ith treatment effect is the average health score of the employees in the business at the ith time period.

Dubnicka et al. [1] give an example as to when this type of design could be used when there are only two treatments. The example they give is comparing the results from two surgical treatments used for high-risk proliferative diabetic retinopathy. In some cases, a patient will have both eyes that need treatment, and in other cases, a patient will have only one eye needing treatment. An individual patient may be considered as a block when both eyes are in need of treatment. For patients needing both eyes treated, one treatment could be randomly assigned to the right eye and one to the left eye. This would be the randomized complete block portion of the experiment. The completely randomized portion of the experiment would consist of patients with only one eye needing treatment and each of these patients being randomly assigned to one of the two treatments. Data are given in Dubnicka et al. [1].

Several well-known nonparametric tests exist for testing differences in treatment effects depending on the type of design used [2,3]. The Mann-Whitney test for two samples [4] or the Kruskal-Wallis test for k samples could be used for testing differences in treatment effects under a completely randomized design (CRD) [5]. The Jonckheere-Terpstra tests the nondecreasing hypothesis in (1) under a CRD [6,7]. Magel et al. [8] developed another test to test the nondecreasing alternative in (1) under a CRD. Their test was designed to protect against situations in

which the treatment effects were different, but were not nondecreasing. Terpstra, Terpstra et al. [9] developed another test for the nondecreasing alternative in a CRD based on Spearman's coefficient. The Wilcoxon test for paired samples [10] or the Friedman test [11,12] tests for a difference in k treatment effects in a randomized complete block design (RCBD). Page's test [13] was designed to test the nondecreasing alternative in (1) under an RCBD design. The Durbin test [14] was designed to test the differences in k treatment effects under a balanced incomplete block design (BICB). Cao et al. [15] proposed a test for the nondecreasing alternative in a balanced incomplete block design. Alvo and Cabilio [16] introduced a test for nondecreasing treatment effects in a randomized block design with missing observations.

Dubnicka et al. [1] developed a nonparametric test for a mixed two-sample design consisting of two independent samples and a paired sample testing the differences in two treatment effects. Their test statistic is formed by adding the unstandardized versions of the Mann-Whitney test [4], and the Wilcoxon paired sample test [10], together and then standardizing this sum by subtracting the combined mean and dividing by the square root of the combined variance. Under the null hypothesis of no difference in the two treatment effects, the test statistic will have an asymptotic standard normal distribution. Magel and Fu [17] introduced a test for this design which consists of adding the standardized versions of the Mann-Whitney test and the paired Wilcoxon test together and then dividing by the square root of two. Their test statistic also has an asymptotic standard normal distribution under the null hypothesis. When the sample size for the completely randomized portion is equal to or higher than the number of blocks, the Magel and Fu test has higher powers. When the number of blocks is higher than the equal sample sizes for the completely randomized design portions, the Dubnicka et al. [1] test has higher powers.

Magel et al. [18] introduced nonparametric tests for the mixed design consisting of an RCBD and a CRB for the general alternative and for the umbrella alternative. Magel et al. [19] introduced nonparametric tests for the same design when testing for nondecreasing treatment effects as in (1). The tests proposed by Magel et al. [18] were for a mixed design consisting of a completely randomized portion and, a randomized complete block portion or a repeated measures complete portion. If the block portion or repeated measures portion had missing observations, then the tests could not be used unless the blocks that were not complete were discarded.

In this research, we will propose a test for a mixed design consisting of a completely randomized portion, and a randomized block portion when there may be missing observations within a block.

Introduction of Two Tests

Alvo and Cabilio [16] introduced an extension of Page's [13] nonparametric test to include both complete and incomplete blocks. Observations are ranked within a block from smallest to largest with missing observations assigned the average rank. The Alvo and Cabilio test statistic is given by

$$AC = \sum i*R_i \quad \text{summed over i=1,2,...,k (time periods)} \qquad (2)$$

with R_i equal to the sum of ranks assigned to treatment i (or time period i).

The variance of the AC test for block i is given by

$$k_i*(k+1)/12*(k_i+1)*\sum (o_{il} - \text{average}(o_i))^2 \qquad (3)$$

where k_i is the number of treatments appearing in the block, k is the number of treatments, average (o_i) is the average of the treatment numbers, not ranks, appearing in each block, and oil is the treatment number. The values of oil range from 1 to k. The asymptotic distributions of both the AC test and Page's test when the null hypothesis is true, are normal.

Two test statistics are proposed for the mixed design consisting of a completely randomized portion and a randomized block portion. The proposed test statistics are combinations of the Jonckheere-Terpstra test and the Alvo and Cabilio test. Let Z_{JT} and Z_{AC} represent the standardized versions of the JT and AC tests, respectively. The JT test will be applied on the completely randomized portion and the AC test will be applied on the randomized block portion. The first proposed test is given by

$$Z_{first} = (Z_{JT} + Z_{AC})/\text{sqrt}(2) \qquad (4)$$

Z_{first} has an asymptotic standard normal distribution when the null hypothesis is true. Let E(JT) and the E(AC) denote the means of the JT and AC tests, respectively. Also, let Var(JT) and Var(AC) denote the variances of the JT and the AC tests, respectively. The second proposed test is given by

$$Z_{last} = ((JT+AC)-(E(JT)+E(AC))/\text{sqrt}(\text{Var}(JT)+\text{Var}(AC))$$

$$= (\text{sqrt}(\text{Var}(JT))*Z_{JT} + \text{sqrt}(\text{Var}(AC))*Z_{AC})/\text{sqrt}((\text{Var}JT)+\text{Var}(AC))$$

The asymptotic distribution of Z_{last} when the null hypothesis is true is also a standard normal distribution.

The question becomes, which test statistic has higher powers? Is Z_{first} better is some cases and Z_{last} better in other cases? It is noted that both Z_{first} and Z_{last} are sums of weighted values of Z_{JT} and Z_{AC}. Z_{first} gives equal weights to both Z_{JT} and Z_{AC}. The weights that Z_{last} give to both Z_{JT} and Z_{AC} vary depending upon the variances of each of these statistics. When the variance of the JT statistic is larger, more weight is given to Z_{JT}. When the variance of the AC statistic is larger, more weight is given to ZAC.

Simulation Study

A simulation study was conducted using SAS [20] to compare Z_{first} with Z_{last}. The tests were compared the basis of estimated powers. Ten thousand simulations were conducted under each of a variety of situations. Estimated powers were calculated for each situation by counting the number of times each test rejects and dividing this number by ten thousand. Significance levels were also estimated in each situation when the treatment effects are the same by counting the number of times each test rejects divided by ten thousand. The stated significance level was always 0.05. The asymptotic distributions of the test statistics were used and the null hypothesis was rejected if the value of the test statistic was greater than 1.645. Three, four and five treatments were considered. The study considered the following underlying distributions: exponential, student's t with 3 degrees of freedom, normal, and Cauchy. Different treatment effects were represented by different shifts in location. Underlying distributions for each treatment differed only in location, if at all. Various combinations of treatment effects were considered including when one treatment effect was different from the others, when two or more treatment effects were equal and different from the others, when there was equal spacing between the treatment effects, and when there was unequal spacing between the treatment effects [21-23].

Situations were also considered when the probability of a missing observation in a block ranged from 0.10 to 0.50 for each treatment, in increments of 0.10. Mixed designs in which the completely randomized portion was 1/4, 1/3, 1/2, 2/3, 1.0, 1.5, 2.0, 3.0, and 4.0 times the

randomized block portion were considered. The number of blocks and the number of observations for each sample in the completely randomized design portion ranged from 5 to 20.

Results

Selected results are given in Tables 1-4. Tables 1-4 give only a small portion of the findings, but are representative as to the overall findings of the study. Additional results may be obtained from the authors. The estimated significance levels were all around 0.05 for the cases that were studied. This is when the number of blocks ranged from 5 to 20 and the equal sample size for the CRD portion was also between 5 and 20. For the situations considered, Z_{first} had greater powers than Z_{last}. This was true for normal distributions, exponential distributions, t distributions with 3 degrees of freedom, and Cauchy distributions. It was true when the number of treatments ranged from 3 to 5. It was also true when the probability that an observation for a treatment was missing within a block ranged from 0.10 to 0.50. When the number of blocks was up to 4 times the equal number of observations in the CRD portion, Z_{first} had higher powers. Table 1 gives estimated powers for both tests for 3 treatments, underlying normal populations, 20 blocks with a probability of an observation being missing within a block for a given treatment being 0.10, and equal sample sizes of size 5 for the completely randomized design. When the treatment effects were 0.00, 0.40, and 0.80, the estimated power for Z_{first} was 0.6560, and the estimated power for Z_{last} was 0.5711. Z_{first} had higher powers then Z_{last} for all of the treatment effect combinations in Table 1. Table 2 gives estimated powers when the number of blocks is 1/2 the number of observations in the random sample of each treatment in the completely randomized design. It is noted that the estimated powers of Z_{first} are always larger than the estimated powers of Z_{last} for the treatment effects studied. Findings were similar when the number of blocks was 1/4 the equal sample size for the CRD portion.

Table 3 and Table 4 illustrate findings when the probability of a missing observation within a block for a particular treatment is 0.50. In both tables, the number of treatments was 5. The underlying distributions in Table 3 were Cauchy, and in Table 4 were normal. In Table 3, the sample size for the CRD portion was the same as the number of blocks. In Table 4, the sample size for the CRD portion was 1/2 the number of blocks. In all cases, Z_{first} has a higher power than Z_{last} when treatment effects differ.

Parameter 1	Parameter 2	Parameter 3	Z_{first}	Z_{last}
0.0	0.0	0.0	0.047	0.045
0.0	0.2	0.4	0.264	0.227
0.0	0.4	0.8	0.656	0.571
0.0	0.1	0.6	0.467	0.403
0.0	0.0	0.8	0.681	0.598
0.0	0.8	0.8	0.620	0.521

Table 1: Normal Distributions; 20 blocks; sample sizes for CRD=5; 10% missing.

Parameter 1	Parameter 2	Parameter 3	Parameter 4	Parameter 5	Z_{first}	Z_{last}
0.0	0.0	0.0	0.0	0.0	0.049	0.048
0.0	0.2	0.5	0.6	0.9	0.667	0.586
0.0	0.0	0.0	0.0	0.9	0.514	0.458
0.0	0.0	0.0	0.8	0.8	0.729	0.668
0.0	0.0	0.8	0.8	0.8	0.722	0.647
0.0	0.8	0.8	0.8	0.8	0.392	0.335

Table 2: T-Distribution with 3 degrees of freedom; 6 blocks; sample sizes for CRD=12; 30% missing.

Parameter 1	Parameter 2	Parameter 3	Parameter 4	Parameter 5	Z_{first}	Z_{last}
0.0	0.0	0.0	0.0	0.0	0.046	0.049
0.0	0.2	0.5	0.6	0.9	0.144	0.097
0.0	0.0	0.0	0.0	0.9	0.122	0.085
0.0	0.0	0.0	0.8	0.8	0.158	0.101
0.0	0.0	0.8	0.8	0.8	0.154	0.098
0.0	0.8	0.8	0.8	0.8	0.104	0.076

Table 3: Cauchy Distributions; k=5; 6 Blocks; sample sizes for CRD=6; 50% missing.

Parameter 1	Parameter 2	Parameter 3	Parameter 4	Parameter 5	Z_{first}	Z_{last}
0.0	0.0	0.0	0.0	0.0	0.051	0.052
0.0	0.2	0.5	0.6	0.9	0.791	0.725
0.0	0.0	0.0	0.0	0.9	0.654	0.596
0.0	0.0	0.0	0.8	0.8	0.860	0.813
0.0	0.0	0.8	0.8	0.8	0.852	0.800
0.0	0.8	0.8	0.8	0.8	0.488	0.415

Table 4: Normal Distributions; k=5; 12 Blocks; sample sizes for CRD=6; 50% missing.

Conclusion

Two nonparametric test statistics were introduced for the mixed design consisting of a completely randomized design portion and a randomized block portion with missing observations. The asymptotic distributions of the test statistics could be used when the number of blocks and the equal sample sizes for the CRD portion are as low as 5 since the estimated significance levels were all around 0.05 in these cases. The simulation study did not include any fewer than 5 blocks or sample sizes lower than 5 for the CRD portion. The test statistic, Z_{first} had higher estimated powers than Z_{last} for all situations considered in the simulation study. Z_{first} gave equal weights to both Z_{JT} and Z_{AC}, the standardized test statistics for the completely randomized and the randomized complete block design portions, respectively. gave more weight to Z_{AC} if the AC test statistic had a larger variance, and more weight to Z_{JT} if the JT test statistic had a larger variance. The study included cases in which the number of blocks was up to 4 times more than the sample sizes for the CRD portion and cases in which the number of blocks was as low as ¼ the sample size for the CRD portion. The results, as far as which test statistic had higher powers, did not depend on the underlying distributions studied.

References

1. Dubnicka SR, Blair RC, Hettmansperger TP (2002) Rank-Based Procedures for Mixed Paired and Two-Sample Designs. J Mod Appl Stat Methods 1: 32-41.

2. Daniel WW (2000) Applied Nonparametric Statistics. (2ndedn), Duxbury,USA.

3. Hollander M, Wolfe DA (1999) Nonparametric Statistical Methods. (2ndedn), John Wiley and Sons, New York, USA.

4. Mann HB, Whitney DR (1947) On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other. Annals of Mathematical statistics 18: 50-60.

5. Kruskal WH, Wallis WA (1953) Use of Ranks in One-Criterion Variance Analysis. J Am Stat Assoc 47: 583-621.

6. Jonckheere AR (1954) A Distribution-Free k-Sample Test against Ordered Alternatives. Biometrika 41: 133 - 145.

7. Terpstra TJ (1952) The Asymptotic Normality and Consistency of Kendall's Test against Trend,When Ties Are Present in One Ranking. Indag Math 14: 327-333.

8. Magel R, Terpstra J, Shams S (2013) A New Nonparametric Test for the Umbrella Alternative. Journal of Statistics and Management Systems.

9. Terpstra JT, Chang CH, Magel RC (2011) On the use of Spearman's

correlationcoefficient for testing ordered alternatives. J Stat Comput Simul 81: 1381-1392.

10. Wilcoxon F (1945) Individual Comparisons by Ranking Methods. Biometrics Bulletin 1: 80-83.

11. Friedman M (1937) The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. J Am Stat Assoc 32: 675-701.

12. Friedman M (1940) A Comparison of Alternative Tests of Significance for the Problem of m Rankings. Annals of Mathematical statistics 11: 86-92.

13. Page EB (1963) Ordered Hypotheses for Multiple Treatments: A Significance Test for linear Ranks. J Am Stat Assoc 58: 216-230.

14. Durbin J (1951) Incomplete Blocks in Ranking Experiments. Br J Clin Psychol 4: 85-90.

15. Cao L, Magel R, Cao L, Ndungu A (2012) Comparing the Durbin, Wilcoxon Signed-RankTest, and a Proposed Test in a Balanced Incomplete Block Design. The International Journal of Science in Society 3: 1-16.

16. Alvo M, Cabilio P (1995) Testing Ordered Alternatives in the Presence of Incomplete Data. J Am Stat Assoc 90: 1015-1024.

17. Magel R, Fu R (2013) A Proposed Test for a Mixed Two-Sample Design.

18. Magel R, Terpstra J, Canonizado K, Park JI (2010) Nonparametric Tests for Mixed Designs. Commun Stat Simul Comput 39: 1228-1250.

19. Magel R, Terpstra J, Wen J (2009) Proposed Tests for the Nondecreasing Alternative in a Mixed Design. Journal of Statistics and Management Systems 12: 963-977.

20. Bailer AJ (2010) Statistical programming in SAS. (1stedn), SAS Institute, NC, USA.

21. Hemmer MT, Magel R (2012) Nonparametric Test for the Umbrella Alternative in the Randomized Complete Block and Balanced Incomplete Block Mixed Design. North Dakota State University.

22. Kim DH, Kim YC (1992) Distribution-Free Tests for Umbrella Alternatives in a Randomized Block Design. Journal of Nonparametric Statistics 1: 277-285.

23. Mathisen D, Magel R (2011) A Comparison of Nonparametric Test Statistics for Nondecreasing Treatment Effects Over a Mixed Model Design, MS. North Dakota State University.