A general problem of statistical interest is the formulation of tests for the hypothesis that one or more response variables are distributed at random with respect to a set of sub-populations (e.g., treatments) within one or more strata based upon control variables (e.g., blocks or clinics). More specifically, if $y_{ijh}$ denotes the observed value for the $j$-th response variable for the $i$-th subject in the $h$-th stratum where $h=1,2,...,q$, $i=1,2,...,n_h$, and $j=1,2,...,r_h$, then this type of hypothesis can be expressed as

$$H_0: \text{null hypotheses for each stratum, } h=1,2,...,q.$$ 

From the finite population randomization model implied by $H_0$, several different types of well-known statistical tests can be formulated without any external assumptions concerning underlying distributions. These include univariate and multivariate rank analysis of variance statistics, rank analysis of covariance statistics, Spearman rank correlation test statistics, and contingency table chi-square test statistics. All of these methods together with certain more general counterparts are discussed in terms of a common underlying methodology. This conceptual framework is then used as the rationale for a SAS Macro for the straightforward computation of these types of non-parametric statistical tests via a standard set of operations.

### Abstract

A general problem of statistical interest is the formulation of tests for the hypothesis that one or more response variables are distributed at random with respect to a set of sub-populations (e.g., treatments) within one or more strata based upon control variables (e.g., blocks or clinics). More specifically, if $y_{ijh}$ denotes the observed value for the $j$-th response variable for the $i$-th subject in the $h$-th stratum where $h=1,2,...,q$, $i=1,2,...,n_h$, and $j=1,2,...,r_h$, then this type of hypothesis can be expressed as

$$H_0: \text{null hypotheses for each stratum, } h=1,2,...,q.$$ 

From the finite population randomization model implied by $H_0$, several different types of well-known statistical tests can be formulated without any external assumptions concerning underlying distributions. These include univariate and multivariate rank analysis of variance statistics, rank analysis of covariance statistics, Spearman rank correlation test statistics, and contingency table chi-square test statistics. All of these methods together with certain more general counterparts are discussed in terms of a common underlying methodology. This conceptual framework is then used as the rationale for a SAS Macro for the straightforward computation of these types of non-parametric statistical tests via a standard set of operations.

## Observational data examples:

1. Autopsied accident studies concerned with the association between driver injury and restraint system usage. Here, strata correspond to no usage, lap belts only, and lap and shoulder belt and the outcome measure corresponds to the extent of injury.
2. Criminal justice studies concerned with the association between court dispositions and type of legal assistance for defendants. Here, strata correspond to such factors as prior criminal record, offense severity, employment status, etc.; sub-populations correspond to no attorney, court-appointed attorney, and private attorney and disposition corresponds to duration of prison sentence (if any).
3. Epidemiological studies concerned with the association between current health status and certain exposure risk factors for their personal behavior, home environment, or work environment. Here, strata correspond to demographic variables, medical history variables, etc.

## Experimental data examples:

4. Dairy cows are randomly assigned to several methods of treatment for mastitis. Data pertaining to the disease status of each quarter of the udder of each cow are obtained both before and after treatment. Here, cows represent the randomized experimental unit.
5. A multi-center clinical trial is undertaken to compare two treatments. For this purpose, a set of appropriately qualified clinics are invited to participate on a judgmental basis. Within each, patients are randomly assigned to two treatments. Data pertaining to various aspects of the medical status of each patient are recorded both before treatment (e.g., baseline) and at weekly visits during treatment. Here, clinics correspond to the strata.

6. A change-over clinical trial is undertaken to compare two treatments. For this purpose, patients are randomly partitioned into sequence groups which define the order according to which they are given the two treatments during successive periods of time; i.e., some patients receive the new treatment first followed by the control treatment while others receive the control treatment first followed by the new treatment. Here, patients represent a randomized experimental unit with respect to sequence group.

Certain randomization chi-square test statistics can be obtained from the GLM and ANOVA procedures of SAS. These include the within stratum one-way analysis of variance and covariance for a univariate response profile. A one-way analysis of covariance is specified by:

```
PROC GLM;
MODEL RESPONSE = COVAR.
OUTPUT RESIDUAL = RESID OUT = PROCGLM;
PROC GLM; CLASSES SUBPOPS;
MODEL RESID = SUBPOPS;
```

Then:

```
Q(Response ICovARiable under H_0) = (n-1)^*SSB/FST,
where n = number of observations in stratum
SSB = subpopulation sum of squares
FST = total sum of squares
```

This represents a within-stratum analysis which can be repeated for each stratum. It can then be followed by the across-strata two way analysis of covariance procedure, thereby completing the randomization hypothesis testing analysis. Although the previously described univariate chi-square statistic can reflect an adjustment for more than one covariable, the evaluation of its underlying assumption of randomness for the multivariate covariable profile cannot be directly undertaken with PROC GLM, but is available indirectly through multiplication of the Pillai trace criterion by (n-1). Other programs producing randomization chi-square statistics include SPLOTA which is concerned with split plot analyses for univariate and multivariate responses without covariable adjustment and PARCAT which produces generalized Mantel Haenszel statistics without covariable adjustment for q x s x (d+1) contingency tables where q = number of tables (strata), s = number of subpopulations, (d+1) = number of response categories. As an alternative to this existing software, a more general randomization procedure called GRMM has been developed as a SAS MACRO which is accessible through PROC MATRIX.

2. General Randomization Model Macro (GRMM)

GRMM produces multivariate one-way and two-way analysis of variance and covariance, non-parametric chi-square statistics as outlined in Appendix A. Because GRMM operates on a stratum by stratum basis, different covariables may be entered for each stratum. In this situation, the across strata statistic of interest is the average stratum adjusted multivariate chi-square statistic for the responses. A practical use for the by-stratum covariable specification is the definition of missing value indicators as discussed in Appendix A.5.

Because missing value indicators represent a special use of covariables, GRMM allows the user the option of distinguishing them from other covariables as indicated below. The user specified matrices are as follows:

```
Required:
_ID_ for stratum and sub-population identification is an n x 2 matrix where n is the total number of observations,
_VAR_ for response variables is an n x d matrix where d is the number of response variables,
```

Optional:
```
_COVAR_ for covariables is an n x c matrix where c is the number of covariables,
_MISS_ for missing value indicators is an n x m matrix where m is the number of missing value indicators,
_COVARH_ for stratum partition map of the covariables is an q x c matrix where q is the number of strata,
_MISSH_ for stratum partition map of the missing value indicators is an q x m matrix,
_OPTION_ for output options which have default values.
```

To invoke the macro, the user needs to initially compile the macro within PROC MATRIX by the statement GRMM; subsequently GRMM may be executed a number of times by specifying the statement LINK ANALYSES; after each set of user specified matrices. Furthermore, matrices that do not change across sets of analyses need not be respecified for them.

The examples in the next section illustrate the GRMM output.

3. Examples

In this section, univariate and multivariate applications of randomization test statistics and the GRMM SAS macro are illustrated for selected aspects of data from a multi-center clinical trial concerned with the effectiveness of a drug for alleviating a certain skin condition. These data (which are fully presented in Stanish, Gillings, and Koch [1978]) represent the experience
of patients randomly assigned to drug or placebo treatments within each of six clinics. Each patient has an initial severity evaluation (coded: fair, poor or exacerbation) of the skin condition in addition to improvement evaluations (coded: rapidly improving, slowly improving, stable, slowly worsening, or rapidly worsening) for at least one of three follow-up visits over time. Because some patients are missing improvement evaluations for one or two visits, a missing value adjustment strategy is used for multivariate analyses via three missing value indicators as concomitant variables.

For these data, the question of interest is whether there is no difference between the two treatments for extent of improvement within each clinic and averaged across clinics. Moreover, its scope included the respective visits in both a separate univariate sense, as well as all of them simultaneously in a multivariate sense, optionally relative to the initial severity of the skin condition and/or the missing value indicators. For this purpose, a SAS dataset named CLINIC is constructed to contain the original data for the six clinics with the substitution of imputed values for the missing improvement evaluations and the addition of corresponding missing value indicators which have the value 0 if the data are observed and 1 if the data are imputed. CLINIC is then sorted to ensure the ordering of all patients within a specific treatment group within a given clinic. Because the univariate and multivariate examples involve an analysis based on ranks, CLINIC undergoes a transformation of the values of the three improvement evaluations and the initial severity evaluation with the corresponding within-clinic ranks. Finally, a number of rank analyses of variance/covariance can now be specified for CLINIC within PROC MATRIX by linking the macro with the appropriate data matrices for each analysis. The following univariate and multivariate examples are specified by the program statements shown in Figure 3.1. The output only shows results from two clinics because of space limitations.

3.2 A univariate analysis of covariance for rank data

This example is concerned with the following hypothesis:

$H_0$: There is no relationship between the two treatments and the degree of improvement at the third evaluation after adjusting for the initial severity of the skin condition in accordance with the underlying randomization model.

It uses selected columns of the data matrix, HODATA, representing 172 patients from 6 clinics in Figure 3.1, reassigned to the following matrices:

$$\text{VAR}_1 = T \quad \text{VAR}_2 = \text{HODATA}(*,1,3)$$

where $T$ is the treatment number, the STRATUM Identifier, and T1 is the treatment number, the SUBPOP identifier.

where $F_3$ is the within-STRATUM rank of the improvement evaluation at visit 3 (redefined to be the Final Visit by replacement of missing values by within subject, nearest adjacent observed values), a RESPONSE VARIABLE.

$$\text{COVAR}_1 = \text{STAGE} = \text{HODATA}(*,2)$$

where STAGE is the within-STRATUM rank of the initial severity evaluation of the skin condition, a COVAR variable.

Following the assignment of the 3 matrices: _ID_, _VAR_ and _COVAR_, shown in Figure 3.1, the GRMM macro is compiled by the statement **GRMM** and linked by the statement **LINK ANALYSIS**. Because elements of the **OPTION** matrix were not reassigned before the **LINK** statement, **GRMM** produces the default list of output, some of which is shown in Figure 3.2.

Evaluation of the **GRMM** output Figure 3.2 suggests that the patterns of ranks for the subpopulation and overall stratum means for the subpopulations (treatments) is similar for both strata (clinics). More specifically, the ranks of the response variable (the third improvement evaluation) are consistently lower (indicating a higher degree of improvement) for treatment 1 (the test drug) while the ranks of the covariable (the initial severity evaluation of the skin condition) vary about a common value (indicating similar initial severities of the skin condition for patients receiving either treatment). Furthermore, the extent to which the sums of the ranks deviate from the expected sums of ranks (cf. subpopulation residual sums of variables table) based on the sample sizes varies about 40% of the expected sum (not shown) for the response variable but only varies about 7% for the covariable. To assess the significance of the observed rank variation, the variance-covariance matrix _VH_ together with the among subpopulation sums of cross-products matrix _SH_ are used to obtain the stratum partial association test statistics of the form (A12) where _QHF_ is the multivariate Kruskal-Wallis test statistic for the joint response variable and covariable,

$QHY$ is the univariate Kruskal-Wallis test statistic for the response variable,

$QHX$ is the univariate Kruskal-Wallis test statistic for the covariable,

$QHYRX$ is the reduction of the multivariate Kruskal-Wallis test statistic by the covariable and represents a univariate response statistic adjusted for covariable differences of form (J11),

$QHG$ is the linear model statistic of form (A12) and represents a univariate response statistic adjusted for covariable differences.

$QHYRX$ and $QHG$ have the same value since both represent a within-stratum adjustment of the re-
response for the covariable. These also yield the same results as the following sequence of SAS analysis of variance statements executed in place of the PROC MATRIX statements in Figure 3.1:

```
PROC GLM DATA=CLINIC;
   MODEL R3=STAGE;
   OUTPUT OUT=PROCGLM RESIDUAL=R3RESIDS;
PROC GLM DATA=PROCGLM;
   CLASSES TRT;
   MODEL R3RESIDS-TRT;
```

For STRATUM 5, the 37 observations have a multiple correlation coefficient of 0.3782 for the second model which is asymptotically equal to QHG/(n-1) or alternatively
\[ QHG = (36)(0.3782) = 13.6158. \]

Due to the non-significance of QHX, the use of QHG and QGRBG is statistically valid in the sense of the Q(y|x under H0) statistic in (A10) where H0 induces the randomization model on x.

The across strata average partial association statistics of the form (A17) are a function of the stratum matrices _VH_ and the residual sums of variables matrices _RSUMS_ , where

- QGRBF is the multivariate generalized randomized block statistic for response variable and covariable,
- QGRBY is the univariate generalized randomized block statistic for the response variable,
- QGRBX is the univariate generalized randomized block statistic for the covariable,
- QYRX is the reduction of the multivariate generalized randomized block statistic to a univariate generalized randomized block statistic for the response variable adjusted for the covariable,
- QGRBG is the linear model univariate generalized randomized block statistic for the response variable adjusted for the covariable.

Because QGRBG maintains the within stratum residual structure, it differs from QGRX which has an across strata residual structure. Furthermore, both QYRX and QGRBG here are valid applications of the average of the Q(y|x under H0) statistic since QGRBX and the stratum set of QYRX statistics are all non-significant. Thus, it can be concluded for the two clinics that the null hypothesis of no treatment effects at the third improvement evaluation after adjusting for the initial severity of skin condition is contradicted under the randomization model in the sense that the test drug is significantly (p<0.001) more effective in improving the skin condition. This result also holds for all six clinics (QGRBG=64.7, p=0.0001).

### 3.3 Multivariate analysis of covariance for rank data

This example is concerned with the follow-
variables table, however, shows the residual sums of ranks for the observed data only. From this table, it can be seen that the pattern of ranks for the two treatment groups (SUBPOP) is the same for each clinic (STRATUM).

Because each clinic (STRATUM) has a different set of missing value indicators (MISS IND), the only reasonable across strata statistic is the analysis of covariance statistic, QGREG. This result indicates that the randomization model hypothesis $H_0$ is contradicted in the sense of significantly ($p < 0.01$) more favorable improvement for the test drug. Similar results are obtained across all six clinics ($QGREG=58.8, p = 0.0001$).

REFERENCES


Forthofer, R. and Koch, G. SPLOTA (program documentation), Department of Biostatistics, University of North Carolina.


ACKNOWLEDGMENTS

This research was supported in part by the U.S. Bureau of the Census through Joint Statistical Agreement JSA 79-16. The authors would like to thank Ronald W. Helms for helpful comments with respect to the development of the macro discussed in this paper. They would also like to express their appreciation to John DeGraffenreidt and Jean Harrison for their conscientious typing of the manuscript.
PROC TRIANl VERSION OF SAS DATASEI, CLINIC;
DATA CLINIC;
INFILE CARDS;
INPUT @1 STAGE 1.
@6 TRT 1.
@8 R1 1.
@10 R2 1.
@12 R3 1.

LABEL STAGE = 'INITIAL SEVERITY STAGE OF SKIN CONDITION';
LABEL TRT = 'TREATMENT GROUP (1 IF TEST, 2 IF PLACEBO)';
LABEL R1 = 'IMPROVEMENT EVALUATION AT 1ST TIME POINT';
LABEL R2 = 'IMPROVEMENT EVALUATION AT 2ND TIME POINT';
LABEL R3 = 'IMPROVEMENT EVALUATION AT 3RD TIME POINT';
LABEL M1 = 'MISSING INDICATOR (1 IF MISS, 0 NOT)';
LABEL M2 = 'MISSING INDICATOR (1 IF MISS, 0 NOT)';
LABEL M3 = 'MISSING INDICATOR (1 IF MISS, 0 NOT)';
LABEL F3 = 'MISSING INDICATOR (1 IF MISS, 0 NOT)';

M1=0; M2=0; M3=0; F3=0.
IF R2=. THEN DO;
   E2=MAX(R1,R3);
   R2=0; M2=1;
   END;
IF R3=. THEN DO;
   F3=R2;
   R3=0; M3=1;
   END;
IF R1=. THEN DO;
   FI=R2;
   R1=0; M1=1;
   END;
CARDS;

PROC SORT DATA=CLINIC;
BY INV;
PROC RANK TIES=MEAN OUT=PROCRRANK DATA=CLINIC;
BY INV;
VAR STAGE R1 R2 R3; F3;
PROC MATRIX;
FETCH HODATA DATA=PROCRRANK COLNAME=VARNAME;
PRINT PROCRRANK VERSION OF SAS DATASET CLINIC;
PRINT HODATA COLNAME=VARNAME;
COMPILE GRAM MACRO;
GRAM;
*UNIVARIATE EXAMPLE SPECIFICATIONS;
   ID = HODATA(*,1);
   VAR = HODATA(*,10);
   COVAR = HODATA(*,2);
*EXECUTE ANALYSIS;
LINK ANALYSIS;
*MULTIVARIATE EXAMPLE SPECIFICATIONS;
   MISS = HODATA(*,6);
   MISS = HODATA(*,7);
*NOTE THAT ID AND COVAR NEED NOT BE RESPECIFIED;
*EXECUTE ANALYSIS;
LINK ANALYSIS;
For each clinic (STRATUM) there is no relationship between the 2 treatment groups (SUBPOP) and degree of improvement at final visit (RESPONSE) after adjusting for initial stage of skin condition (COVAR) in the sense that the observed partition of the improvement scores into treatment groups can be regarded as equivalent to a successive set of random samples of \( n_1 \) and \( n_2 \) patients.

### SUBPOPULATION AND OVERALL STRATUM (0,0) MEANS

<table>
<thead>
<tr>
<th>MEANS</th>
<th>STRATUM</th>
<th>SUBPOP</th>
<th>N</th>
<th>RESPONSE</th>
<th>COVAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROW1</td>
<td>5</td>
<td>1</td>
<td>19</td>
<td>12.9</td>
<td>20.1</td>
</tr>
<tr>
<td>ROW2</td>
<td>5</td>
<td>2</td>
<td>18</td>
<td>25.4</td>
<td>17.8</td>
</tr>
<tr>
<td>ROW3</td>
<td>0</td>
<td>0</td>
<td>37</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>ROW1</td>
<td>6</td>
<td>1</td>
<td>17</td>
<td>12.2</td>
<td>16.2</td>
</tr>
<tr>
<td>ROW2</td>
<td>6</td>
<td>2</td>
<td>16</td>
<td>22.1</td>
<td>17.8</td>
</tr>
<tr>
<td>ROW3</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td>17</td>
<td>17</td>
</tr>
</tbody>
</table>

### WITHIN STRATUM VARIANCE - COVARIANCE

<table>
<thead>
<tr>
<th>STRATUM</th>
<th>RESPONSE</th>
<th>COVAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROW1</td>
<td>103.7</td>
<td>-0.676</td>
</tr>
<tr>
<td>ROW2</td>
<td>-0.676</td>
<td>89.9</td>
</tr>
</tbody>
</table>

### SUBPOPULATION RESIDUAL SUMS OF VARIABLES AMONG SUBPOPULATIONS SUMS OF CROSS PRODUCTS

<table>
<thead>
<tr>
<th>RSUMS</th>
<th>STRATUM</th>
<th>SUBPOP</th>
<th>N</th>
<th>RESPONSE</th>
<th>COVAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROW1</td>
<td>5</td>
<td>1</td>
<td>19</td>
<td>-116</td>
<td>21.5</td>
</tr>
<tr>
<td>ROW2</td>
<td>5</td>
<td>2</td>
<td>18</td>
<td>116</td>
<td>-21.5</td>
</tr>
<tr>
<td>ROW1</td>
<td>6</td>
<td>1</td>
<td>17</td>
<td>-81.5</td>
<td>-13</td>
</tr>
<tr>
<td>ROW2</td>
<td>6</td>
<td>2</td>
<td>16</td>
<td>81.5</td>
<td>13</td>
</tr>
</tbody>
</table>

### STRATUM PARTIAL ASSOCIATION STATISTICS

<table>
<thead>
<tr>
<th>QSTAT_</th>
<th>STRATUM</th>
<th>Q_STAT</th>
<th>DF</th>
<th>P_VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multivariate Kruskal-Wallis:</td>
<td>Q(R3 STAGE)</td>
<td>QHF</td>
<td>5</td>
<td>24.1572</td>
</tr>
<tr>
<td>Univariate Kruskal-Wallis:</td>
<td>Q(R3)</td>
<td>QHY</td>
<td>5</td>
<td>13.6831</td>
</tr>
<tr>
<td>Univariate Kruskal-Wallis:</td>
<td>Q(STAGE)</td>
<td>QNX</td>
<td>5</td>
<td>0.5414</td>
</tr>
<tr>
<td>Reduced Kruskal-Wallis:</td>
<td>Q(R3 STAGE under ( H_0 ))</td>
<td>QHYRX</td>
<td>5</td>
<td>13.6158</td>
</tr>
<tr>
<td>Analysis of Covariance:</td>
<td>Q(R3)</td>
<td>QHG</td>
<td>5</td>
<td>13.6158</td>
</tr>
</tbody>
</table>

### AVERAGE PARTIAL ASSOCIATION STATISTICS

<table>
<thead>
<tr>
<th>QSTAT_</th>
<th>APA_STAT</th>
<th>DF</th>
<th>P_VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multivariate Friedman:</td>
<td>Q(R3)</td>
<td>QGRHF</td>
<td>23.0425</td>
</tr>
<tr>
<td>Univariate Friedman:</td>
<td>Q(R3)</td>
<td>QGRHY</td>
<td>22.7062</td>
</tr>
<tr>
<td>Univariate Friedman:</td>
<td>Q(STAGE)</td>
<td>QGRNX</td>
<td>0.0533</td>
</tr>
<tr>
<td>Reduced Friedman:</td>
<td>Q(R3 STAGE under ( H_0 ))</td>
<td>QRYRX</td>
<td>22.9892</td>
</tr>
<tr>
<td>Analysis of Covariance:</td>
<td>Q(R3)</td>
<td>QGRBG</td>
<td>22.3356</td>
</tr>
</tbody>
</table>

Note: \( H_0 \) is the randomization model for the specific variables of interest in the sense that there is no association between these variables and the subpopulation to which patients were assigned.
**Figure 3.3. Multivariate Analysis of Covariance for Rank Data Example**

**H₀:** For each clinic (STRATUM) there is no relationship between the 2 treatment groups (SUBPOP) and degree of improvement over the 3 time points (RESPONSE) after adjusting for initial stage of the skin condition (COVAR) and any missing value substitutions (MISS IND) in the sense that the observed partition of the improvement scores into treatment groups can be regarded as equivalent to a successive set of random samples of n1 and n2 patients.

**SUBPOPULATION AND OVERALL STRATUM (0,0) MEANS**

<table>
<thead>
<tr>
<th>STRATUM</th>
<th>SUBPOP</th>
<th>N</th>
<th>RESPONSE</th>
<th>RESPONSE</th>
<th>RESPONSE</th>
<th>COVAR</th>
<th>MISS IND</th>
<th>MISS IND</th>
<th>MISS IND</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROW1</td>
<td>5</td>
<td>19</td>
<td>13.3</td>
<td>13.5</td>
<td>18.6</td>
<td>20.1</td>
<td>0.05</td>
<td>0.16</td>
<td>0.21</td>
</tr>
<tr>
<td>ROW2</td>
<td>5</td>
<td>18</td>
<td>25.0</td>
<td>25.8</td>
<td>19.4</td>
<td>17.8</td>
<td>0</td>
<td>0.11</td>
<td>0.50</td>
</tr>
<tr>
<td>ROW3</td>
<td>0</td>
<td>37</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>0.03</td>
<td>0.14</td>
<td>0.35</td>
</tr>
<tr>
<td>ROW1</td>
<td>6</td>
<td>17</td>
<td>12.1</td>
<td>14.1</td>
<td>17.1</td>
<td>16.2</td>
<td>0.12</td>
<td>0.06</td>
<td>0.38</td>
</tr>
<tr>
<td>ROW2</td>
<td>6</td>
<td>16</td>
<td>22.2</td>
<td>20.1</td>
<td>16.9</td>
<td>17.8</td>
<td>0.18</td>
<td>0.19</td>
<td>0.21</td>
</tr>
<tr>
<td>ROW3</td>
<td>0</td>
<td>33</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>0.15</td>
<td>0.21</td>
<td>0.21</td>
</tr>
</tbody>
</table>

**SUBPOPULATION RESIDUAL SUMS OF VARIABLES**

<table>
<thead>
<tr>
<th>STRATUM</th>
<th>SUBPOP</th>
<th>N</th>
<th>RESPONSE</th>
<th>RESPONSE</th>
<th>RESPONSE</th>
<th>COVAR</th>
<th>MISS IND</th>
<th>MISS IND</th>
<th>MISS IND</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROW1</td>
<td>5</td>
<td>19</td>
<td>-108.5</td>
<td>-104.5</td>
<td>-7.5</td>
<td>21.5</td>
<td>0.49</td>
<td>0.43</td>
<td>-2.68</td>
</tr>
<tr>
<td>ROW2</td>
<td>5</td>
<td>18</td>
<td>108.5</td>
<td>104.5</td>
<td>7.5</td>
<td>-21.5</td>
<td>-0.49</td>
<td>-0.45</td>
<td>2.68</td>
</tr>
<tr>
<td>ROW1</td>
<td>6</td>
<td>17</td>
<td>-83.5</td>
<td>-50</td>
<td>-2</td>
<td>-13</td>
<td>-0.58</td>
<td>-2.61</td>
<td>2.61</td>
</tr>
<tr>
<td>ROW2</td>
<td>6</td>
<td>16</td>
<td>83.5</td>
<td>50</td>
<td>2</td>
<td>13</td>
<td>0.58</td>
<td>2.61</td>
<td>2.61</td>
</tr>
</tbody>
</table>

**STRATUM PARTIAL ASSOCIATION STATISTICS**

<table>
<thead>
<tr>
<th>Q_STAT</th>
<th>STRATUM</th>
<th>PA_STAT</th>
<th>DF</th>
<th>P_VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q(R1 R2 M1)</td>
<td>QHE</td>
<td>5</td>
<td>20.8351</td>
<td>7</td>
</tr>
<tr>
<td>Q(R1 R2 M2)</td>
<td>QHY</td>
<td>5</td>
<td>16.6129</td>
<td>3</td>
</tr>
<tr>
<td>Q(M1 M2 M3)</td>
<td>QHM</td>
<td>5</td>
<td>7.3503</td>
<td>3</td>
</tr>
<tr>
<td>Q(R1 R2 R3 STAGE</td>
<td>QRM</td>
<td>5</td>
<td>13.4749</td>
<td>4</td>
</tr>
<tr>
<td>Q(STAGE M1 M2 M3)</td>
<td>QHR</td>
<td>5</td>
<td>7.8271</td>
<td>4</td>
</tr>
<tr>
<td>Q(R1 R2 R3 STAGE M1 M2 M3)</td>
<td>QHG</td>
<td>5</td>
<td>13.0881</td>
<td>3</td>
</tr>
</tbody>
</table>

**AVERAGE PARTIAL ASSOCIATION STATISTICS**

<table>
<thead>
<tr>
<th>Q_STAT</th>
<th>APA_STAT</th>
<th>DF</th>
<th>P_VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q(R1 R2 M1 M2 M3)</td>
<td>QGRHE</td>
<td>26.449</td>
<td>3</td>
</tr>
<tr>
<td>Q(R1 R2 M3)</td>
<td>QGRMS</td>
<td>18.955</td>
<td>3</td>
</tr>
</tbody>
</table>

Note: H₀ is the randomization model for the specific variables of interest in the sense that there is no association between these variables and the subpopulations to which patients were assigned.

*These quantities should be ignored for this application because they involve the 0-value imputation for missing data. The other statistics shown here without *+ are adjusted for this imputation (see Appendix A.4)
Appendix: Statistical Theory

This Appendix is concerned with the statistical theory underlying multivariate randomization test statistics. For this purpose, finite population sampling principles are used to obtain expected values and covariance matrices with respect to which test statistics with approximate chi-square distributions are formulated.

A.1. One way multivariate analysis of variance randomization test statistic for a sub-population partition of a finite population. Let \( \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \) denote \((sx1)\) vectors for the response variables for the subjects in some finite population of size \( n \). Let \( i, i', \ldots, s \) index a set of sub-populations for which the relationship to the \( d \) response variables is to be investigated. Let \( n_1, n_2, \ldots, n_s \) denote the corresponding numbers of subjects so that \( \sum n_i = n \). If the sub-populations are homogeneous in the sense of covariance to random labels, then the distribution of the \( \mathbf{y}_i \) among them is compatible with the hypothesis (A1) of no association:

**H**: There is no relationship between sub-populations and responses in the sense of equally likely realizations for the \((n/n_i)\) possible exhaustive, random partitions of the \( n \) data vectors \( \mathbf{x}_i \) into successive random samples of sizes \( n_1, n_2, \ldots, n_s \) for the \( s \) sub-populations. (A1)

Let \( \mathbf{U} \) be a \((n/s)\) matrix of sub-population allocation indicators

\[
U_{ij} = \begin{cases} 1 & \text{if subject } j \text{ belongs to } i\text{-th sub-population} \\ 0 & \text{otherwise} \end{cases}
\]

Under the hypothesis \( H_0 \), it follows that

\[
E(\mathbf{U}_{ij} | H_0) = (n_i/n)
\]

\[
\text{Cov}(\mathbf{U}_{ij}, \mathbf{U}_{ij'}, | H_0) = -\frac{n_i(1-n_j/n)}{n(n-1)} \quad \text{if } i \neq i', \, j \neq j'
\]

\[
= \frac{n_i(n_j/n)}{n(n-1)} \quad \text{if } i \neq i', \, j = j'
\]

\[
= \frac{n_i}{n(n-1)} \quad \text{if } i = i', \, j = j'
\]

(A4)

The expected value and covariance structure in (A3) and (A4) is a consequence of the theory for simple random sampling from a finite population since the \( U_{ij} \) are random variables which describe the sampling process implied by \( H_0 \).

The response vectors \( \mathbf{y}_i \) are assumed here to be obtained without measurement error and to have values under \( H_0 \) that do not depend in any way on the sub-population to which the \( i \)-th subject is assigned. In other words, the \( \mathbf{y}_i \) are regarded as fixed constants in a framework like that for finite population sampling. More generally, consideration would be given to random response vectors \( \mathbf{y}_i \) which are mutually independent of each other and also the \( U_{ij} \) and have conditional expected values \( E(\mathbf{y}_i | U_{ij}, \mathbf{y}_j) = y_{ij} \) which are the same for all the sub-populations; i.e., a situation where the sampling and measurement random processes are independent.

Let \( \mathbf{y}_i = \sum_{j=1}^{n_i} \frac{1}{n_i} \mathbf{y}_{ij} \) denote the mean vector for the \( i \)-th sub-population with the \( y_{ij} \) being either fixed constants or conditional expected values with values which are unrelated under \( H_0 \) to the assigned sub-population. From (A3) and (A4), it follows that

\[
E(\mathbf{y}_i | H_0) = \sum_{j=1}^{n_i} \frac{\mathbf{y}_{ij}}{n_i} = \mathbf{y}_i
\]

(A5)

\[
\text{Cov}(\mathbf{y}_i, \mathbf{y}_j | H_0) = \frac{(\delta_{ij} n_i - 1)}{n_i(n-1)} \mathbf{I}
\]

(A6)

where \( \mathbf{I} = \mathbf{I}_{(n-1)^2} \). \( \delta_{ij} = 1 \) if \( i=j \) or \( 0 \) if \( i \neq j \).

Here \( \mathbf{y} \) is the finite population mean vector and \( \mathbf{y}_i \) is the finite population covariance matrix.

Let \( \mathbf{E} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_s) \) denote the compound vector of means for the \( s \) sub-populations. Then, (A7) and (A8) imply \( E(\mathbf{E} | H_0) = \mathbf{0} \) and

\[
\text{Var}(\mathbf{E} | H_0) = (\mathbf{y}_1(n-1)) \otimes (\mathbf{I}_{(n-1)^2})^{-1}
\]

(A9)

Thus, \( Q = \mathbf{g}'(\text{Var}(\mathbf{E} | H_0))^{-1} \mathbf{g} \) is a test statistic for \( H_0 \) which has an approximate chi-square distribution with \( D.F. = s(s-1) \) provided that the sample sizes \( n_i \) are sufficiently large for \( \mathbf{g} \) to have an approximate normal distribution. The test statistic \( Q \) can be interpreted as a one-way multivariate analysis of variance statistic since

\[
Q = \mathbf{g}'(\mathbf{y}_1(n-1)) \otimes (\mathbf{I}_{(n-1)^2})^{-1} \mathbf{g} \quad \text{(A10)}
\]

where \( \mathbf{g} = \sum_{i=1}^{s} \mathbf{g}_i \) and \( \mathbf{g}_i \) is the among sub-populations sums of products matrix.

If the data matrix for the \( \mathbf{y}_i \) corresponds to across subjects ranks for each of the respective responses, then \( Q \) in (A10) is the...
multivariate Kruskal-Wallis statistic (generalized to handle ties via midranks). Also, if the data matrix corresponds to binary indicators, then \( Q = \frac{1}{n} \sum_{i=1}^{q} \frac{1}{n_i} \) where \( Q \) is the Pearson chi-square statistic. Also, if the data matrix is non-redundant, then \( Q \) is distributed as \( \chi^2 \) with degrees of freedom \( D.F. = (q-1)(r-1) \) provided that the within stratum \( \chi^2 \) test statistic is used for the \( \chi^2 \) test statistic.

A.2. One-way multivariate analysis of covariance randomization test statistics. In Section A.1, attention was directed at multivariate test statistics for a set of response variables \( Y \). However, for many investigations data are also obtained for covariates \( X \) pertaining to sources of background information concerning subject heterogeneity. Moreover, these covariates are either known (by experimental design) or interpreted (by preliminary tests of significance) to be distributed at random with respect to the subject populations in the sense of \( H_0 \) in (Al). For this reason, the response variables and the covariates may be viewed as having opposite roles with respect to \( H_0 \) in the sense that the detection of significant differences is of interest for the response variables while the verification of the non-existence of such differences is of interest for the covariates.

The hypothesis \( H_0 \) in (Al) can be tested for the combined set of response variables and covariates via the one-way (m-variate) analysis of variance statistic \( Q(y|x) \) like that defined in (Al). Similarly, it can be tested for the \( t \) covariates via the one-way \( t \)-variate analysis of covariance statistic \( Q(z) \). Since the \( \{z_t\} \) are expected to be uncorrelated with \( Q(z) \), the statistic

\[
Q(y|x) = Q(y) - Q(z) = Q(z) + \sum_{t=1}^{T} \frac{1}{n_t} \sum_{i=1}^{n_t} (y_{it} - z_{it})^2/n_t
\]

represents a multivariate conditional reduction statistic for \( y \) which has an approximate chi-square distribution with degrees of freedom \( D.F. = (q-1)(r-1) \) and \( (d+1) - \sum_{t=1}^{T} \frac{1}{n_t} \sum_{i=1}^{n_t} z_{it}^2/n_t \) degrees of freedom for the null hypothesis.

The chi-square partition of \( Q(y|x) \) and \( Q(y) \) can be interpreted as weighted least squares arguments. Here, \( Q(y) \) is the residual goodness of fit statistic for a model constrained to satisfy \( H_0 \) with respect to the \( \{y_i\} \) randomization and \( Q(y|x) \) is the test statistic for null values on the parameter vector of that model. Furthermore, if this framework is simplified (by a substantial amount of algebra), it can be shown that

\[
Q(y|x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{n_i} \sum_{i=1}^{n_i} (y_{ix} - \bar{y}^x_{i})^2
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{n_i} \sum_{i=1}^{n_i} (y_{ix} - \bar{y}^x_{i})^2
\]

where \( \bar{y}^x_{i} = \frac{1}{n} \sum_{j=1}^{n_i} y_{ij} \) and \( \bar{y}^x_{i} = \frac{1}{n} \sum_{j=1}^{n_i} x_{ij} \) and \( \bar{y}^x_{i} = \frac{1}{n} \sum_{j=1}^{n_i} y_{ij} \) being appropriate components of \( y \). Thus, given that the data are non-redundant so that \( \bar{y}^x_{i} \) exists, \( Q(y|x) \) is a one-way multivariate analysis of covariance statistic. Otherwise, usage of \( Q(y|x) \) only requires verification of the compatibility of \( Q(x) \) with \( H_0 \) for the \( \{x_t\} \); more specifically, its usage does not require any assumptions concerning parallel regression surfaces.

If the data matrix for the \( x_j \) are across subject ranks for each covariable and response variable, then \( Q(x|x) \) is equivalent to the rank analysis of covariance methods proposed by Quade [1967].

A.3. Partial association (two-way analysis of variance) randomization test statistics for a set of sub-population partitions of a stratified population. If there are \( q \) strata indexed by \( h = 1, 2, ..., q \) within which \( H_0 \) in (Al) is expressed simultaneously, then two test strategies of interest are total partial association statistics and average partial association statistics. The former is obtained as \( Q = \sum_{h=1}^{q} Q_h \) where \( Q_h \) is some type of within stratum statistic like (AIO) or (Al). It has an approximate chi-square distribution with \( D.F. = q(d+1) \) provided that within stratum sample sizes \( \{n_{h,i}\} \) are sufficiently large for the respective \( \{Q_h\} \) to have approximate chi-square distributions.

However, for most applications, the test statistics of interest for the stratified versions of \( H_0 \) are based on across strata summary measures

\[
E_{h=1}^{q} \sum_{h=1}^{q} Q_h \quad \text{for which}
\]

\[
E(Q_h | H_0) = \sum_{h=1}^{q} n_{hi} \bar{v}_h
\]

\[
Cov(Q_h, Q_h') = \sum_{h=1}^{q} n_{hi} \left( \bar{v}_h - \bar{v}_h' \right) \bar{v}_h / (n_{hi} - 1)
\]

Thus, if \( \bar{v} = (\bar{v}_1, ..., \bar{v}_q) \), it follows that

\[
E(\bar{v}) = \sum_{h=1}^{q} n_{hi} \left( \bar{v}_h - \bar{v} \right)
\]

\[
\text{Var}(\bar{v}) = \sum_{h=1}^{q} \frac{n_{hi}}{(n_{hi} - 1)} \left( \bar{v}_h - \bar{v} \right) \left( \bar{v}_h - \bar{v} \right)
\]

The average partial association (or generalized randomized block) statistic for \( H_0 \) has the form

\[
Q_{GRB} = \langle Q_{GRB} \rangle^q
\]

\[
Q = (D \otimes C)[Var(\bar{v}) | H_0](D \otimes C)^{-1}
\]

Under \( H_0 \), \( Q_{GRB} \) has an approximate chi-square distribution with \( D.F. = q(d+1) \) provided that the \( n_{hi} = \frac{1}{n} \) are sufficiently large for \( \bar{v} \) to have an approximate multivariate normal distribution, the research design is connected in a within stratum sense, and the data matrix is non-redundant.
in the sense that $Y_{ij}$ is non-singular.

The test statistic $Q_{GRB}$ is more practical than $Q_2$ because its power is directed at average differences and its large sample chi-square distribution only requires a large overall sample rather than large within stratum samples. In this sense, it is a two-way multivariate analysis of variance test statistic.

Two special cases of $Q_{GRB}$ are of some interest. If all the $v_{ijh}$ and the $x_{h}$ are within-stratum ranks for the respective responses, then $Q_{GRB}$ is the multivariate Friedman statistic. Also, if the data matrix corresponds to $d$ binary indicators for a $(d+1)$ category response, then $Q_{GRB}$ is the Mantel-Haenszel statistic with D.F.$=d(s-1)$ for sets of a $(d+1)$ contingency tables.

Covariates $X_{ij}$ can be taken into account in several ways with respect to partial association test statistics. In particular, if the $Y_{ijh}$ are non-singular for all $h$, then (A15-A19) can be applied to residual vectors $\hat{e}_{ij}$ like those defined in reference to (A12). Similarly, if some of the $Y_{ijh}$ are singular, then the corresponding $z_{ij}$ can be defined with respect to reduced sets of covariates which do not have to be the same across strata; i.e., the covariates used with stratum $h$ may be partially different from those used with stratum $h'$. Finally, if some or many of the $Y_{ijh}$ are singular and if adjustment for all covariates is of interest, then the average partial association reduction statistic

$$Q_{\hat{e}}(y|x) = Q_{GRB}(y|x) - Q_{GRB}(x)$$

(A20) can be used where $Q_{GRB}(y|x)$ and $Q_{GRB}(x)$ have definitions like that for $Q_{GRB} = Q_{GRB}(y)$ in (A15-A19). As is the case with $Q_{GRB}$, the statistic $Q_{\hat{e}}(y|x)$ has an approximate chi-square distribution with D.F.$=d(s-1)$ under $H_0$ provided the $\{n_{ihj}\}$ are large.

otherwise, $Q_{GRB}$ is not a generalized randomized block statistic in the usual sense since it is not expressed as a quadratic form for weighted sums of differences between observed and expected mean vectors or the inverse of the sum of the corresponding covariance matrices.

A.4. Missing data adjustment. A common source of difficulty for many statistical situations is the occurrence of missing data for one or more of the response variables under study for some subjects. However, for hypotheses of randomness like $H_0$ in (A1), such missing data simply represent additional phenomena which may be distributed at random. If the assumption that missing data are distributed at random among the sub-populations is realistic (i.e., their occurrences are independent of the conceptual values they may have had if observed), then the hypothesis of randomness for the data which are present is meaningful and can be tested by multivariate statistics.

These randomization test statistics are directed at the sub-population mean vectors for the observed data for the respective sub-populations in a sense analogous to (A10), but involve a covariance adjustment in the sense of (A11-A12) for the pattern of missing data with respect to the estimation of their corresponding covariance matrix.

The specific formulation for multivariate randomization test statistics with adjustment for randomly missing data can be conveniently expressed in terms of operations on a data array for which missing data are assigned the value 0 and their prevalence is linked to complementary binary indicators; i.e., the data array has the form

$$Y_{ijk} = \begin{cases} y_{ijk} & \text{if observed} \\ 0 & \text{if missing} \end{cases}$$

(A21)

Then, for this framework, the statistic $Q(y|x)$, as given by (A11-A12) with $Q(y|x)$ and $Q(x)$ obtained via (A10) for the array (A21), is the multivariate randomization test statistic for $H_0$ with respect to the observed $Y_{ijk}$ given adjustment for the presumed random pattern of missing data. Accordingly, it has an approximate chi-square distribution with D.F.$=d(s-1)$ provided that the sample sizes $n_{ij}$ are sufficiently large.

The test statistic $Q(y|x)$ described here can be verified to be invariant with respect to the assigned value (such as 0) for the missing data because the assignment of any arbitrary value $\bar{e}_{ij}$ to the missing data for the $j$-th response is equivalent to a linear transformation

$$y_{ijk} = y_{ijk} + \bar{e}_{ij}$$

Furthermore, it can be verified that the $\bar{e}_{ij}$ and $y_{ijk}$ in (A21) on which $Q(y|x)$ is based do not involve the $\{e_{ij}\}$ in any manner whatsoever. Thus, $Q(y|x)$ only depends upon the observed values of the $\{y_{ijk}\}$ and their pattern $\{y_{ijk}\}$

If the data array under consideration is the set of rank for the observed values of each of the respective response variables (with 0's assigned to missing data), then $Q(y|x)$ can be interpreted as an incomplete data multivariate Kruskal-Wallis statistic. In this regard, it can be noted that its value is the same as would occur for the linear transformed rank score array

$$y_{ijk} = y_{ijk} + \frac{(m+1)}{2} - \frac{m^2}{2} - \frac{1}{2} (1 - i_{ijk})$$

obtained by viewing missing data as tied with all observed data and all other missing data; i.e., the modified ranks for observed data are the observed ranks plus half the number of missing values for the corresponding response variable and the modified ranks for missing data are all $\{(m+1)/2\}$. For reasons stated previously, these incomplete data extensions of the multivariate Kruskal-Wallis statistic only depend upon the ranks for the observed data and the pattern for the missing data. Finally, the methods described here can be generalized to situations involving covariates $X_{ij}$ and stratified populations by applying the strategies described in Sections A.2-A.3.