A SAS Macro to Extend the Friedman Test for the Unbalanced Case
Plus a Multiple Comparisons Procedure

Kirk A. Easley, The Cleveland Clinic Foundation

ABSTRACT

A SAS Macro utilizing the method of m rankings (or within block rankings) generates the Benard-van Elteren test statistic which is an extension of the Friedman test that permits empty cells and any number of observations per cell. A multiple comparisons procedure for pairwise comparisons of treatments is included for use in conjunction with the Benard-van Elteren test. Additional applications of this Macro include the Kruskal-Wallis test and the model.

INTRODUCTION AND PURPOSE

One of the most commonly used statistical design structures is the randomized complete block design (RCBD). The RCBD is any blocking scheme in which the number of experimental units within a block is a multiple of the number of treatments and thus a complete set of treatments can be assigned completely at random to c experimental units in each block (Milliken and Johnson, 1984). A proper randomization scheme distinguishes experimental from observational studies and is a fundamental assumption underlying parametric and nonparametric analysis of variance techniques for the RCBD. However, observational studies are quite common and data arising from such studies can be analyzed using statistical analysis techniques appropriate for the RCBD. The interpretation of results from these analyses may require greater caution because of the lack of randomization.

Furthermore, parametric analysis-of-variance techniques for the RCBD are justified when the assumptions of independence, normality, equal variance and additivity hold. Fortunately the robust nature of parametric ANOVA techniques still provide reliable results with minor departures from the model assumptions. If the underlying assumptions are severely violated or if interaction effects are present, transformations on the data can be attempted to alleviate the problem. Berry (1987) proposes a logarithmic transformation whereby an additive constant c is chosen when transforming data x to y = log(x + c). Conover and Iman (1981) suggest the use of rank transformation procedures whereby the parametric procedure is applied to the ranks of the data instead of the data themselves. However, nonparametric distribution-free analysis strategies including multiple comparisons for randomized block studies are still desirable and appropriate.

Another difficulty with the RCBD is the occurrence of one or more missing observations. Missing data are commonly encountered in many disciplines but present no difficulty when using a regression approach to the RCBD. The technique simply requires setting up the regression model for the observations available and fitting the model to the data. The nonparametric analog to the RCBD analysis of variance is the Benard-van Elteren test (Benard and van Elteren, 1953). The Benard-van Elteren test is an extension of the Friedman test (1937) which utilizes the method of m rankings (or within block rankings) and allows any cell to be empty or to contain any number of observations. Recent applications of the Benard-van Elteren test have been reported by Norwood et al. (1989) and Eiseck and Uthoff (1989).

Additionally, Norwood et al. (1989) have developed a multiple comparisons procedure to be used in conjunction with the Benard-van Elteren test. The purpose of this project was to develop and benchmark a SAS macro using the methodology of Benard and van Elteren (1953) and Norwood et al. (1989) to generate the Benard-van Elteren test statistic for the RCBD and the associated multiple comparisons. The macro allows missing data as well as multiple observations per cell. The Supplemental Library MRANK procedure not available under Version 6 of the SAS System and PROC FREQ (using CMH2 SCORES=RANK as options on the TABLES statement) also provide the Benard-van Elteren test statistic but not the multiple comparisons procedure of Norwood et al (1989).

MACRO CALL:

%BENARD(TRT, ID, RESPONSE, DATASET, MULTCOMP=O)

The Macro BENARD generates the Benard-van Elteren (BVE) test statistic for the RCBD and multiple comparisons using the Tukey-Kramer method or a Bonferroni procedure for up to five levels of a treatment factor. The BVE statistic has an asymptotic Chi-Square distribution with t-1 degrees of freedom (Benard and van Elteren, 1953) where t is the number of treatments. The multiple comparisons procedure is a large-sample test using average centered ranks, and the normal range statistic for pairwise comparisons (Norwood et al., 1989). Both multiple comparisons procedures control the experimentwise type I error rate (Hayter, 1984; D'Agostino and Heeren, 1991).

Many applications of the RCBD involve ties and incomplete or missing data. Incomplete data may be related to the underlying data values and not simply missing at random. The occurrence of ties and disproportionate incomplete data in the RCBD can violate the underlying assumption of the multiple comparisons procedure which requires the variances of the pairwise differences of the average centered ranks to be equal, in order to ensure the procedure is asymptotically correct. Implementation of the procedure with missing data should include an evaluation of these variances to ensure an appropriate application. The variances of the differences of the mean centered ranks are always printed by the macro BENARD. Norwood et al. (1989) indicate the multiple comparisons procedure is most likely conservative when the number of treatments is three or four regardless of the form of the variances and covariances of the treatment mean centered ranks. Note that each block is assumed to have at least two observed treatments, since blocks with a single treatment do not provide information about treatment comparisons or change any of the statistics. SAS/IML® is necessary for more than three levels of treatment.

In the case of the balanced RCBD with p >1 replicates per treatment per block, the variances of the differences are equal (see Examples two and four). Tied observations within a block do...
not affect the equality of variances of the differences since the effect of ties is a proportional reduction of variances and covariances for all treatments in that particular block.

PARAMETER DEFINITIONS:

TRT is a macro variable identifying the treatment variable or factor contained in a SAS data set. The treatment factor must be numeric and the value of the variable must be an integer between one and five.

ID is a macro variable identifying the blocking variable within a SAS data set. The blocking variable must be numeric and the value of the variable must be an integer.

RESPONSE is a macro variable which identifies the measured response or outcome variable in a SAS data set. The outcome must be a numeric variable.

DATASET is a macro variable naming the SAS data set containing the variables to be used in the analysis. These four positional parameters must be given values with each macro invocation.

MULTCOMP is a keyword macro variable indicating whether or not a Bonferroni procedure (MULTCOMP=1) is to be used for all pairwise comparisons. The default value is 0 which uses the Tukey-Kramer method outlined by Norwood et al. (1989).

OUTPUT FROM BENARD:

Two temporary SAS data sets are output from BENARD. The first data set is called BVE and contains the following SAS variables which are displayed using PROC PRINT.

```
BVE STAT  -The Benard-van Elteren test statistic.
PVVALUE  -The significance level for the Benard-van Elteren test statistic. This is provided by
           using the PROBCHI function.
```

The second SAS data set is called MC for multiple comparisons and contains the following SAS variables when the number of treatments are three. These results are also displayed using PROC PRINT. Naming convention follows that used by Norwood et al. (1989).

```
VARUB12  -The variance of the differences of the mean-centered ranks for treatments one and two
          (VARUB12), treatments one and three (VARUB13) and treatments two and three
          (VARUB23). These variances will be the same for the balanced case.
D12, D13, D23 -The resulting statistics from the multiple comparisons among treatments. The .05 and .01 critical values were obtained from the
                 studentized range critical values with df and are displayed with the TITLE statement following PROC PRINT (Miller, 1981, page 177).
```

EXAMPLES

Four examples have been selected to demonstrate the accuracy and output from the macro BENARD. The first example is taken from Norwood et al. (1989). A clinical trial was designed to evaluate three treatments (suprofen at 400 and 200 mg placebo) for primary dysmenorrhea. Each woman was to use a different treatment over three consecutive menstrual cycles. Eighty-six women completed all three cycles and 12 completed only two cycles resulting in missing data. After using a treatment during a menstrual cycle, each woman rated that medication with regard to degree of relief (none, minimal, moderate or complete). Degree of relief responses were rank ordered as detailed by Norwood et al. 1989.

The Benard-van Elteren test statistic and associated p-value, the variances of the differences of the mean centered ranks and the resulting statistics from the Tukey-Kramer multiple comparisons procedure are provided in the appendix. There is a highly significant difference among treatments with both dosages of suprofen providing greater relief than placebo. No differences were detected in the 200 and 400 mg dosages of suprofen. Notice the variances of the differences are almost identical. The multiple comparisons procedure controls the experimentwise type I error rate and is the recommended procedure for pairwise comparisons by D'Agostino and Heeren (1991).

Example two uses data presented by Elswick and Uthoff (1989). Twelve patients with gastroesophageal reflux disease were enrolled in a two treatment, two period, four-sequence crossover trial. Each patient was randomly assigned to one of four sequences: AB, BA, AA or BB where A and B represent metoclopramide and placebo respectively. The objective was to determine whether metoclopramide helped to prevent acid reflux. The number of reflux episodes occurring within each period of observation was recorded. After forming differences within each patient, the Benard-van Elteren test statistic can be used with the transformed data to construct a test of the equality of the direct effects of the treatment (Elswick and Uthoff, 1989). The validity of a test for the treatment effect in the four-sequence model does not depend on equal carryover effects. The resulting Benard-van Elteren test statistic was 4.67 (p=0.031; 1 df) suggesting the metoclopramide group had fewer reflux episodes than the placebo group. See the paper by Elswick and Uthoff (1989) for additional detail.

Example three uses the contrived data reported by Hutchinson (1977). Six professional truck drivers were asked to rank five makes of lorry cabs as to which they preferred driving. Not all drivers had experience driving all makes so some observations were missing. The Benard-van Elteren test statistic was 9.96 (p=0.04; 4 df) suggesting a difference between lorry cabs regarding degree of driving comfort. Additional caution is necessary when the number of blocks is small since the sampling distribution under the null hypothesis for the test statistic is approximately a chi-square (k-1 degrees of freedom) with reasonably large sample sizes. A clear presentation of the computational detail for calculating the variances and covariances of the treatment centered rank sums was provided by Hutchinson (1977).

Example four uses two data sets from Woolson (1987) to benchmark the use of the BENARD macro for the Kruskal-Wallis test plus all pairwise comparisons and the Friedman test plus all pairwise comparisons. For the Kruskal-Wallis example, twenty-
nine pigeons were randomized to one of four groups: amphetamine (n=8); amphetamine plus low dose of chlorpromazine (n=7); amphetamine plus middle dose of chlorpromazine (n=5); amphetamine plus high dose of chlorpromazine (n=9). The information recorded for each pigeon was number of pecks per unit of time. This measurement was thought to reflect behavioral effects of chlorpromazine on amphetamine treated pigeons. The question of interest was whether the data provided evidence to conclude that the four groups differ. To use the BENARD macro, one must declare a blocking factor set equal to one for each pigeon (i.e., experimental unit). This is necessary since the Kruskal-Wallis test is a special case of the Benard-van Elteren test where the number of blocks is one. The test statistic is identical to that reported in Woolson (1987, page 369), resulting in insufficient evidence to reject the null hypothesis of no group differences. Similarly the Bonferroni procedure for all pairwise comparisons provides identical results to those reported by Woolson. The Bonferroni procedure is activated at macro invocation by setting the keyword parameter MULTCOMP equal to one. Woolson (1987) notes the error rate will be no greater than $\alpha$ for the entire set of comparisons jointly. Furthermore, the procedure tends to be overly conservative in that usually the error rate will be much less than $\alpha$.

For illustration of the Friedman test, sixteen boys grouped on the basis of age, length of illness and severity of illness were used in a comparative study of four methods of treating hyperactivity. The four treatments were randomly assigned to one of the four boys in each block. The boys were rated by their school teachers using a structured schedule for assessment of hyperactivity. The intrablock ranks of hyperactivity scores were used to compute the Benard-van Elteren test statistic of 10.42 (p = .015, 3 df) which is identical to that reported by Woolson (1987, page 377). The Bonferroni multiple comparisons procedure found none of the pairs to be significantly different. Results from the Tukey-Kramer method were similar to the Bonferroni except somewhat less conservative.

The SAS code and output for examples two, three and four are available from the author at the following address:

Kirk A. Easley
Department of Biostatistics and Epidemiology
Cleveland Clinic Foundation
9500 Euclid Avenue
Cleveland, Ohio 44195-5196

The preferential mode of file transfer is IBM-PC ASCII files on 3.5\* (720K or 1.44M) or 5.25\* (360K or 1.2M) diskette.

REFERENCES


APPENDIX

* EXAMPLE ONE Taken from NORWOOD ET AL, 1989 "A Multiple Comparisons Procedure For Use in Conjunction with the Benard·van Elteren Test" (BIOMETRICS 45, p. 1175-1182).

```plaintext
DATA NORWOOD;
  INPUT ID TRT RESPONSE @@;
  * trt = 1= s400, trt = 2= s200 and trt = 3 = pl;
cards;
1 1 3 1 2 2 1 3 1
2 1 3 2 2 2 3 1
3 1 3 3 2 3 3 3 1
4 1 3 4 2 2 4 3 1
5 1 3 5 2 2 5 3 1
etc.
%INCLUDE 'C:\BENARD\BENARD.MAC';
%BENARD(TRT,ID,RESPONSE,NORWOOD, MULTCOMP=0)
```

OUTPUT FOR EXAMPLE 1

```
BENARD-van ELTEREN Test Statistic P value from the CHI-SQUARE distribution
61.4448 4.5408E-14
```

VARIANCES of the DIFFERENCES of the MEAN CENTERED RANKS (MCR)
These variances should be equal or nearly equal

```
VARIANCE (MCR1-MCR2) VARIANCE (MCR1-MCR3) VARIANCE (MCR2-MCR3)
0.016358 0.016182 0.016187
```

RESULTING STATISTICS FROM MULTIPLE COMPARISONS AMONG GROUPS
0.05 and 0.01 critical values of 3.314 and 4.120
These Critical Values Control the Type 1 Experimentwise Error Rate using a Tukey-Kramer Procedure (see Miller, 1981, p. 177).

```
D1,2 for D1,3 for D2,3 for
group group group
1 vs 2 1 vs 3 2 vs 3
2.45003 10.5782 8.11371
```

1139