A SAS PROGRAM FOR THE EXTENDED MANTEL-HAENSZEL PROCEDURE
Alan C. Fisher and Daniel M. Wildman, Hoechst-Roussel Pharmaceuticals Inc.

ABSTRACT

The extended Mantel-Haenszel procedure may be used to test for treatment group differences when data are obtained across strata. This nonparametric procedure is appropriate for response levels that are either dichotomous, ordinal, or continuous.

Input to our SAS program may be either contingency tables or the observed data for each subject. Scores assigned to each response level may be either integer, marginal rank, marginal percentile, combined rank, or user-specified scores. As output, in addition to the extended Mantel-Haenszel statistics, the program provides the user with summary statistics for each stratum along with a nonparametric test for treatment differences.

For dichotomous responses the extended Mantel-Haenszel procedure is equivalent to the standard Mantel-Haenszel procedure. For a single stratum, using rank or percentile scoring, it is equivalent to the Wilcoxon-Mann-Whitney test. The extended Mantel-Haenszel procedure therefore allows generalization of standard nonparametric tests to multi-stratum situations.

INTRODUCTION

We use the extended Mantel-Haenszel procedure, originally proposed by Mantel (1963), in clinical trials to test for treatment group differences when data are obtained across strata. We employ this procedure when response levels are either dichotomous, ordinal, or continuous. For multi-center clinical trials, each investigator is usually considered as an individual stratum. For single center clinical trials, the strata are usually some important baseline characteristic such as initial severity.

We may view the data of the ith stratum as a 2 by k contingency table where k designates the number of possible response levels. A typical table is shown below:

<table>
<thead>
<tr>
<th>Response Levels</th>
<th>1</th>
<th>2</th>
<th>...</th>
<th>k</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>a1</td>
<td>a2</td>
<td>...</td>
<td>a_k</td>
<td>a1+k</td>
</tr>
<tr>
<td>Group B</td>
<td>b1</td>
<td>b2</td>
<td>...</td>
<td>b_k</td>
<td>b1+k</td>
</tr>
<tr>
<td>Total</td>
<td>m1</td>
<td>m2</td>
<td>...</td>
<td>m_k</td>
<td>T_i</td>
</tr>
</tbody>
</table>

The ordering of the response levels are taken into account by assigning scores Y_{ij} to response level j within stratum i, j=1, 2, ..., k; i=1, 2, ..., g. Various scoring procedures have been discussed by Landis et al. (1979).

In our program, the following scoring procedures may be used:

- integer scores in which the score j is assigned to the jth level of response. Thus, Y_{ij} = j for all i, j.
- marginal rank scores in which separately for each stratum the scores assigned to each response level are determined by the ranks of the observations. The score assigned is the mid-rank of all observations at that response level. Thus, for
  \[ j \leq i : Y_{ij} = \frac{(m_{1j} + 1)/2}{j-1} \]
  \[ j > i : Y_{ij} = \frac{(m_{1j} + (m_{j1} + 1))/2}{m_{1j} + (m_{j1} + 1)/2} \]
- marginal percentile (ridit) scores which are marginal rank scores divided by the total sample size \( T_i \) for the stratum.
- combined rank scores in which the same set of rank scores are used for each stratum. To obtain these scores the combined data set over all strata are used. Thus, for
  \[ j \leq i : Y_{ij} = \frac{(\sum_i m_{1j} / T_i + 1)/2}{\sum_i m_{1j} / T_i + (\sum_i m_{j1} + 1)/2} \]
  \[ j > i : Y_{ij} = \frac{\sum_i m_{1j} / T_i + (\sum_i m_{j1} + 1)/2}{\sum_i m_{1j} / T_i} \]
- user-specified scores.

The extended Mantel-Haenszel chi-square test statistic, with one degree of freedom, is defined to be

\[ \chi^2 = \sum_{i=1}^{g} \sum_{j=1}^{k} \left( \frac{E_{i1j} Y_{ij}^2 - \hat{\lambda}_{ij} E_{i1j}^2}{E_{i1j}} \right) \]

where

\[ \hat{\lambda}_{ij} = \frac{E_{i1j} + E_{j1i}}{T_i} \]

and

\[ \text{Var} \left( \sum_{j=1}^{k} Y_{ij} \right) = \frac{\sum_{i=1}^{g} \sum_{j=1}^{k} \left( \frac{E_{i1j} Y_{ij}^2 - \hat{\lambda}_{ij} E_{i1j}^2}{E_{i1j}} \right)}{T_i} \]

The extended Mantel-Haenszel procedure may be used when the response levels are dichotomous. It yields an equivalent result to the original procedure for dichotomous results proposed by Mantel and Haenszel (1959) without the continuity correction factor if the scoring procedure employed is either integer, marginal percentile, or combined rank scoring. However, the extended Mantel-Haenszel procedure with marginal rank scoring is not equivalent to the standard Mantel-Haenszel procedure.
The extended Mantel-Haenszel procedure may also be used when there is only one stratum. Regardless of the number of response levels, when there is only one stratum the extended Mantel-Haenszel procedure with either marginal rank or percentile scoring is equivalent to the Wilcoxon-Mann-Whitney test.

**SAS PROGRAM**

We have developed a SAS program for the extended Mantel-Haenszel procedure (EHMF) which performs the same calculations as the SAS procedure PROC PARCAT, as described by Fleming et al. (1982). However, our SAS program, due primarily to the display of the output and several of the optional input and output features, is more suitable to the standard analysis and reporting of clinical trial data.

**EHMF** is a SAS program based on the concept of major routines with embedded macros. The user supplies control and formatting information in "user defined macros" which are employed by the routines during processing. The structural aspects of this program are similar to that of the Pharmacokinetic Analysis System developed by Wildman and Chen (1982).

The program is divided into three main modules, two for input and one for processing and output. Each module is an independent member of a partitioned data set (PDS) and is "INCLUDED" by a macro call issued in the user's program. A separate PDS member is automatically invoked which supplies the default definitions for the user defined macros.

**EHMF** handles any number of variables in a single run as long as they share a common range of response levels. This is particularly useful with repeated measures where the variable names are indexed chronologically (e.g., REL1-REL5, relief scores at .5, 1, 2, 4, and 8 hours). This feature can also be used for a series of dichotomous variables.

The input data for **EHMF** may be either individual scores or frequency counts. One input module (RANDATA) draws data from a SAS dataset which includes raw scores for each subject. The other input module (INPEREQ) accepts a sequential file consisting of frequency counts for each combination of variable, stratum, and treatment. Figure 1 illustrates a typical example of **EHMF** when the raw data of each subject is input. Figure 2 illustrates, for the same data set, an example in which the input data is frequency counts. Note that there are data for two variables, representing two time points.

The output for both of these examples is shown in Figure 3. **EHMF** produces three summary sections for each variable. In the first, frequency counts and treatment group means are displayed. However, if the number of response levels exceeds 30, only the means appear. The second section shows the results of within-stratum pairwise comparisons, including means differences, chi-squares, and p-values. The final section summarizes the treatment group means and pairwise comparisons over all strata and includes tests for interaction and treatment effects. This section does not appear for single-stratum runs.

For combining statistical results over many runs, or to format a special-purpose table, an output dataset may be specified containing mean differences, chi-squares, and p-values for each within-strata and overall treatment comparison.

**USER DEFINED MACROS**

In this section we provide details of the user defined macros which supply control and formatting information to the major routines. Figure 4 summarizes these macros and their defaults.

**HEADER1**

MACRO HEADER1 prints the first line of the title. It is usually used to identify the protocol or study number of the specific clinical trial.

**HEADER2**

MACRO HEADER2 prints the second line of the title. It is usually used to either identify the investigator or to state that the study is multi-center.

**FINOTE**

MACRO FINOTE prints a footnote at the bottom of the output. In this example the footnote denotes that substitute values were inserted for missing values by the last score substitution method.

**STRTA**

MACRO _STRTA supplies the variable name used to identify each of the strata. In this example the strata are identified as centers. In the case of only one stratum, the user specifies "NOSTRATA".

**SKIP**

MACRO _SKIP is applicable when there are more than two treatment groups. It identifies the pairwise comparisons that should not be statistically analyzed. In this illustrative example the macro is employed to exclude the comparison of group 1 (TRT.A) to 2 (TRT.B).

**VARNAMES**

MACRO _VARNAMES supplies a label to be printed above the data summary in the output.

**SHOWOPT**

MACRO _SHOWOPT determines the appearance of intermediate results. These intermediate results are in addition to the output shown in Figure 3 and are primarily intended for diagnostic purposes. Possible options are:
INTROPT

MACRO INTROPT determines the type of test for interaction to be performed. Possible options are:

0: no intermediate results
1: partial intermediate results
2: complete intermediate results

WEIGHT

MACRO WEIGHT specifies the scoring procedure to be employed to assign scores to each of the response levels. Possible options are:

RIDI for marginal percentile, also known as marginal ridit scoring
RANK for marginal rank scoring
INTEGER for integer scoring
CRANK for combined rank scoring
USER for user specified scores

Note that combined ridit-type scoring would yield the same result as combined rank scoring.

If user specified scores are employed, the scores for each response level of each stratum follow the word USER.

OUTPUT

MACRO OUTPUT supplies the name of an output SAS dataset to be created containing the statistical results.

When the input data to EMHP are frequency counts, the following statements are used to call the input and processing modules after specifying the user-defined macros:

_DATA_
_DATA_

The next line specifies the observed scores of the response levels. Up to 15 response levels are allowed when the input is frequency counts. The observed scores start in column 10 and four columns are allowed for each score.

The next set of lines specifies the frequency counts at each response level for a specific treatment group within a specific stratum for a specific variable. This information is in the following columns:

VARNUM: columns 1-2
STRATUM: columns 4-5
TREATMENT: columns 7-8
COUNTS: start in column 10 with four columns allowed for each frequency count.

When the input data to EMHP are the raw data of each subject, a SAS data set needs to be created. Up to 60 unique response levels are allowed when the input is raw data. In the example shown in Figure 2, the data set name is RELIEF. This data set consists of four variables: SUBJECT, TGRP, RELI, and REL2. It is mandatory that variables named SUBJECT and TGRP appear in this data set.

In addition to the previously described MACROs, the following two additional MACROs are required when the input is raw data:

DATASET

MACRO DATASET identifies the data set to be analyzed by EMHP.

DEPVAR

MACRO DEPVAR identifies the dependent variables in the data set to be analyzed. In this example the dependent variables are RELI and REL2, the relief scores at HOUR 1 and HOUR 2.

When the input to EMHP is raw data, the following two statements are employed after specifying the user defined macros to call the input and processing routines:

_DATA_
_DATA_

The example in Figure 2 illustrates the utilization of a unique feature in EMHP. MACRO SPECIAL allows for special assignments of values in the data set. In this example we use MACRO SPECIAL to assign subjects to either CENTER 1 or CENTER 2 depending upon whether the identification number of the subject is less than or greater than 100. Any complete SAS statement or group of statements can be supplied here to manipulate incoming raw data.

Finally, a PROC FORMAT must appear before calling the MANTEL routine to specify the labels for each value of TGRP (treatment group), VARNUM (variable number), and STRATA (stratum). The VARNUM formats are particularly important for identifying variable names or time points for multiple-variable runs. Any of these formats may be specified as blank to inhibit the appearance of labels.
AVAILABILITY

Inquiries about obtaining DHMP should be directed to the first author, c/o Hoechst-Roussel Pharmaceuticals Inc., Medical Research Department, Route 202-206 North, Somerville, New Jersey 08876. Although the program has received extensive testing, no warranty, expressed or implied, is made to the accuracy and functioning of the program. No responsibility is thus assumed by the authors.

FOOTNOTES

1. The user can modify these constraints by changing the columns input statement in the SAS source to allow, for example, 20 response levels of 3 columns.
2. The RENAME is useful for this purpose.

REFERENCES


DATA RELIEF;
INPUT SUBJECT TILRP RELI REL2;
CARDS;
  1 1 1 6
  2 3 3 3
  3 3 2 2
...
145 3 3 3
MACRO _HEADERI 'ANALGESIC DENTAL PAIN STUDY'
MACRO _HEADER2 'MULTI-CENTER STUDY [2 CENTERS]'s
MACRO _HEADER3 'LAST SCORER SUBSTITUTION' /
   DIO METHOD EMPLOYED'
MACRO _DATASET RELIEF X
MACRO _SPECIAL
   IF SUBJECT <10a THEN CENTER = 1: ELSE CENTER = 2: C
MACRO _VARNAME 'PAIN RELIEF SCORES':
MACRO _STRATA CENTER X
MACRO _FREQ TGRP1_1 AND TGRP2_2 X
PROC FORMAT;
   VALUE TGRP 1- 'TRT. A' 2- 'TRT. B' 3- 'PLACEBO';
   VALUE VARMNum 1_ 'HOUR 1' 2- 'HOUR 1';
   VALUE STRATA 1- 'ONE' 2- 'TWO':
MACRO _SHOWOPT 1
MACRO _INTROPT 1
MACRO _WEIGHT 'RIOIT'
INFREQ - MANTEL

FIGURE 1
EXAMPLE FOR RAW DATA INPUT

FIGURE 2
EXAMPLE FOR FREQUENCY COUNT INPUT

MACRO _HEADERS 'ANALGESIC DENTAL PAIN STUDY'
MACRO _HEADERS 'MULTI-CENTER STUDY [2 CENTERS]'s
MACRO _THNOTE 'LAST SCORER SUBSTITUTION' /
   DIO METHOD EMPLOYED'
MACRO _DATASET RELIEF X
MACRO _SPECIAL
   IF SUBJECT <10a THEN CENTER = 1: ELSE CENTER = 2: X
MACRO _VARNAME 'PAIN RELIEF SCORES':
MACRO _STRATA CENTER X
MACRO _FREQ TGRP1_1 AND TGRP2_2 X
PROC FORMAT;
   VALUE TGRP 1-'TRT. A' 2- 'TRT. B' 3- 'PLACEBO';
   VALUE VARMNum 1_ 'HOUR 1' 2- 'HOUR 1';
   VALUE STRATA 1- 'ONE' 2- 'TWO':
MACRO _VARNAME 'PAIN RELIEF SCORES'
MACRO _SHOWOPT 1
MACRO _INTROPT 1
MACRO _WEIGHT 'RIOIT'
INFREQ - MANTEL

//FREQS DD *
  1 2 3 4 5
  1 1 1 0 1 4 9 1
  1 2 4 4 7 2 2
  1 3 3 3 4 3 1
  1 2 2 5 5 5 2
  1 2 2 2 4 0 8 0
  1 2 0 0 0 0 0 0
  1 1 0 1 2 3 4 0
  2 3 4 0 8 4 2
  2 1 3 3 4 2 3 1
  7 3 1 9 5 0 4 1
  2 2 2 1 2 7 1
  2 2 3 4 3 5 2 1
### ANALGESIC DENTAL PAIN STUDY
#### MULTI-CENTER STUDY (2 CENTERS)

#### HOURS 1

**CENTER ONE**

<table>
<thead>
<tr>
<th>TREATMENT GROUP</th>
<th>PAIN RELIEF SCORES 1-5</th>
<th>N</th>
<th>MEAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRT. A</td>
<td>4 1 7 2 2</td>
<td>15</td>
<td>3.67</td>
</tr>
<tr>
<td>TRT. B</td>
<td>3 3 4 3 1</td>
<td>14</td>
<td>2.71</td>
</tr>
</tbody>
</table>

**COMPARISON**

<table>
<thead>
<tr>
<th>MEAN DIFFERENCE</th>
<th>CHI-SQUARE</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) vs (2)</td>
<td>12.095</td>
<td>0.029*</td>
</tr>
</tbody>
</table>

**CENTER TWO**

<table>
<thead>
<tr>
<th>TREATMENT GROUP</th>
<th>PAIN RELIEF SCORES 1-5</th>
<th>N</th>
<th>MEAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRT. A</td>
<td>2 5 1 2 2</td>
<td>16</td>
<td>3.81</td>
</tr>
<tr>
<td>TRT. B</td>
<td>3 4 0 0 3</td>
<td>15</td>
<td>3.00</td>
</tr>
<tr>
<td>PLACEDO</td>
<td>2 3 8 8 0</td>
<td>14</td>
<td>2.80</td>
</tr>
</tbody>
</table>

**COMPARISON**

<table>
<thead>
<tr>
<th>MEAN DIFFERENCE</th>
<th>CHI-SQUARE</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) vs (2)</td>
<td>0.03125</td>
<td>0.759</td>
</tr>
</tbody>
</table>

**OVERALL**

<table>
<thead>
<tr>
<th>TREATMENT GROUP</th>
<th>PAIN RELIEF SCORES 1-5</th>
<th>N</th>
<th>MEAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRT. A</td>
<td>31 1 2 2 2</td>
<td>15</td>
<td>3.67</td>
</tr>
<tr>
<td>TRT. B</td>
<td>50 2 2 2</td>
<td>16</td>
<td>2.80</td>
</tr>
<tr>
<td>PLACEDO</td>
<td>29 2 2 2</td>
<td>14</td>
<td>2.66</td>
</tr>
</tbody>
</table>

**COMPARISON**

<table>
<thead>
<tr>
<th>MEAN DIFFERENCE</th>
<th>INTERACTION</th>
<th>TREATMENT EFFECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) vs (2)</td>
<td>CHI-SQUARE</td>
<td>P-VALUE</td>
</tr>
<tr>
<td>TRT. A vs PLACEBO</td>
<td>12.095</td>
<td>0.029*</td>
</tr>
<tr>
<td>TRT. B vs PLACEBO</td>
<td>0.03125</td>
<td>0.759</td>
</tr>
</tbody>
</table>

**TREATMENT EFFECT**

<table>
<thead>
<tr>
<th>CHI-SQUARE</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.09</td>
<td>0.079</td>
</tr>
</tbody>
</table>

**ALL P-VALUES ARE TWO-TAILED**

* * 0.05
** ** * 0.01
*** *** * 0.001

**LAST SCORE SUBSTITUTION**

**METHOD EMPLOYED**

---

**FIGURE 3**

**MANTEL-HAENSZEL OUTPUT**

**EXTENDED MANTEL-HAENSZEL PROCEDURE**

**METHOD OF WEIGHTING: RIDITS**
## EMHP USER DEFINED MACROS AND THEIR DEFAULTS

<table>
<thead>
<tr>
<th>MACRO NAME</th>
<th>DESCRIPTION</th>
<th>EXAMPLES</th>
<th>DEFAULT</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEADER1</td>
<td>Text to appear at top left of printout. Often used to identify study. Quotes optional but are recommended.</td>
<td>MACRO HEADER1 'SUPERDRUG STUDY 101' %</td>
<td>BLANK</td>
</tr>
<tr>
<td>HEADER2</td>
<td>Text to appear just below HEADER1. Often used to identify investigator. Same requirements as HEADER1.</td>
<td>MACRO HEADER2 'DR. SMART' %</td>
<td>BLANK</td>
</tr>
<tr>
<td>FTNOTE</td>
<td>Text to appear as footnote at column 10. Must be bound by quotes. You may include a second footnote line by inserting <code>/* @OM</code> before the text for that line.</td>
<td>MACRO _FTNOTE 'WORST SCORE METHOD' %</td>
<td>BLANK</td>
</tr>
<tr>
<td>VARNAME</td>
<td>Label to appear over frequency counts and means in printout. Usually identifies variable. Must be in quotes.</td>
<td>MACRO _VARNAME 'RELIEF SCORES'</td>
<td>BLANK</td>
</tr>
<tr>
<td>SKIP</td>
<td>Controls omission of treatment group comparisons. Must skip display of DRUG 2 versus PLACEBO. Syntax should be in terms of TGRP1-N1 AND TGRP2-N2.</td>
<td>MACRO SKIP TGRP1-2 AND TGRP2-9 %</td>
<td>TORP1-2 AND TGRP2-N2</td>
</tr>
<tr>
<td>DATASET</td>
<td>For use with raw-data input, specifies name of SAS dataset to be analyzed.</td>
<td>MACRO _DATASET OS.A101REL %</td>
<td>LAST</td>
</tr>
<tr>
<td>DEPVAR</td>
<td>For use with raw-data input, supplies names of dependent variables to be analyzed. May be a single variable name, a list of names, or a range of names. Care should be taken in the use of the form VARA-VARZ, since the ordering of variables will follow order of creation.</td>
<td>MACRO _DEPVAR TOTPAIN %</td>
<td>NO DEFAULT</td>
</tr>
<tr>
<td>STRATA</td>
<td>Name of variable which represents stratum level. If there is only one stratum, or if no within-stratum results are desired, specify &quot;NOSTRATA&quot;.</td>
<td>MACRO _STRATA LOCATION %</td>
<td>NOSTRATA</td>
</tr>
<tr>
<td>SPECIAL</td>
<td>All-purpose macro applied to input dataset to perform any special assignments, selection of observations, or other data manipulation. Any number of complete SAS statements may be included.</td>
<td>MACRO SPECIAL IF OBJECT&gt;200 THEN DELETE; IF INVEST=100 AND SUBJECT&gt;43 THEN DO; INVEST_100; TGRP_3; END; %</td>
<td>EMPTY MACRO</td>
</tr>
<tr>
<td>OUTPUT</td>
<td>Name to be given to an output dataset containing statistical results of EMHP.</td>
<td>MACRO _OUTPUT RESULTO1 %</td>
<td>NULL</td>
</tr>
<tr>
<td>WEIGHT</td>
<td>Supplies one of 5 weighting options. Choose from: RANK, RIDIT, INTEGER, CRANK, USER. Must be enclosed in quotes, and may be shortened to first two letters. &quot;USER&quot; Must be followed by a series of numbers representing the desired weights. There must be as many numbers as response levels in the data. Separate with spaces or commas. See text for further details.</td>
<td>MACRO _WEIGHT 'RANK' %</td>
<td>RIDIT</td>
</tr>
<tr>
<td>INTROPT</td>
<td>Interaction test to be used for cross-strata analysis. Donone, 2test described by Fleiss, 2-non-parametric (see text).</td>
<td>MACRO _INTROPT 0 %</td>
<td>1</td>
</tr>
<tr>
<td>SHOWOPT</td>
<td>Specify option for printing intermediate calculations and temporary datasets. Donone, 1=partial, 2=extensive printout. Useful for error diagnosis and exploratory analyses.</td>
<td>MACRO _SHOWOPT 1 %</td>
<td>0</td>
</tr>
</tbody>
</table>