1982

Nonparametric Analysis of Right Censored Data with Multiple Comparisons

Hwei-Weng Shih
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/gradreports

Part of the Applied Statistics Commons

Recommended Citation
Shih, Hwei-Weng, "Nonparametric Analysis of Right Censored Data with Multiple Comparisons" (1982). All Graduate Plan B and other Reports. 1170.
https://digitalcommons.usu.edu/gradreports/1170

This Report is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Plan B and other Reports by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.
NONPARAMETRIC ANALYSIS OF RIGHT CENSORED DATA WITH MULTIPLE COMPARISONS

by

Hwei-Weng Shih

A report submitted in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE in
Applied Statistics
(Plan B)

UTAH STATE UNIVERSITY
Logan, Utah
1982
ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my major professor, David L. Turner, for his patient guidance, and help in organizing the paper. Thanks are also extended to my committee members, Dr. Ronald V. Canfield and Dr. Gregory Jones, for their help and criticism.

I would also like to express my sincere thanks to my parents for their encouragement and support in my graduate studies at Utah State University.

Hwei-Weng Shih
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>iv</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>v</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. METHODOLOGY</td>
<td>3</td>
</tr>
<tr>
<td>A Special Comparison for Two Exponential Distributions</td>
<td>3</td>
</tr>
<tr>
<td>The Wilcoxon Two Sample Rank Sum Test and the Kruskal-Wallis Test</td>
<td>4</td>
</tr>
<tr>
<td>A Generalized Kruskal-Wallis Test for Comparing K Censored Samples</td>
<td>7</td>
</tr>
<tr>
<td>III. MULTIPLE COMPARISONS</td>
<td>12</td>
</tr>
<tr>
<td>IV. EXAMPLES</td>
<td>14</td>
</tr>
<tr>
<td>The Methods Used</td>
<td>14</td>
</tr>
<tr>
<td>Example 1</td>
<td>14</td>
</tr>
<tr>
<td>Example 2</td>
<td>21</td>
</tr>
<tr>
<td>V. CONCLUSIONS</td>
<td>28</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>30</td>
</tr>
<tr>
<td>APPENDIXES</td>
<td>31</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Data and $\hat{\lambda}_i$ for sorted Monte Carlo data</td>
<td>16</td>
</tr>
<tr>
<td>2. The results of example 1 using F test for censored data in Table 1</td>
<td>17</td>
</tr>
<tr>
<td>3. The ranks for data censored at 20 from Table 1</td>
<td>18</td>
</tr>
<tr>
<td>4. The results of example 1 using Kruskal-Wallis test and the ranking in Table 3</td>
<td>19</td>
</tr>
<tr>
<td>5. Statistics $S^{**}$, $S^*$ and $S$ and the results of example 1 for F test, Kruskal-Wallis test, Breslow's method and Bonferroni method</td>
<td>20</td>
</tr>
<tr>
<td>6. The values $n_i$, $\hat{\lambda}_i$, $X_i$ and $\hat{\lambda}_i$ of each treatment</td>
<td>24</td>
</tr>
<tr>
<td>7. The results for example 2 using F test</td>
<td>24</td>
</tr>
<tr>
<td>8. The results of example 2 using Kruskal-Wallis test</td>
<td>25</td>
</tr>
<tr>
<td>9. The results of example 2 using F test, Kruskal-Wallis test, Breslow's method and Bonferroni method</td>
<td>26</td>
</tr>
</tbody>
</table>
ABSTRACT

Nonparametric Analysis of Right Censored Data with Multiple Comparisons

by

Hwei-Weng Shih, Master of Science
Utah State University, 1982

Major Professor: Dr. David L. Turner
Department: Applied Statistics

This report demonstrates the use of a computer program written in FORTRAN for the Burroughs B6800 computer at Utah State University to perform Breslow's (1970) generalization of the Kruskal-Wallis test for right censored data. A pairwise multiple comparison procedure using Bonferroni's inequality is also introduced and demonstrated. Comparisons are also made with a parametric F test and the original Kruskal-Wallis test. Application of these techniques to two data sets indicate that there is little difference among the procedures with the F test being slightly more liberal (too many differences) and the Kruskal-Wallis test corrected for ties being slightly more conservative than Breslow's test statistic.

(30 pages)
CHAPTER I
INTRODUCTION

Statistical relationships between variables must sometimes be estimated from incomplete data or data which has been censored. Censoring occurs when an experiment is stopped before the event of interest occurs. When this happens the recorded data do not provide direct information about the event. In this paper we shall consider only samples censored on the right. This means that the only information about the censored observations is their total number and the fact that each is greater than some known value. For example, if we were studying survival time of a patient or animal under a set of experimental conditions, the data would be analyzed while some patients or animals are still alive. According to Lagakos (1979) the analysis of censored data can be used to obtain as much information as an uncensored experiment would yield.

A fundamental problem in many life testing problems is a comparison of the survival-time distributions from two or more samples of censored data. Norman Breslow (1970) reviews a generalization of Wilcoxon's statistic for comparing two populations as proposed by Gehan (1965) for use when the observations are subject to arbitrary right censorship. Breslow also discusses Mantel's (1967) further generalization to the case of arbitrarily restricted observations, or left and right censorship. Both Mantel and Gehan base their calculations on the permutation distribution of the statistic, conditional on the observed censoring pattern for the combined sample.
Breslow (1970) extended Gehan's generalization of Wilcoxon's test to allow for testing the equality of K continuous distribution functions when observations are subject to arbitrary right censorship. Breslow's generalization is an extension of the Kruskal-Wallis test, and is the "state of the art" nonparametric test of equality of K groups with possibly differing distributions for the censoring variables.

Breslow's development of this extended Kruskal-Wallis test involves some very complicated formulae. He gives two "easy" approximations but even these would be very laborious to compute.

This report demonstrates the use of a computer program written for the Burroughs B6800 computer which translates Breslow's formulae into a form which may actually be used. A pairwise multiple comparison procedure using Bonferroni's inequality is also developed and demonstrated. Two sets of data will be analyzed using Breslow's procedure and the Bonferroni multiple comparison procedure. Comparisons will also be made with the parametric (F test) procedure for the case of data from exponential distributions. The nonparametric Kruskal-Wallis test for uncensored data will also be applied using the modifications for tied data discussed in Ott (1977).
CHAPTER II
METHODOLOGY

In this report the major method used to analyze the hypothesis is Breslow's generalization of the Kruskal-Wallis test. In addition to Breslow's method, the Kruskal-Wallis test is also used to test the hypothesis that $K \geq 2$ populations are identical using modifications when there are ties in the data. An $F$ test for the case of two exponential distributions is also performed. Comparisons will then be made among the various methods when results are known using a set of generated or Monte Carlo Data. The methods are then applied to a real set of data.

A Special Comparison for Two Exponential Distributions

Let the two exponential distributions with parameters $\lambda_1$ and $\lambda_2$ have probability density function

$$f(X_{ij}; \lambda_i) = \frac{1}{\lambda_i} \exp(-X_{ij}/\lambda_i) \quad i = 1, 2; j = 1, 2, \ldots, n_i$$

Then

$$\hat{\lambda}_1 = \frac{\sum_{j=1}^{n_i} X_{ij}}{\sum_{j=1}^{n_i} \delta_{ij}} = \frac{X_{i\cdot}}{\delta_{i\cdot}}.$$

is the maximum likelihood estimate of $\lambda_i$, $i = 1, 2$, where $X_{ij}$ equals the true value or censored value depending on whether $\delta_{ij}$ equals 1 (uncensored) or 0 (censored). Then
\[ R = \frac{\lambda_1}{\lambda_2} \]

is an F distributed random variable with 2\( \delta_1 \) and 2\( \delta_2 \) degrees of freedom. This result may be used to test \( H_0: \lambda_1 = \lambda_2 \).

The Wilcoxon Two Sample Rank Sum Test and the Kruskal-Wallis Test

The Wilcoxon rank sum test provides a nonparametric test of the hypothesis that two populations are identical, since the experimenter has obtained two samples from possibly different populations, and we wish to use a statistical test to see if we can reject the null hypothesis that the two populations are identical. That is, we wish to detect differences between the two populations on the basis of random samples from those populations. An approach to the two-sample problem is to rank the combined data from lowest to highest. We let \( R_1 \) denote the sum of the ranks for sample 1. \( R_1 \) can take on values ranging from \( n_1(n_1 + 1)/2 \) to \( (n_1 + n_2)(n_1 + n_2 + 1)/2 - n_2(n_2 + 1)/2 \). Intuitively, if \( R_1 \) is close to either extreme, we would have evidence to reject the null hypothesis that the two populations are identical, since sample 1 would then be all close to the bottom or the top of the ranked distribution.

The concept of a rank sum test was extended to a comparison of more than two populations by Kruskal and Wallis (1952). The \( K \geq 2 \) random samples have been obtained from each of \( K \) possibly different populations, and we want to test the null hypothesis that all of the populations are identical against the alternative that some of the populations tend to furnish greater observed values than other populations.
To perform the test, the $K \geq 2$ samples are combined into a single ordered sample, then ranks are assigned to the sample values from the smallest value to the largest, without regard to which population each value came from. Let $N$ denote the total number of observations,

$$N = \sum_{i=1}^{K} n_i$$

where $n_i$ is the number of observations from sample $i$. Let $R(X_{ij})$ denote the rank assigned to $X_{ij}$, $R_i$ be the sum of the ranks assigned to the $i$th sample,

$$R_i = \sum_{j=1}^{n_i} R(X_{ij}) \quad i=1, 2, \ldots, K.$$ 

Note that

$$\sum_{i=1}^{K} R_i = 1 + 2 + \ldots + N = \frac{N(N+1)}{2}.$$ 

If there are several observations tied or equal to each other, the average of their ranks is assigned to each of the tied observations.

The large sample approximation for the test statistic $T$ is based on the fact that $R_i$ is the sum of $n_i$ random variables. So the mean and variance of $R_i$ are given by

$$E(R_i) = \frac{n_i(N+1)}{2},$$

and

$$\text{Var}(R_i) = \frac{n_i(N+1)(N-n_i)}{2}.$$
Therefore

\[ \frac{R_j - E(R_j)}{\sqrt{\text{Var}(R_j)}} \]

is approximately distributed as a standardized normal random variable when \( n_i \) is large enough. Thus

\[ \left( \frac{R_j - E(R_j)}{\sqrt{\text{Var}(R_j)}} \right)^2 = \frac{(R_j - [n_i(N + 1)/2])^2}{n_i(N + 1)(N - n_i)/12} \]

is approximately distributed as a chi-square random variable with one degree of freedom. If the \( R_i \) were independent of each other the distribution of the sum

\[ T = \sum_{i=1}^{K} \frac{(R_i - [n_i(N + 1)/2])^2}{n_i(N + 1)(N - n_i)/12} \]

could be approximated using the chi-square distribution with \( K \) degrees of freedom. However, since the sum of the \( n_i \)'s is \( N \), there is some dependence among the \( R_i \)'s. Kruskal (1952) showed that if the \( i \)th term in \( T \) is multiplied by \((N - n_i)/N\) for \( i = 1, 2, \ldots, K \), then the result

\[ T = \sum_{i=1}^{K} \frac{(R_i - [n_i(N + 1)/2])^2}{n_i(N + 1)N/12} \]

is asymptotically distributed as a chi-square random variable with \( K - 1 \) degrees of freedom. Since \( \sum_{i=1}^{K} R_i = N(N + 1)/2 \), \( T \) may be written as

\[ T = \sum_{i=1}^{K} \frac{(R_i - [n_i(N + 1)/2])^2}{n_i(N + 1)N/12} \]

\[ = \frac{12}{N(N + 1)} \sum_{i=1}^{K} \frac{1}{n_i} \left[ R_i^2 - R_i n_i(N + 1) + \frac{1}{4} n_i^2(N + 1)^2 \right] \]
\[
= \frac{12}{N(N+1)} \sum_{i=1}^{K} \frac{R_i^2}{n_i} - \frac{12}{N(N+1)} \left[ \frac{N(N+1)}{2} \cdot (N+1) - \frac{N}{4(N+1)^2} \right]
\]

\[
= \frac{12}{N(N+1)} \sum_{i=1}^{K} \frac{R_i^2}{n_i} - 3(N+1),
\]

is an equivalent form for \( T \), and is usually more convenient to use.

A modification proposed by Ott uses \( T' \) rather than \( T \) when there are groups of tied ranks. To do this we form the \( g \) groups composed of identical ranks, where the \( j \)th group contains \( t_j \) (\( j = 1, \ldots, g \)) ties. The statistic \( T' \) is then close to a chi-square random variables with \( K - 1 \) degrees of freedom where

\[
T' = \frac{T}{1 - [\frac{\sum (t_j^3 - t_j)/N^3}{N}]}.
\]

A Generalized Kruskal-Wallis Test for Comparing \( K \) Censored Samples

Although the Kruskal-Wallis test assumes only continuous underlying distributions, it does not do very well if there are large numbers of ties. This is especially so for censored data when the ties may lie among the upper values of the ranks.

To handle problems of right censored data, Breslow (1970) generalized the Kruskal-Wallis test. Let \( X_{ij}^0 \) be the true observation for the \( j \)th individual obtained from the \( i \)th population (\( j = 1, \ldots, N_i; i = 1, \ldots, K \)). Variable \( Z_{ij} \) is used to censor \( X_{ij}^0 \), so sometimes the true observation \( X_{ij}^0 \) may not be observed. The observed data which we can get from a real sample is \( X_{ij} = \min(X_{ij}^0, Z_{ij}) \). \( X_{ij} \) should indicate with a variable \( \delta_{ij} \) whether or not \( X_{ij} \) is in fact censored: i.e., \( \delta_{ij} = 1 \) when \( X_{ij}^0 = Z_{ij} \) (uncensored); \( \delta_{ij} = 0 \) when \( X_{ij} = Z_{ij} < X_{ij}^0 \).
(censored). \( N = N_1 + \ldots + N_k \) is the total sample size and \( \lambda_1 = \frac{N_1}{N} \) is the proportion of the \( i \)th sample size to the total sample size.

\( F_i \) is the \( i \)th cumulative distribution function. The null hypothesis to be tested is \( H_0: F_1 = \ldots = F_k \), which specified that \( K \) populations have equal distribution functions.

Breslow (1970) defined a scoring function \( x \) for comparing two observations \( X_{ij} \) and \( X_{ij'} \) by

\[
\begin{align*}
    x(X_{ij}, \delta_{ij}; X_{ij'}, \delta_{ij'}) = & \begin{cases} 
        -1 & X_{ij} < X_{ij'}; \delta_{ij} = 1, \delta_{ij'} = 1 \\
        -1 & X_{ij} < X_{ij'}; \delta_{ij} = 1, \delta_{ij'} = 0 \\
        +1 & X_{ij} > X_{ij'}; \delta_{ij} = 1, \delta_{ij'} = 1 \\
        +1 & X_{ij} > X_{ij'}; \delta_{ij} = 0, \delta_{ij'} = 1 \\
        0 & \text{otherwise.}
    \end{cases}
\end{align*}
\]

The \( x \) function is then used in computing a vector score statistic \( W \). The \( i \)th component of this vector score statistic is defined to be the total score comparing the \( i \)th sample with the remaining \( K - 1 \) samples,

\[
W_i = \sum_{j=1}^{K} \sum_{i'=1}^{N_i} x(X_{ij}, \delta_{ij}; X_{ij'}, \delta_{ij'}).
\]

For uncensored data sets, \( W_i = 2[R_i - (1/2)N_i(N + 1)] \). Large negative values of \( W_i \) mean that observations in the \( i \)th sample are smaller than those from other samples and large positive values of \( W_i \) would indicate that the \( i \)th sample had larger than average values. The total of \( W_i \) should be equal to 0.

Breslow (1970) goes on to use this \( W \) vector to form test statistics for testing the equality of \( K \) distribution functions. His first statistic refers to Rao (1965) which shows that the well-known large sample theory for chi-square statistics holds for the statistic
\[ S^* = \sum_{i'=1}^{K} \sum_{j'=1}^{N_i} \delta_{i',j'} \left( \sum_{i=1}^{K} \sum_{j=1}^{N_i} e(X_{ij} - X_{ij'}) \right)^2 \]

where

\[ e(x) = \begin{cases} 
1 & \text{if } x > 0 \\
0 & \text{if } x \leq 0.
\end{cases} \]

Under general regularity conditions, \( S^* \) can be shown to have an asymptotic chi-square distribution with \( K - 1 \) degrees of freedom.

Breslow recommends evaluation of \( S^{**} \) in order to check on computational accuracy of \( S^* \). \( S^{**} \) is a lower bound for \( S^* \) and is easily computed as:

\[ S^{**} = 3N^{-2} \sum_{i=1}^{K} (W_i^2/N_i). \]

Breslow (1970) goes on to develop a statistic which is calculated as follows. A covariance matrix \( \Sigma \) must be computed. Individual terms \( \sigma_{ii'} \) can be calculated from

\[ \begin{align*}
N \sigma_{ii'} = & - \sum_{i'=1}^{K} \sum_{j=1}^{N_i} \delta_{i',j} \sum_{j=1}^{N_{i''}} \left( X_{ij} - X_{ij''} \right) e(X_{ij}, X_{ij''}) \\
\sigma_{ii'} = & - \sum_{i' \neq i}^{N} \sigma_{ii'}
\end{align*} \]

where \( e(x) = 1 \) if \( x > 0 \), \( 0 \) if \( x \leq 0 \).

The covariance matrix \( \Sigma \) is then decomposed into \( K - 1 \) vectors

\[ \Sigma_i = (\xi_{i1}, \ldots, \xi_{ik})' \quad (i=1, \ldots, K-1) \]

such that \( \xi_{i1}' \Sigma \xi_i = 1 \) and \( \xi_{i}' \Sigma \xi_j = 0 \) (\( i \neq j \)). The vectors \( \Sigma_i \) may be easily found by using the Gram-Schmidt orthogonalization process. We can use the \( K - 1 \) vectors

\[ Y_1 = (1, 0, \ldots, 0)', \quad Y_2 = (1, 1, 0, \ldots, 0)', \quad \ldots, \]

\[ Y_{K-1} = (1, 0, \ldots, 0, 1)' \]
\[ Y_{K-1} = (1, \ldots, 1, 0)' \]

as a starting point for the Gram-Schmidt process. We denote the inner product of two vectors by

\[(X, Y) = X' \hat{\circ} Y.\]

The Gram-Schmidt process proceeds as follows:

1. Calculate \( \xi_1 = \frac{Y_1}{\|Y_1\|} \) where \( \|Y_1\| = \sqrt{(Y_1, Y_1)} = \sqrt{Y_1' \hat{\circ} Y_1} \)

2. Use \( \xi_1 \) to calculate \( Z_2 = Y_2 - (Y_2, \xi_1)\xi_1 \) and \( \xi_2 = \frac{Z_2}{\|Z_2\|} \)

   where \( \|Z_2\| = \sqrt{(Z_2, Z_2)} = \sqrt{Z_2' \hat{\circ} Z_2} \)

3. Use above information to calculate \( Z_3 = Y_3 - (Y_3, \xi_1)\xi_1 - (Y_3, \xi_2)\xi_2 \)

   and \( \xi_3 = \frac{Z_3}{\|Z_3\|} \) where \( \|Z_3\| = \sqrt{(Z_3, Z_3)} = \sqrt{Z_3' \hat{\circ} Z_3} \)

   (\( K-1 \))st. Step. Use above information to calculate

   \[ Z_{K-1} = Y_{K-1} - (Y_{K-1}, \xi_1)\xi_1 - (Y_{K-1}, \xi_2)\xi_2 - \cdots \]

   \[ - (Y_{K-1}, \xi_{K-2})\xi_{K-2} \] and

   \[ \xi_{K-1} = \frac{Z_{K-1}}{\|Z_{K-1}\|} \] where \( \|Z_{K-1}\| = \sqrt{(Z_{K-1}, Z_{K-1})} = \sqrt{Z_{K-1}' \hat{\circ} Z_{K-1}} \)

The statistics

\[ S_1 = N^{-3/2} \xi_1 \hat{\circ} W \]

are easily found and a combined test statistic is calculated as
\[ S = \sum_{i=1}^{K-1} S_i^2 \]

which is a chi-square random variable with \( K - 1 \) degrees of freedom.

Breslow suggests calculating \( S^{**} \) as a lower bound to \( S^* \). \( S \) and \( S^* \) are asymptotically equivalent statistics, but \( S^* \) is computationally far easier to compute. The "easier" \( S^* \) and \( S^{**} \) are needed only if a computer program is not available to calculate \( S \). A computer program is given in the Appendix which translates Breslow's formulae into a FORTRAN IV program for the Burroughs B6800 computer.
CHAPTER III
MULTIPLE COMPARISONS

The procedures described in Chapter II provide an overall test of the equality of $K \geq 2$ distributions. For $K > 2$, if the populations are declared significantly different, then a multiple comparison procedure is needed to isolate the differences.

The Bonferroni inequality provides one method of simultaneously estimating several confidence intervals. Let $A_1$ denote the first event, say a $1 - \alpha_1$ confidence interval, and let $A_2$ denote the second event also a $1 - \alpha_2$ confidence interval. We can then use the Bonferroni inequality to get the probability of both events of $A_1$ and $A_2$ occurring simultaneously. We already know that

$$P(A_1 \cap A_2) = 1 - P(\bar{A}_1) - P(\bar{A}_2) + P(\bar{A}_1 \cap \bar{A}_2)$$

and since $P(\bar{A}_1 \cap \bar{A}_2) \geq 0$, we obtain the Bonferroni inequality:

$$P(A_1 \cap A_2) \geq 1 - P(\bar{A}_1) - P(\bar{A}_2).$$

For this situation, the joint confidence is

$$P(A_1 \cap A_2) \geq 1 - \alpha_1 - \alpha_2.$$

The Bonferroni inequality can easily be extended to $K$ simultaneous confidence intervals with family confidence coefficient $1 - \alpha$ by requiring $P(\bar{A}_i) = \alpha_i$; and $\sum \alpha_i = \alpha$ which then gives

$$\sum_{i=1}^{K} P(\bar{A}_i) \geq 1 - \alpha.$$
For example, let

\[ A_{12} \] be a 99% confidence interval for \( \mu_1 - \mu_2 \),

\[ A_{13} \] be a 99% confidence interval for \( \mu_1 - \mu_3 \),

and \( A_{23} \) be a 99% confidence interval for \( \mu_2 - \mu_3 \).

The Bonferroni inequality then guarantees us a family or simultaneous confidence interval of at least 97 percent that the three intervals based on the same sample are simultaneously correct, i.e.,

\[ P( A_{12} \cap A_{13} \cap A_{23} ) \geq .97. \]

If \( K \) interval estimates are desired with a family confidence coefficient \( 1 - \alpha \), constructing each interval estimate with statement confidence coefficient \( 1 - \alpha/K \) will suffice. The Bonferroni technique is ordinarily most useful when the number of simultaneous estimates is not too large. Note that different statement confidence coefficients also could be calculated, as long as \( \sum_{i=1}^{K} P(\bar{A}_i) = \alpha \).

For instance, the event \( A_1 \) may be a 98 percent confidence interval and the event \( A_2 \) could be a 97 percent confidence interval. The family confidence coefficient would then be at least 95 percent.
CHAPTER IV
EXAMPLES

The Methods Used

In this chapter two examples are given to illustrate the analysis of censored data. We will apply the $F$ test for exponential data, the Kruskal-Wallis test for ranked data and Breslow's method for censored data. The Bonferroni method will then be used to test pairwise comparisons.

Example 1

In the first example, three groups of data were generated from known exponential distributions. The procedure to get the three data sets uses an integral transform, i.e., if $F(X)$ is the distribution function for a random variable $X$, and if $X_1, \ldots, X_n$ is a random sample from $F(\cdot)$ then $U_i = F(X_i)$ for $i = 1, \ldots, n$ will be a random sample of uniform random variables over the interval $(0, 1)$. It follows then that if $U_1, \ldots, U_n$ is a random sample from a uniform distribution, then $X_i = F^{-1}(U_i)$ for $i = 1, \ldots, n$ will be a random sample from $F(\cdot)$.

It is easy to use a computer to generate uniform random numbers, and then we may use the integral transformation technique for finding random numbers from a given distribution. For this example we got three groups of uniform random numbers from 0 to 1 using MINITAB. If $F(\cdot)$ is a negative exponential distribution, $F_X(X) = 1 - e^{-X/\lambda} = \mu$, then $X = F_X^{-1}(\mu) = -\lambda \ln (1 - \mu)$ has a negative exponential distribution.
with parameter \( \lambda \). i.e., the density function of \( X \) is \( f_X(x) = (1/\lambda)e^{-x/\lambda} \), which is a negative exponential distribution. Table 1 presents such samples from three negative exponential distributions. Each sample has been sorted for ease in censoring at an arbitrary value of 20.

If we ignore the fact that the data is censored, we can get \( \hat{\lambda}_i (\hat{\lambda}_i = X_i / \delta_i) \) for each group and then use \( R = \hat{\lambda}_i / \hat{\lambda}_j \) \((i \neq j)\) to do an F test with \( 2\delta_i \) and \( 2\delta_j \) degree of freedom. All possible F tests are listed in Table 2.

If we ignore the censoring in Table 1, Table 3 then gives the \( R_i \)'s, the sum of the ranks for each group, and the Kruskal-Wallis tests are listed in Table 4.

Since Breslow's method includes many complicated formulae, the computer program listed in the Appendix was used to get the statistics \( S^{**}, S^* \) and \( S \). We used \( S \) to do the chi-square test and calculated \( S^* \) and \( S^{**} \) for illustrative purpose only. Using an experimentwise error rate of .05, the three pairwise comparisons are \( G_1 = G_2 \), \( G_1 = G_3 \) and \( G_2 = G_3 \). The Bonferroni procedure then uses \( 1 - .05/3 = .9833 \) as confidence coefficient for the individual intervals. We list the results of the F test, the Kruskal-Wallis test, Breslow's method and the Bonferroni method in Table 5.
Table 1. Data and $\lambda_i$ for sorted Monte Carlo data.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>True $\lambda_i$</td>
<td>12</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>$n_i$</td>
<td>15</td>
<td>12</td>
<td>18</td>
</tr>
<tr>
<td>Sorted Uncensored Data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X (Uncensored)</td>
<td>20.81</td>
<td>9.158</td>
<td>5.7466</td>
</tr>
<tr>
<td>$\delta_i = \sum_{j=1}^{n_i} \delta_{ij}$</td>
<td>9</td>
<td>9</td>
<td>17</td>
</tr>
<tr>
<td>$X_i = \sum_{j=1}^{n_i} X_{ij}$ for data censored at 20</td>
<td>197.798</td>
<td>88.544</td>
<td>103.191</td>
</tr>
<tr>
<td>$\lambda_i = X_i / \delta_i$</td>
<td>21.978</td>
<td>9.838</td>
<td>6.07</td>
</tr>
</tbody>
</table>
Table 2. The results of example 1 using F test for censored data in Table 1.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>$G_1 = G_2$</th>
<th>$G_1 = G_3$</th>
<th>$G_2 = G_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R = \lambda_i / \lambda_j$</td>
<td>2.234&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3.621&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.621</td>
</tr>
<tr>
<td>Degrees of Freedom</td>
<td>18, 18</td>
<td>18, 34</td>
<td>18, 34</td>
</tr>
<tr>
<td>P Value&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.04845</td>
<td>0.00060</td>
<td>0.11002</td>
</tr>
</tbody>
</table>

<sup>a</sup>Significant at $\alpha = 0.05$.

<sup>b</sup>Run STATPAC/DIST to get the probability of an F value larger than observed when the degrees of freedom are $2n_i$ and $2n_j$, respectively.
Table 3. The ranks for data censored at 20 from Table 1.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>$G_1 = G_2$</th>
<th>$G_1 = G_3$</th>
<th>$G_2 = G_3$</th>
<th>$G_1 = G_2 = G_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groups</td>
<td>$G_1$</td>
<td>$G_2$</td>
<td>$G_1$</td>
<td>$G_3$</td>
</tr>
<tr>
<td>$R(X_{ij})$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>2</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>3</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>4</td>
<td>19</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>5</td>
<td>21</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>7</td>
<td>22</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>8</td>
<td>24</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>11</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>17</td>
<td>26</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>23</td>
<td>30</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>23</td>
<td>30</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>23</td>
<td>30</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>30</td>
<td>16</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>30</td>
<td>17</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>30</td>
<td>18</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td></td>
<td></td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td></td>
<td></td>
<td>30</td>
</tr>
</tbody>
</table>

$R_i = \sum_{j=1}^{n_i} R(X_{ij})$

<table>
<thead>
<tr>
<th></th>
<th>251</th>
<th>127</th>
<th>343</th>
<th>218</th>
<th>167.5</th>
<th>297.5</th>
<th>474</th>
<th>216.5</th>
<th>344.5</th>
</tr>
</thead>
</table>
Table 4. The results of example 1 using Kruskal-Wallis test and the ranks given in Table 3.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>$G_1 = G_2$</th>
<th>$G_1 = G_3$</th>
<th>$G_2 = G_3$</th>
<th>$G_1 = G_2 = G_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>4.0024</td>
<td>10.1229</td>
<td>0.6134</td>
<td>9.70</td>
</tr>
<tr>
<td>$T'$</td>
<td>4.1546$^a$</td>
<td>10.2185$^a$</td>
<td>0.6148</td>
<td>9.8066$^a$</td>
</tr>
<tr>
<td>$P$ Value$^b$</td>
<td>0.04152</td>
<td>0.00139</td>
<td>0.43299</td>
<td>0.00742</td>
</tr>
</tbody>
</table>

$^a$Significant at $\alpha = .05$.

$^b$Probability of a $X^2$ value larger than observed.
Table 5. Statistics S**, S* and S, and the results of example 1 for F test, Kruskal-Wallis test, Breslow's method and Bonferroni method.

<table>
<thead>
<tr>
<th>Method</th>
<th>Hypothesis</th>
<th>$G_1 = G_2$</th>
<th>$G_1 = G_3$</th>
<th>$G_2 = G_3$</th>
<th>$G_1 = G_2 = G_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breslow and Bonferroni</td>
<td>S**</td>
<td>4.1506</td>
<td>10.4296</td>
<td>0.6338</td>
<td>9.9127</td>
</tr>
<tr>
<td></td>
<td>S*</td>
<td>4.5410</td>
<td>11.0086</td>
<td>0.6679</td>
<td>10.3524</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>5.0971$^a$</td>
<td>11.7467$^b$</td>
<td>0.7309</td>
<td>9.7876$^a$</td>
</tr>
<tr>
<td>Kruskal-Wallis and Bonferroni</td>
<td>$T'$</td>
<td>4.1546$^a$</td>
<td>10.2185$^b$</td>
<td>0.6148</td>
<td>9.8066$^a$</td>
</tr>
<tr>
<td></td>
<td>$\chi^2_{K-1}$</td>
<td>3.84</td>
<td>3.84</td>
<td>3.84</td>
<td>5.991</td>
</tr>
<tr>
<td>F Test</td>
<td>R</td>
<td>2.234$^c$</td>
<td>3.621$^c$</td>
<td>1.621</td>
<td></td>
</tr>
<tr>
<td>P Values</td>
<td>S</td>
<td>0.02397</td>
<td>0.00061</td>
<td>0.39259</td>
<td>0.00749</td>
</tr>
<tr>
<td></td>
<td>$T'$</td>
<td>0.04152</td>
<td>0.00139</td>
<td>0.43299</td>
<td>0.00742</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>0.04845</td>
<td>0.00060</td>
<td>0.11002</td>
<td></td>
</tr>
</tbody>
</table>

$^a$Significant for chi-square value with $\alpha = .05$.

$^b$Significant for chi-square value with $\alpha = .05/3 = .01667$.

$^c$Significant for F value with $\alpha = .05$. 
From Table 5, it is easy to see that the statistic $S^{**}$ is a lower bound to $S^*$ and it also is a lower bound to $S$ except for the case $G_1 = G_2 = G_3$. $T'$, the tie-corrected Kruskal-Wallis test statistic, is very close to $S^{**}$ in this example. To allow easy comparison of the $S$, $T'$ and $R$ test results, p-values were obtained by running STATPAC/DIST to get the probability of a large chi-square or $F$ statistic. For this data set the statistic $R$ always got the smallest probability except for the test of $G_1 = G_2$. This means that when we test the null hypothesis of equality of two groups, the $R$ value is possibly too liberal, i.e., too easy to reject. The probability of $S$ and $T'$ listed in Table 5 show that the p-value for $S$ is always smaller than the p-value of $T'$ except the case $G_1 = G_2 = G_3$. In this example, the three statistics $S$, $T'$ and $R$ yielded the same conclusions, i.e., the "significant" difference between group 1 and groups 2 and 3. These results are somewhat surprising since there is a relatively small difference between the $\chi_i$'s for groups 1 and 2. Since this was Monte Carlo data, these differences may be ascribed to chance. Further Monte Carlo work would undoubtedly tend to "smooth" these unexpected differences.

Example 2

The second example is a nutrition experiment conducted by Susan Collinge who was a graduate student in Nutrition Food Science Department, USU, in 1981. Susan looked at how many samples of meat products were bad each day when they were put in 27°C (80.6°F) temperature room. Each of the nine treatments contained twenty-five sealed bags of meat with different chemical additives. During each day a count of the number of swollen bags was made. The swelling indicated spoilage
of the contents. After 100 days the experiment was terminated, resulting in some treatments having censored data. The results of the nine treatments are compared below to see what kind of chemical combination added to the meat will keep the meat from spoilage for the longest period of time. The nine treatments were:

- Treatment 1. Control - no chemicals
- Treatment 2. Nitrite only
- Treatment 3. Nitrite + 20 ppm FeCl₃
- Treatment 4. Nitrite + Myoglobin
- Treatment 5. Nitrite + 200 ppm EDTA + Myoglobin
- Treatment 6. Nitrite + Nytrosylmyoglobin
- Treatment 7. Nitrite + 200 ppm EDTA
- Treatment 8. Nitrite + 200 ppm EDTA + 20 ppm FeCl₃
- Treatment 9. Nitrite + 200 ppm EDTA + 40 ppm FeCl₃

We are interested in the following specific comparisons:

1. Treatment 1 vs. treatment 2 through 9.
2. Treatment 2 vs. treatment 3 through 9.
3. Treatment 3 vs. treatment 4.
4. Treatment 3 vs. treatment 8.
5. Treatment 4 vs. treatment 6.
6. Treatment 5 vs. treatment 7.
7. Treatment 7 vs. treatment 8.
8. Treatment 8 vs. treatment 9.

Table 6 gives the values \( n_i \), \( \delta_i \), \( X_i \), and \( \lambda_i \), and the results of the F test are listed in Table 7. The \( R_i \)'s and results of the Kruskal-Wallis test are shown in Table 8. Table 9 shows all the results of the
F test, the Kruskal-Wallis test, Breslow's method and the Bonferroni method. From Table 9 we find the different methods yield the same results, i.e., only treatment 3 and treatment 4 are homogeneous.
Table 6. The values of $n_i$, $\xi_i$, $X_i$, and $\hat{x}_i$ of each treatment.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>$T_1$</th>
<th>$T_2$</th>
<th>$T_3$</th>
<th>$T_4$</th>
<th>$T_5$</th>
<th>$T_6$</th>
<th>$T_7$</th>
<th>$T_8$</th>
<th>$T_2 - T_8$</th>
<th>$T_3 - T_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi_i$</td>
<td>25</td>
<td>25</td>
<td>23</td>
<td>25</td>
<td>14</td>
<td>13</td>
<td>1</td>
<td>1</td>
<td>11</td>
<td>88</td>
</tr>
<tr>
<td>$X_i$</td>
<td>100</td>
<td>559</td>
<td>587</td>
<td>473</td>
<td>1457</td>
<td>1893</td>
<td>2424</td>
<td>2421</td>
<td>2416</td>
<td>2421</td>
</tr>
<tr>
<td>$\lambda_i = \frac{X_i}{\xi_i}$</td>
<td>4</td>
<td>22.36</td>
<td>25.52</td>
<td>78.92</td>
<td>104.071</td>
<td>145.615</td>
<td>2424</td>
<td>164.182</td>
<td>2421</td>
<td>102.83</td>
</tr>
</tbody>
</table>

Table 7. The results for example 2 using F test.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>$T_1$ vs $T_2 - T_8$</th>
<th>$T_2$ vs $T_3 - T_8$</th>
<th>$T_3$ vs $T_4$</th>
<th>$T_4$ vs $T_6$</th>
<th>$T_5$ vs $T_6$</th>
<th>$T_6$ vs $T_7$</th>
<th>$T_7$ vs $T_8$</th>
<th>$T_8$ vs $T_9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R = \frac{\hat{x}_1}{\hat{x}_j}$</td>
<td>25.7075&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5.621&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.3488&lt;sup&gt;a&lt;/sup&gt;</td>
<td>6.433&lt;sup&gt;a&lt;/sup&gt;</td>
<td>7.696&lt;sup&gt;a&lt;/sup&gt;</td>
<td>23.222&lt;sup&gt;a&lt;/sup&gt;</td>
<td>14.764&lt;sup&gt;a&lt;/sup&gt;</td>
<td>14.746&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Degrees of Freedom</td>
<td>226,50</td>
<td>175,50</td>
<td>46,50</td>
<td>22,46</td>
<td>26,50</td>
<td>2,28</td>
<td>2,22</td>
<td>2,22</td>
</tr>
<tr>
<td>P Values&lt;sup&gt;b&lt;/sup&gt;</td>
<td>&lt;.00001</td>
<td>&lt;.00001</td>
<td>.015046</td>
<td>.00001</td>
<td>&lt;.00001</td>
<td>.000001</td>
<td>.000009</td>
<td>.000009</td>
</tr>
</tbody>
</table>

<sup>a</sup>Significant at $\alpha = .05$.

<sup>b</sup>RUN STATPAC/DIST to get the p values.
Table 8. The results of example 2 using Kruskal-Wallis test.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>T₁ vs T₂</th>
<th>T₂ vs T₃</th>
<th>T₃ vs T₄</th>
<th>T₄ vs T₅</th>
<th>T₅ vs T₆</th>
<th>T₆ vs T₇</th>
<th>T₇ vs T₈</th>
<th>T₈ vs T₉</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = Σnᵢ</td>
<td>225</td>
<td>200</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Rᵢ</td>
<td>325  25100</td>
<td>1143.5  18956.5</td>
<td>584.5  690.5</td>
<td>399.5  875.5</td>
<td>364.5  910.5</td>
<td>476.5  798.5</td>
<td>761  514</td>
<td>514  761</td>
</tr>
<tr>
<td>T'</td>
<td>70.5512ᵃ</td>
<td>27.8846ᵃ</td>
<td>1.0617ᵃ</td>
<td>22.1017ᵃ</td>
<td>28.4833ᵃ</td>
<td>14.8667ᵃ</td>
<td>10.2335ᵃ</td>
<td>10.2335ᵃ</td>
</tr>
</tbody>
</table>

| P Values for T' | <.00001 | 0.00001 | 0.30263 | 0.00001 | 0.00001 | 0.00012 | 0.00138 | 0.00138 |

ᵃSignificant at α = 0.05.
Table 9. The results of example 2 using F test, Kruskal-Wallis test, Breslow's method and Bonferroni method.

<table>
<thead>
<tr>
<th>Method</th>
<th>Hypothesis</th>
<th>$T_1$ vs $T_2$ $T_1 - T_3$</th>
<th>$T_2$ vs $T_3$ $T_2$ vs $T_4$ $T_5$ vs $T_6$</th>
<th>$T_3$ vs $T_5$ $T_5$ vs $T_7$ $T_7$ vs $T_8$</th>
<th>$T_8$ vs $T_9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>and Bonferroni</td>
<td>$S^*$</td>
<td>74.8929</td>
<td>29.0374</td>
<td>1.2002</td>
<td>24.4947 30.8067 16.0582 10.7245 10.7245</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>2659.5745$^a$</td>
<td>49.5521</td>
<td>1.2317</td>
<td>28.6045 38.3626 16.6990 10.9237 10.9237</td>
</tr>
<tr>
<td>Kruskal-Wallis and</td>
<td>$T'$</td>
<td>70.5512$^a$</td>
<td>27.8846$^a$</td>
<td>1.0617</td>
<td>22.1017 28.4833 14.8667 10.2335 10.2335</td>
</tr>
<tr>
<td>Bonferroni</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P Values</td>
<td>$T'$</td>
<td>&lt; .00001</td>
<td>&lt; .00001</td>
<td>0.26708</td>
<td>&lt; .00001 0.00005 0.00005 0.00095 0.00095</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>&lt; .00001</td>
<td>&lt; .00001</td>
<td>0.30283</td>
<td>&lt; .00001 0.00001 0.00012 0.00138 0.00138</td>
</tr>
</tbody>
</table>

$^a$ Means significant for chi-square value with $\alpha = .05$.

$^b$ Means significant for chi-square value with $\alpha = .05/6 = .00833$.

$^c$ Means significant for F value with $\alpha = .05$. 26
From Table 9 we see that for this example $S^{**}$ was a lower bound to $S^*$ and $S$. The statistic $T'$ was between $S^{**}$ and $S^*$, sometimes it was close to $S^{**}$ and sometimes close to $S^*$. In this example the $p$ value for $R$ was the smallest value for all of the cases, whereas the $T'$ value of Kruskal-Wallis test had the biggest probability. This suggests that the $R$ test may be too liberal (too easy to reject) while the Kruskal-Wallis test (corrected for ties) may be too conservative. Since the true population values are unknown for this case, it is impossible to say for certain.
CHAPTER V

CONCLUSIONS

A fundamental problem in many biological and medical investigations is a comparison of the survival distributions from two or more samples of censored data. The hypothesis of interest is the equality of survival time distribution functions across samples. In this paper we discussed this topic and analyzed censored data by using four different methods, namely: the F test, the Kruskal-Wallis test, Breslow's generalization of the Kruskal-Wallis test and a Bonferroni multiple comparison method. The F test is restricted to two exponential distributions, so it cannot be used widely. The Kruskal-Wallis test is suitable for two or more populations, but for censored data there will usually be a lot of ties. This violates the assumptions made in developing the Kruskal-Wallis test. For the Bonferroni method we use a given value $\alpha$ to do the $K$ multiple comparisons, then for each single case will only use $1 - \alpha/K$ to test the hypothesis. In this situation the given confidence interval is so large that it is hard to reject the null hypothesis. If a computer is available, we can translate the formulae of Breslow's method to a computer program as given in the Appendix. It will then be easy to analyze censored data. For all the reasons stated above, we prefer to use Breslow's method if a computer is available. If not, the Kruskal-Wallis procedure corrected for ties seemed to give almost the same results. There are only two examples in this paper; if
we want to get more information to tell the exact differences between the Kruskal-Wallis test and Breslow's method, more examples and a computer program for the Kruskal-Wallis test should be developed.

A Monte Carlo study could give enough different situations involving different distributions and different values of the parameters to help decide on the best overall procedure. A more exact multiple comparison procedure could also be developed using the asymptotical distribution of Breslow's vector of $W_i$'s rather than using the Bonferroni inequality. Since the Bonferroni method seems to work fairly well in these examples, further work might not be terribly worthwhile. Further Monte Carlo research could help in the decision on whether to pursue this matter in more detail.
REFERENCES


APPENDIXES
WORKFILE: PROG (12/10/81) PROG (1/14/82)

100 $SET LINEINFO AUTOBIND
200 $BIND# FROM IMSL/
300 C* THE IMS (INTERNATIONAL MATHEMATICAL AND STATISTICAL
400 C* LIBRARY CONSISTS OF A SUBSTANTIAL COLLECTION
500 C* OF SUBROUTINES AND FUNCTIONS SUBPROGRAMS IN THE AREAS OF
600 C* MATHEMATICS AND STATISTICS,
700 FILE $KIND=DISK, FILETYPE=F*
800 C*
900 C* $BIOMETRIKA (1970), 73, P. 579
1000 C* A GENERALIZED KRUSKAL-WALLIS TEST FOR COMPARING
1100 C* K SAMPLES SUBJECT TO UNEQUAL PATTERNS OF
1200 C* CENSORSHIP, BY NORMAN BRESLOW
1300 C* FORTRAN PROGRAM WRITTEN FOR THE BURROUGHS 6600
1400 C* COMPUTER AT USU BY HWEI-SHENG SHIH
1500 C* IN PARTIAL
1600 C* FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
1700 C* OF MASTER OF SCIENCE IN APPLIED STATISTICS AT
1800 C* UTAH STATE UNIVERSITY, 1981
1900 C*
2000 C* DIMENSION N(10), W(10), X(10,200), D(10,200), SGE(10,10),
2100 C* $SIG(10,10), M1(10), D1(10), Y(10,10), X(10,10),
2200 C* N(10), X(10,200), D(10,200), ND(10), INDX(2)
2300 C* THE DATA ARE ENTERED TO THE DATA FILE BY FOLLOWING STEPS:
2400 C* 1. ENTER K, K IS THE NUMBER OF SAMPLES
2500 C* 2. ENTER N(1), N(1) IS THE NUMBER OF OBSERVATIONS
2600 C* OF THE FIRST SAMPLE
2700 C* 3. ENTER PAIRS DATA X(1,1), D(1,1) X(1,2), D(1,2)
2800 C* 4. ENTER N(2), N(2) IS THE NUMBER OF OBSERVATIONS
2900 C* OF THE SECOND SAMPLE
3000 C* 5. ENTER PAIRS DATA X(2,1), D(2,1) X(2,2), D(2,2)
3100 C*........X(2,N(2)), D(2,N(2))
3200 C* 6. ENTER N(K), N(K) IS THE NUMBER OF OBSERVATIONS
3300 C* OF THE LAST SAMPLE
3400 C* 7. ENTER PAIRS DATA X(K,1), D(K,1) X(K,2), D(K,2)
3500 C*........X(K,N(K)), D(K,N(K))
3600 C* READ(5,/) K
3700 C* NTBK
3800 C* DO 50 I=1,K
3900 C* READ(5,/) N(I)
4000 C* NT=N+1
4100 C* READ(5,/) (X(I,J), D(I,J), J=1, N(I))
4200 C* 50 CONTINUE
4300 C** DENOTE BY X(I,J) THE TRUE OBSERVATION FOR THE (I,J)TH
4400 C** INDIVIDUAL IN THE (I)TH SAMPLE (J=1,...,N(I), I=1,...,K),
4500 C** SINCE THIS OBSERVATION MAY BE CENSORED BY A VARIABLE
4600 C** Z(I,J), IT CANNOT ALWAYS BE OBSERVED, RATHER ONE
4700 C** OBSERVES X(I,J)=MIN(XO(I,J),Z(I,J))
4800 C** ALONG WITH THE INDICATOR VARIABLE
4900 C**
5900 C** D(I,J)=1 IF X(I,J)="XO(I,J)
6000 C** D(I,J)=0 IF X(I,J)="XO(I,J)
6100 C** N(I)=# OBSERVATION OF EACH SAMPLE
6200 C** NT=NT+3
6300 WRITE(6,101)
6400 WRITE(6,100) NT,NT1
6500 100 FORMAT(1X, NT=THE TOTAL SAMPLE SIZE=I5,/,NT1=THE CUBIC OF TOTAL SAMPLE SIZE=I5)
6600 101 FORMAT(1X,///,50(I4,/,)
6700 C** P1 IS THE FIRST SUBROUTINE AND IT DEFINE A SCORING
7100 C** FUNCTION P81, THEN COMPUTE THE VECTOR w(I),
7200 C** CALL P1(K,N,W,X,D)
7300 C** P2 IS THE SECOND SUBROUTINE AND IT COMPUTE THE
7400 C** INDIVIDUAL TERMS OF A COVARIANCE MATRIX SIGMA,
7500 C** CALL P2(K,N,NT1,W,D,SIGE,SIG)
7600 C** P3 IS THE THIRD SUBROUTINE AND IT COMPUTE THE
7700 C** STATISTICS S AND SS,
7800 C** CALL P3(K,N,NT1,W,D,SIGE1,SIGE2)
7900 C** P4 IS THE FOURTH SUBROUTINE AND IT COMPUTE THE
8000 C** STATISTIC S2,
8100 C** CALL P4(K,N,NT1,W,D,SIGE2,SIGE3)
8200 C** MDCCH IS AN INSL SUBROUTINE WHICH IS USED TO GET
8300 C** THE PROBABILITY OF A CHI-SQUARE DISTRIBUTION,
8400 C** THE FORM IS CALL MDCCH(CS,DF,P,IER),
8500 C** CS = INPUT VALUE FOR WHICH THE PROBABILITY IS
8600 C** COMPUTED,
8700 C** DF = INPUT NUMBER OF DEGREES OF FREEDOM OF THE
8800 C** CHI-SQUARE DISTRIBUTION,
8900 C** P = OUTPUT PROBABILITY THAT A RANDOM VARIABLE
9000 C** WHICH FOLLOWS THE CHI-SQUARE DISTRIBUTION
9100 C** WITH OF DEGREES OF FREEDOM IS LESS THAN OR
9200 C** EQUAL TO CS,
9300 C** IER = ERROR PARAMETER,
9400 C** CALL MDCCH(S3,K=1,P3,IER)
9500 WRITE(4,111)
9600 111 FORMAT(1X,///,----- TESTING HYPOTHESIS -----/)
12000 10a FORMAT(1X,10H1) STAT. AS A LOWER BOUND TO STAT. S AND STAT. S* WILL BE
12010 * 10B 10C 10H ASYMMETRICALLY EQUIVALENT STATISTICS,!
12020 * 10D 10E 10F S* IS COMPUTATIONALLY SIMPLER THAN S,!
12030 * 10G 10H 10I HOWEVER, ONLY S WILL BE AS ASYMMETRICALLY!
12040 * 10J 10H 10I VALID STATISTIC UNDER HYPOTHESIS,!
12050 WRITE(b,101)
12060 IF(( L(2),3) GO TO 500
12070
12080 C C C THIS PART USE BONFERRONI MULTIPLE COMPARISON METHOD
12090 C C C NPWC=NUMBER OF PAIRWISE COMPARISON
13000 C C C ALPHA=CONFIDENT COEFFICIENT
13010 C C INDX(1) AND INDX(2) ARE THE TWO GROUPS WHICH
13020 C C WANT TO COMPARE
13030 C C C
13040 READ(5,/) NPWC,ALPHA
13050 DO 300 KK=1,NPWC
13060 READ(5,/) INDX(1),INDX(2)
13070 WRITE(b,118)
13080 140 FORMAT(1X,5(1H1))
13090 WRITE(b,119) INDX(1),INDX(2)
14000 150 FORMAT(1X,15I7) ===== 'I',II,' VS 'I',II,' =====/',
14100 NR=0
14200 200 DO 200 =1,2
14300 ND(INDX(I),J)=0
14400 00 200 J=1,N(INOX(I))
14500 DX(I,J)=0
14600 500 1=1,2
14700 NT=NT+N(INOX(I))
14800 NT1=NT1+1
14900 WRITE(b,100) NT,NT1
15000 CALL Pt(2,NO,w,ox,oa)
15100 CALL P2(2,ND,NT1,w,ox,oa)
15200 CALL P3(2,NO,NT,NT1,w,w1,dl,dx,d0,s3,s2)
15300 CALL P4(2,NT1,YY,XX,sig,ox,oa)
15400 WRITE(b,111)
15500 CALL MDCH(S,1,PS,J,J)
15600 WRIT(b,102) 53
15700 WRITE(b,107) S3,PS3
15800 WRITE(b,108) ALPHA,NPWC,ALPHA,NPWC
15900 WRITE(b,110) /// BONFERRONI CRITICAL
16000 107 FORMAT(1X,2X,1 P CHI SQUARE(1) =\',F9.6,') =\',F8.6)
16100 CALL MDCH(S3,1,PS2,IER)
16200 PS2=1.0PS2
16300 WRITE(b,109) PS2
16400 WRITE(b,109) S3,PS3
16500 WRITE(b,109) S3,PS3
16600 WRITE(b,109) S3,PS3
16700 TEST=1.0PC
16800 WRITE(b,109) TEST,PC
16900 WRITE(b,109) TEST,PC
17000 WRITE(b,109) TEST,PC
17100 125 FORMAT(1X, ' REJECT ',F8.6, ' IF ',F8.6, '<',F5.2,'/',',I1,1,1=',F8.6)
17200 * '/ AS ASSUMES ALPHA=',F5.2,'/','I2', 'PAIRWISE COMPARISONS'),
17300 * ', 'CHI-SQUARE FOR ID', (1 =',F5.2,'/',',I1,1,1=',F8.6)
17400 * ', 'ALPHA=NPWC,ALPHA=NPWC
17500 108 FORMAT(1X, ' BONFERRONI CRITICAL VALUE',
17600 * '/ ',AS ASSUMES ALPHA=',F5.2,'/','I2', 'PAIRWISE COMPARISONS'),
17700 * ', 'CHI-SQUARE FOR ID', (1 =',F5.2,'/',',I1,1,1=',F8.6)
17800 * ', 'ALPHA=NPWC,ALPHA=NPWC
17900 Ps1=ALPHA/NO
18000 WRITE(b,109) Ps1
18100 C * MDCH IS AN IMSL SUBROUTINE WHICH IS USED TO GET
18200 C THE INVERSE VALUE OF A CHI-SQUARE DISTRIBUTION,
THE FORM IS: CALL MDCHI(P,DF,X,IER)

P = INPUT PROBABILITY,

DF = INPUT NUMBER OF DEGREES OF FREEDOM,

X = OUTPUT CHI-SQUARE VALUE, SUCH THAT A RANDOM

VARIABLE, DISTRIBUTED AS CHI-SQUARE WITH DF

DEGREES OF FREEDOM, WILL BE LESS THAN OR

EQUAL TO X WITH PROBABILITY P,

IER = ERROR PARAMETER,

CALL MDCHI(P,DF,X,IER)

WRITE(6,101) CHI,P

300 CONTINUE

STOP

END

K = THE NUMBER OF SAMPLES

N(I) = THE NUMBER OF OBSERVATIONS OF THE ITH SAMPLE

X(I,J) AND D(I,J) =

IF X(I,J) IS AN UNCENSORED DATA THEN D(I,J)=1

IF X(I,J) IS A CENSORED DATA THEN D(I,J)=0,

SUBROUTINE PICK,N,w,x)

THIS SUBROUTINE DEFINE A SCORING FUNCTION PSI, THEN

WE DEFINE THE SCORING FUNCTION PSI FROM EQUATION (3)

OF BRESLOW FOR COMPARING TWO OBSERVATION X(I,J)

AND X(I',J') BY

THE ITH COMPONENT, W(I), OF THE VECTOR SCORE

STATISTIC IS DEFINED TO BE THE TOTAL SCORE

COMPARING THE ITH SAMPLE WITH THE REMAINING

SAMPLES,

W(I) = SUM

1=1, K

OF PSI(X(I,J), DELTA(I,J))

X(I',J'), DELTA(I',J')

DIMENSION W(10),N(10),X(10,200),D(10,200)

WRITE(6,140)
24200 IF(T1, EQ, 0) GO TO 130
24300 IF(T2, EQ, 0) GO TO 10
24400 IF(X(I,J), LT, X(IP,JP)) GO TO 20
24500 IF(X(I,J), GT, X(IP,JP)) GO TO 30
24600 GO TO 150
24700 10 IF(D(I,J), GT, D(IP,JP)) GO TO 40
24800 IF(X(I,J), GT, X(IP,JP)) GO TO 30
24900 GO TO 150
25000 40 IF(X(I,J), LT, X(IP,JP)) GO TO 20
25100 GO TO 150
25200 20 W(I)=W(I)+1
25300 GO TO 130
25400 30 W(I)=W(I)+1
25500 CONTINUE
25600 120 CONTINUE
25700 WRITE(6,102) I,W(I)
25800 102 FORMAT(1X,I,12,1,F7,0)
25900 SUM=SUM+W(I)
26000 110 CONTINUE
26100 WRITE(6,135) SUM
26200 135 FORMAT(1X, THE SUM OF W(I) EQUAL ',F2,0)
26300 RETURN
26400 END
26500 C K = THE NUMBER OF SAMPLES
26600 C N = THE NUMBER OF OBSERVATIONS OF THE ITH SAMPLE
26700 C NT1 = THE CUBIC OF TOTAL SAMPLE SIZE
26800 C W(I) = THE VECTOR SCORE STATISTIC
26900 C XI,J AND D(I,J) = SOURCE DATA
27000 SUBROUTINE P2CK,N,NT1,W,X,O,SIG)
27100 THIS SUBROUTINE COMPUTE THE INDIVIDUAL TERMS
27200 C OF A COVARIANCE MATRIX SIGMA
27300 C SIG(I,I), IN COVARIANCE MATRIX SIGMA
27400 C MAY BE FOUND FROM THE FORMULAE (8)
27500 CONTINUE
27600 C OF BESLAW
27700 C
27800 C I UNEQUAL I'
27900 C (N-5)SIG(SIG(I,I'))
28000 C = SIG(I,I')+(SUM J=1,N(I) OF DELTA(I',J)')
28100 C = SIG(I,I')-(SUM J=1,N(I) OF E(X(I,J)*X(I,J')))
28200 C = E(X)*I IF X < 0
28300 C = E(X)*0 IF X = 0
28400 C = (SIG(I,I')=SUM N I UNEQUAL I OF SIG(I,I'))
28500 C SIG(I,I')=SUM N I UNEQUAL I OF SIG(I,I')
28600 C DIMENSION N(10),W(10),X(10,200),D(10,200),SIGE(10,10),
28700 SIG(10,10)
28800 29000 DO 210 I=1,N
28900 29100 DO 210 J=1,N
29000 IF((I,J), EQ, 0) GO TO 210
29100 DO 220 I=1,N
29200 29300 IF((I,J), EQ, 0) GO TO 210
29400 DO 220 I=1,N
29500 29600 DO 220 J=1,N
29700 29800 IF((I,J), EQ, 0) GO TO 220
29900 EPS1=0
30000 DO 230 J=1,N
30100 IF(X(I,J), LT, X(IP,JP)) GO TO 230
30200 EPS1=EPS1+1
30300 210 CONTINUE
30400 220 CONTINUE
30500 230 CONTINUE
CONTINUE
EPS2=0
DO 240 JPI1,N(IP)
IF((X(I,JPI1),LE,X(I,IPP,JPI1)) GO TO 240
EPS2=EPS2+1
CONTINUE
EPS=EPS+EPS2
TOTAL=TOTAL+EPS
CONTINUE
SIGE(I,IP)=TOTAL/NT1
00 2
DO 200 JP=1,N(IP)
IF(DCIP,JP),EQ,O) GO TO 200
SIGE(I,IP)=SIGE(I,IP)
CONTINUE
SIGE(I,IP)=SIGE(I,IP)
CONTINUE
WRITE(6,204)
WRITE(6,205) (SIGE(I,J), =1,K)
RETURN
DIMENSION W(10),W1(10),0L(10),X(10,200),D(10,200)
REAL N(10),NT
S=0
DO 310 IP=1,K
00 310 JP=1,N(IP)
IF(DCIP,JP),EQ,O) GO TO 310
EPS:O
DO 320
J=1,K
DO 20
J=1,NCO
If(X(I,J),L ,XCIP,JP)) GO TO 320
EPS=EPS+l
320 CONTINUE
CONTINUE
WRITE(6,204)
WRITE(6,205) (SIGE(I,J), =1,K)
RETURN
C* K = THE NUMBER OF SAMPLES
C* NT = THE NUMBER OF OBSERVATIONS OF THE (I)TH SAMPLE
C* NT1 = THE CUBIC OF TOTAL OBSERVATIONS
C* W(I) = THE VECTOR SCORE STATISTIC
C* X(I,J) AND D(I,J) = SOURCE DATA
SUBROUTINE P3(Kx,NT,NT1,K1,DL,X,D)
C* S* IN ORD R TO CHECK ON COMPUTATIONS,
C* S* IN ORDER TO CHECK ON COMPUTATIONS,
C* (SUM I=1,K OF W(I)**2/LAMBD(A(I))
C* S* IN ORDER TO CHECK ON COMPUTATIONS,
36400  SA=D(IP,J)*EPS**2
36500  SB=SB+SA
36600  310 CONTINUE
36700  SC=0
36800  DO 330 I=1,K
36900   DL(I)=N(I)/NT
37000   M(I)=M(I)**(1/2)
37100   SC=SC+M(I)**2/DL(I)
37200  330 CONTINUE
37300  S3=SC/NT1
37400  S2=SC/S8
37500  RETURN
37600  END
37700  C= K = THE NUMBER OF SAMPLES
37800  C=
37900  C= SIGMA = THE COVARIANCE MATRICE
38000  C= N(I) = THE VECTOR SCORE STATISTIC
38100  SUBROUTINE P4(K,NT1,YV,XX,S4GE,Xt,v1,Y2,w,4Ns,v,s,TOT)
38200  C= THIS SUBROUTINE COMPUTE THE STATISTIC S.
38300  C= A GENERALIZED KRUSKAL-WALLIS TEST FOR COMPARING
38400  C= K SAMPLES FROM P,583 OF BRESLOW.
38500  C= K** SIGMA MATRIX HAS RANK K=1 PROVIDED THAT EACH OF
38600  C= THE K SAMPLES CONTAINS AT LEAST ONE UNCEBREOED
38800  C= OBSERVATION, FOR SUCH SIGMA THERE EXIST K=1 VECTOR
38900  C= (X(I)=XX(I,1),..., XX(I,K))! (I=1,..., K=1) SUCH
39000  C= THAT X(I)**SIGMA(X(J)) EQUALS ONE OR ZERO ACCORDING
39100  C= AS I AND J ARE EQUAL OR UNEQUAL, CONSEQUENTLY THE
39300  C= STATISTICS
39400  C= (S(I)=N**(.3/2) . (X(I)**W))
39500  C= WHERE W=(X(1),..., X(K))!
39600  C= WILL BE ASYMPTOTICALLY UNCORRELATED WITH MEAN 0
39700  C= AND UNIT VARIANOE, N**(.3/2) HAS ASYMPTOTICALLY
39800  C= A MULTIDIMENSIONAL NORMAL DISTRIBUTION, FROM
39900  C= THIS IT FOLLOWS THAT
40000  C= SUM I=1, K=1 OF (S(I)**2)
40200  C= IS ASYMPTOTICALLY DISTRIBUTED IN A CHI-SQUARED
40300  C= DISTRIBUTION WITH K=1 DEGREES OF FREEDOM, THE
40400  C= STATISTICS S WILL BE USED TO TEST THE HYPOTHESIS.
40500  C=
40600  C= IN PRACTICE, THE X(I) ARE FOUND BY MEANS OF THE
40700  C= GRAM-SCHMIDT ORTHOMOINALIZATION PROCESS WITH THE
40800  C= INNER PRODUCT OF TWO VECTORS A AND B BY
40900  C= A.GP(A(SIGMA)B, THE K=1 VECTORS
41000  C= Y(1)=(1,0,0,0,0,0,0,0,0,0,0), Y(2)=(1,1,0,0,0,0,0,0,
41100  C= Y(K-1)=(1,...,1,0)
41200  C= MAY BE USED AS A STARTING POINT FOR THE GRAM-
41300  C= SCHMIDT PROCESS.
41400  C=
41500  C= GRAM-SCHMIDT PROCESS
41600  C= 1, X(1)Y(1)//X(1)Y(1), Y(1)//=SRTCXY(1)Y(1))
41700  C= 2, Z(2)Y(2)=Y(2),X(1)X(1)
41800  C= 3, X(2)Z(2)//X(2)Z(2), /X(2)//=SRTCXY(2)Z(2))
41900  C= 4, X(3)Z(3)//X(3)Z(3), /X(3)//=SRTCXY(3)Z(3))
4200  C= 5, Z(4)Y(4)=Y(4)K(4),X(1)X(1)
42100  C= 6, Z(5)Y(5)=Y(5)K(5),X(1)X(1)
42200  C= 7, Z(6)Y(6)=Y(6)K(6),X(1)X(1)
42300  C= 8, Z(8)Y(8)=Y(8)K(8),X(1)X(1)
42400  C= 9, Z(9)Y(9)=Y(9)K(9),X(1)X(1)
FUNCTION VAL(K,Y1,Y2,ANS,SIGE)
C*  CALCULATE THE INNER PRODUCT OF TWO VECTORS A AND B

DIMENSION Y1(10),Y2(10),ANS(10),SIGE(10,10)

* BY * A, B = A' (SIGMA) B

REAL NT1

AN=0.0

DO 410 I=1,K
  DO 420 J=1,K
    VV(I,J)=0
  CONTINUE

410 CONTINUE

DO 420 J=1,K
  Y2(J)=YY(I,J)
  CONTINUE

VAL(K,Y1,Y2,ANS,SIGE)

V(N)=VA

CONTINUE

DO 450 L=1,I
  DO 460 J=1,K
    Y1(J)=Y1(J)*XX(L,J)
  CONTINUE

450 CONTINUE

CONTINUE

460 CONTINUE

DEN=SQRT(VA)

CONTINUE

S(I)=SUM

CONTINUE

470 CONTINUE

480 CONTINUE

RETURN END
48600 460 ANS(M)=ANS(M)+Y1(J)*SIGE(J,M)
48700   VAL=0
48800   DO 470 I=1,K
48900   470   VAL=VAL+ANS(I)*Y2(I)
49000   RETURN
49100   END
49200   C*NT1=THE CUBIC OF TOTAL OBSERVATIONS
<table>
<thead>
<tr>
<th>Time</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>3</td>
</tr>
<tr>
<td>200</td>
<td>15</td>
</tr>
<tr>
<td>300</td>
<td>1,6162,1</td>
</tr>
<tr>
<td>400</td>
<td>3,3108,1</td>
</tr>
<tr>
<td>500</td>
<td>6,3982,1</td>
</tr>
<tr>
<td>600</td>
<td>8,7001,1</td>
</tr>
<tr>
<td>700</td>
<td>9,5813,1</td>
</tr>
<tr>
<td>800</td>
<td>10,3311,1</td>
</tr>
<tr>
<td>900</td>
<td>11,3729,1</td>
</tr>
<tr>
<td>1000</td>
<td>11,6698,1</td>
</tr>
<tr>
<td>1100</td>
<td>14,7173,1</td>
</tr>
<tr>
<td>1200</td>
<td>20,0</td>
</tr>
<tr>
<td>1300</td>
<td>20,0</td>
</tr>
<tr>
<td>1400</td>
<td>20,0</td>
</tr>
<tr>
<td>1500</td>
<td>20,0</td>
</tr>
<tr>
<td>1600</td>
<td>20,0</td>
</tr>
<tr>
<td>1700</td>
<td>20,0</td>
</tr>
<tr>
<td>1800</td>
<td>12</td>
</tr>
<tr>
<td>1900</td>
<td>0,3113,1</td>
</tr>
<tr>
<td>2000</td>
<td>0,3193,1</td>
</tr>
<tr>
<td>2100</td>
<td>0,3793,1</td>
</tr>
<tr>
<td>2200</td>
<td>0,9393,1</td>
</tr>
<tr>
<td>2300</td>
<td>1,4585,1</td>
</tr>
<tr>
<td>2400</td>
<td>2,6174,1</td>
</tr>
<tr>
<td>2500</td>
<td>2,9732,1</td>
</tr>
<tr>
<td>2600</td>
<td>2,5377,1</td>
</tr>
<tr>
<td>2700</td>
<td>12,9082,1</td>
</tr>
<tr>
<td>2800</td>
<td>20,0</td>
</tr>
<tr>
<td>2900</td>
<td>20,0</td>
</tr>
<tr>
<td>3000</td>
<td>20,0</td>
</tr>
<tr>
<td>3100</td>
<td>18</td>
</tr>
<tr>
<td>3200</td>
<td>1,9777,1</td>
</tr>
<tr>
<td>3300</td>
<td>1,5408,1</td>
</tr>
<tr>
<td>3400</td>
<td>1,7251,1</td>
</tr>
<tr>
<td>3500</td>
<td>2,1613,1</td>
</tr>
<tr>
<td>3600</td>
<td>2,8669,1</td>
</tr>
<tr>
<td>3700</td>
<td>3,0430,1</td>
</tr>
<tr>
<td>3800</td>
<td>3,1428,1</td>
</tr>
<tr>
<td>3900</td>
<td>3,5533,1</td>
</tr>
<tr>
<td>4000</td>
<td>3,9252,1</td>
</tr>
<tr>
<td>4100</td>
<td>4,2688,1</td>
</tr>
<tr>
<td>4200</td>
<td>4,3705,1</td>
</tr>
<tr>
<td>4300</td>
<td>4,7837,1</td>
</tr>
<tr>
<td>4400</td>
<td>7,7971,1</td>
</tr>
<tr>
<td>4500</td>
<td>7,9731,1</td>
</tr>
<tr>
<td>4600</td>
<td>8,2049,1</td>
</tr>
<tr>
<td>4700</td>
<td>9,2274,1</td>
</tr>
<tr>
<td>4800</td>
<td>11,2949,1</td>
</tr>
<tr>
<td>4900</td>
<td>20,0</td>
</tr>
<tr>
<td>5000</td>
<td>5,0,05</td>
</tr>
<tr>
<td>5100</td>
<td>1,2</td>
</tr>
<tr>
<td>5200</td>
<td>1,3</td>
</tr>
<tr>
<td>5300</td>
<td>2,3</td>
</tr>
</tbody>
</table>
THE TOTAL SAMPLE SIZE = 45
THE CUBIC OF TOTAL SAMPLE SIZE = 9125

VECTOR SCORE STATISTIC

\[ W(1) = 258, \]
\[ W(2) = -119, \]
\[ W(3) = 139, \]
THE SUM OF W(I) EQUAL 0,

SIGMA MATRIX

\[
\begin{pmatrix}
0.0761 & 0.0272 & 0.0489 \\
0.0272 & 0.0504 & 0.0232 \\
0.0489 & 0.0232 & 0.0721 \\
\end{pmatrix}
\]

GRAM SCHMIDT ORTHOGONALIZATION VECTORS

\[
\begin{pmatrix}
1 \\
1.2437 \\
0.0000 \\
0.0000 \\
\end{pmatrix}
\]

TESTING HYPOTHESIS

\[ S** = 9,912700 \]
\[ P(\text{CHI-SQUARE}) \geq 9,912700 \Rightarrow 0.007039 \]
\[ S** = 10,352355 \]
\[ P(\text{CHI-SQUARE}) \geq 10,352355 \Rightarrow 0.005650 \]
\[ S** = 9,787615 \]
PLCHI~SQUARE(2) = 9.787615 \approx 0.007493

S** AS A LOWER BOUND TO S*, S AND S** WILL BE
ASYMPTOTICALLY EQUIVALENT STATISTICS,
S* IS COMPUTATIONALLY SIMPLER THAN S,
HOWEVER, ONLY S WILL BE AN ASYMPTOTICALLY
VALID STATISTIC UNDER HYPOTHESIS.

************

************

----- 1 VS 2 ----- 

NT = THE TOTAL SAMPLE SIZE = 27
NT1 = THE CUBIC OF TOTAL SAMPLE SIZE = 19683

----- VECTOR SCORE STATISTIC ----- 

W(1) = e2,
W(2) = -e2,
THE SUM OF W(I) EQUAL 0.

----- SIGMA MATRIX ----- 

0.0670 0.0670
0.0670 0.0670

----- GRAM-SCHMIDT ORTHOGONALIZATION VECTORS ----- 

1
3.16229
0.00000

----- TESTING HYPOTHESIS ----- 

S*** = 4.150617
P(Chi-Square(1) \geq 4.150617) \approx 0.041619
$s = 4.540970$

$P(\text{CHI}^2(1) \geq 4.540970) = 0.033093$

$s = 5.097801$

$P(\text{CHI}^2(1) \geq 5.097801) = 0.023956$

Using Bonferroni inequality, reject $0.023956 < 0.05/3 = 0.016667$

$s^*$ as a lower bound to $s$, $s$ and $s^*$ will be asymptotically equivalent statistics. $s^*$ is computationally simpler than $s$. However, only $s$ will be an asymptotically valid statistic under hypothesis.

Bonferroni critical value
(Assumes alpha = 0.05, 3 pairwise comparisons)
$\text{CHI}^2$ for 1 df, $(1 = 0.05/3)100(\text{TH})$

$P(\text{CHI}^2(1) \leq 5.737029) = 0.983333$

******************************************************************************

******************************************************************************

----- 1 vs 3 -----  

N = The total sample size = 33
N1 = The cubic of total sample size = 35937

----- Vector score statistic -----  

$W(1) = 176$

$W(2) = -176$

The sum of $W(1)$ equal 0.

----- Sigma matrix -----  

0.0734 0.0734
0.0734 0.0734
--- GRAM-SCHMIDT ORTHOGONALIZATION VECTORS ---

1
3.69161
0.00000

--- TESTING HYPOTHESIS ---

$S** = 10.429630$
$P(\chi^2(1) \geq 10.429630) = 0.001240$

$S** = 11.008594$
$P(\chi^2(1) \geq 11.008594) = 0.000000$

$S = 11.74682$
$P(\chi^2(1) \geq 11.74682) = 0.000010$

Using Bonferroni inequality reject $0.000010$ if $0.000010 < 0.05/3 = 0.01667$

$S^*$ as a lower bound to $S$, $S$ and $S^*$ will be asymptotically equivalent statistics, $S^*$ is computationally simpler than $S$. However, only $S$ will be an asymptotically valid statistic under hypothesis.

Bonferroni critical value
(assumes $\alpha = 0.05$, 3 pairwise comparisons)
$P(\chi^2$ for 10f, $1 = 0.05/3)100(\%)$
$P(\chi^2(1) \leq 5.737029) = 0.003333$

**************************

**************************

--- 2 VS 3 ---

$N_1$: THE TOTAL SAMPLE SIZE = 30
$N_1$: THE CUBIC OF TOTAL SAMPLE SIZE = 27000
VECTOR SCORE STATISTIC

\[ W(1) = -37, \]
\[ W(2) = 37, \]

The sum of \( W(1) \) equal 0.

SIGMA MATRIX

\[
\begin{bmatrix}
0.0694 & 0.0694 \\
0.0694 & 0.0694 \\
\end{bmatrix}
\]

GRAM-SCHMIDT ORTHOGONALIZATION VECTORS

1
3.79676
0.00000

TESTING HYPOTHESIS

\[
P(\chi^2(1)) = 0.633796 \]

\[
P(\chi^2(1)) = 0.667857 \]

\[
P(\chi^2(1)) = 0.730913 \]

Using Bonferroni inequality

Reject 0.392587 if 0.392587 < 0.05/3 = 0.016667

AS A LOWER BOUND TO \( S^* \), \( S^* \) AND \( S_{\alpha} \) WILL BE ASYMPTOTICALLY EQUIVALENT STATISTICS.

\( S^* \) IS COMPUTATIONALLY SIMPLER THAN \( S^* \).

HOWEVER, ONLY \( S^* \) WILL BE AN ASYMPTOTICALLY VALID STATISTIC UNDER HYPOTHESIS.

BONFERRONI CRITICAL VALUE
(ASSUMES ALPHAS = 0.05, 3 PAIRWISE COMPARISONS)

\[
P(\chi^2 = 5.737029) = 0.016667
\]
NT = THE TOTAL SAMPLE SIZE = 225
NT1 = THE CUBIC OF TOTAL SAMPLE SIZE = 11390625

VECTOR SCORE STATISTIC

\[ W(1) = 5000, \]

\[ W(2) = 5000, \]

THE SUM OF \( W(1) \) EQUAL 0,

SIGMA MATRIX

\[
0.0008 \quad -0.0008
\]

\[
-0.0008 \quad 0.0008
\]

GRAM-SCHMIDT ORTHOGONALIZATION VECTORS

\[
1
\]

\[
34.61047
\]

\[
0.00000
\]

TESTING HYPOTHESIS

**\( S** = 66.666667 \)

\[ P(\text{CHI-SQUARE}(1)) = 66.666667 \approx 0.000000 \]

**\( S** = 74.892909 \)

\[ P(\text{CHI-SQUARE}(1)) = 74.892909 \approx 0.000000 \]

\[ S = 2659.574468 \]

\[ P(\text{CHI-SQUARE}(1)) = 2659.574468 \approx 0.000000 \]

**\( S** AS A LOWER BOUND TO \( S \), \( S \) AND \( S* \) WILL BE
ASYMPTOTICALLY EQUIVALENT STATISTICS,

\( S* \) IS COMPUTATIONALLY SIMPLER THAN \( S \).
HOWEVER, ONLY S WILL BE AN ASYMPTOTICALLY VALID STATISTIC UNDER HYPOTHESIS.
NT=THE TOTAL SAMPLE SIZE= 200
NT=THE CUBIC OF TOTAL SAMPLE SIZE= 8000000

----- VECTOR SCORE STATISTIC -----
W(1)= 2738,
W(2)= 2738,
THE SUM OF W(i) EQUAL 0,

----- SIGMA MATRIX -----
0.0189 0.0189
0.0189 0.0189

----- GRAM-SCHMIDT ORTHOGONALIZATION VECTORS -----
1
7.27181
0.00000

----- TESTING HYPOTHESIS -----
S** 25.702779
P(Chi-Square( 1 ) >= 25.702779 ) = 0.000000

S* 29.037435
P(Chi-Square( 1 ) >= 29.037435 ) = 0.000000

S 49.552139
P(Chi-Square( 1 ) >= 49.552139 ) = 0.000000

S** AS A LOWER BOUND TO S*, S AND S* WILL BE
ASYMPTOTICALLY EQUIVALENT STATISTICS,
S* IS COMPUTATIONALLY SIMPLER THAN S,
However, only $S$ will be an asymptotically valid statistic under hypothesis.

***************
| Value | 100 | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 | 1000 | 1100 | 1200 | 1300 | 1400 | 1500 | 1600 | 1700 | 1800 | 1900 | 2000 | 2100 | 2200 | 2300 | 2400 | 2500 | 2600 | 2700 | 2800 | 2900 | 3000 | 3100 | 3200 | 3300 | 3400 | 3500 | 3600 | 3700 | 3800 | 3900 | 4000 | 4100 |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
**THE TOTAL SAMPLE SIZE = 225**

**THE CUBIC OF TOTAL SAMPLE SIZE = 11390625**

--- VECTOR SCORE STATISTIC ---

| \( w(1) \) | \( 5000 \) |
| \( w(2) \) | \( 2113 \) |
| \( w(3) \) | \( 1944 \) |
| \( w(4) \) | \( 1945 \) |
| \( w(5) \) | \( 1055 \) |
| \( w(6) \) | \( 1694 \) |
| \( w(7) \) | \( 3272 \) |
| \( w(8) \) | \( 1727 \) |
| \( w(9) \) | \( 3254 \) |

THE SUM OF \( w(i) \) EQUAL 0.

--- SIGMA MATRIX ---

\[
\begin{pmatrix}
0.0008 & 0.0001 & 0.0001 & 0.0001 & 0.0001 & 0.0001 & 0.0001 & 0.0001 & 0.0001 \\
0.0001 & 0.0230 & 0.0025 & 0.0027 & 0.0034 & 0.0035 & 0.0036 & 0.0035 & 0.0036 \\
0.0001 & 0.0025 & 0.0234 & 0.0029 & 0.0035 & 0.0035 & 0.0037 & 0.0035 & 0.0037 \\
0.0001 & 0.0027 & 0.0029 & 0.0251 & 0.0038 & 0.0039 & 0.0039 & 0.0038 & 0.0039 \\
0.0001 & 0.0034 & 0.0035 & 0.0347 & 0.0058 & 0.0062 & 0.0057 & 0.0062 & 0.0057 \\
0.0001 & 0.0035 & 0.0035 & 0.0058 & 0.0364 & 0.0068 & 0.0062 & 0.0067 & 0.0062 \\
0.0001 & 0.0036 & 0.0037 & 0.0039 & 0.0062 & 0.0068 & 0.0382 & 0.0066 & 0.0073 \\
0.0001 & 0.0035 & 0.0035 & 0.0057 & 0.0062 & 0.0066 & 0.0360 & 0.0066 & 0.0360 \\
0.0001 & 0.0036 & 0.0037 & 0.0039 & 0.0062 & 0.0067 & 0.0073 & 0.0066 & 0.0382 \\
\end{pmatrix}
\]

--- GRAM-SCHMIDT ORTHOGONALIZATION VECTORS ---
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>34.01047</td>
<td>0.82409</td>
<td>0.91310</td>
<td>1.01961</td>
<td>1.09893</td>
<td>1.37721</td>
<td>1.94955</td>
<td>3.29697</td>
</tr>
<tr>
<td></td>
<td>0.00000</td>
<td>6.59750</td>
<td>0.72631</td>
<td>0.85743</td>
<td>1.07114</td>
<td>1.36545</td>
<td>1.96068</td>
<td>3.27845</td>
</tr>
<tr>
<td></td>
<td>0.00000</td>
<td>0.00000</td>
<td>6.58051</td>
<td>0.82522</td>
<td>1.01374</td>
<td>1.36521</td>
<td>1.95909</td>
<td>3.28122</td>
</tr>
<tr>
<td></td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>6.41692</td>
<td>1.09559</td>
<td>1.36753</td>
<td>1.95578</td>
<td>3.28701</td>
</tr>
<tr>
<td></td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>5.54332</td>
<td>1.35728</td>
<td>1.97038</td>
<td>3.26213</td>
</tr>
<tr>
<td></td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>5.56225</td>
<td>1.97524</td>
<td>3.25345</td>
</tr>
<tr>
<td></td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>5.77519</td>
<td>3.24642</td>
</tr>
<tr>
<td></td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>6.76711</td>
</tr>
<tr>
<td></td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
</tbody>
</table>

Testing Hypothesis

$** = 154,753470$

$P(\text{CHI-SQUARE}(8) \geq 154,753470) = 0.000000$

$** = 173,849062$

$P(\text{CHI-SQUARE}(8) \geq 173,849062) = 0.000000$

$= 2766.360697$

$P(\text{CHI-SQUARE}(8) \geq 2766.360697) = 0.000000$

$**$ as a lower bound to $S$, $S$ and $S$ will be asymptotically equivalent statistics. $S$ is computationally simpler than $S$. However, only $S$ will be an asymptotically valid statistic under hypothesis.

**************

**************

3 vs 4

NOT THE TOTAL SAMPLE SIZE = 50
NOTHE CUBIC OF TOTAL SAMPLE SIZE = 125000
----- VECTOR SCORE STATISTIC -----  
\[ W(1) = 106, \]
\[ W(2) = 106, \]

THE SUM OF W(I) EQUAL 0.

----- SIGMA MATRIX -----  
\[
\begin{pmatrix}
0.0730 & 0.0730 \\
0.0730 & 0.0730 
\end{pmatrix}
\]

----- GRAM-SCHMIDT ORTHOGONALIZATION VECTORS -----  
\[ 1 \]
\[ 3,70177 \]
\[ 0,00000 \]

----- TESTING HYPOTHESIS -----  
\[ S** = 1,078056 \]
\[ P(\text{CHI-SQUARE}(1) \geq 1,078056) = 0.248998 \]
\[ S* = 1,200299 \]
\[ P(\text{CHI-SQUARE}(1) \geq 1,200299) = 0.273262 \]
\[ S = 1,231747 \]
\[ P(\text{CHI-SQUARE}(1) \geq 1,231747) = 0.267067 \]

USING BONFERRONI INEQUALITY  
REJECT 0.267067 IF 0.267067 < 0.05/6 = 0.008333

\[ S** \text{ AS A LOWER BOUND TO } S*, S \text{ AND } S# \text{ WILL BE} \]
\[ \text{ASYMPTOTICALLY EQUIVALENT STATISTICS,} \]
\[ S* \text{ IS COMPUTATIONALLY SIMPLER THAN } S, \]
\[ \text{HOWEVER, ONLY } S \text{ WILL BE AN ASYMPTOTICALLY} \]
\[ \text{VALID STATISTIC UNDER HYPOTHESIS.} \]

BONFERRONI CRITICAL VALUE  
(ASSUMED ALPHA = 0.05, 6 PAIRWISE COMPARISONS)  
\[ \text{CHI-SQUARE FOR 1DF, } (1 - 0.05/6)100(\%): \]
\[ P(\text{CHI-SQUARE}(1) \leq 6,981694) = 0.991667 \]
----- 3 VS 8 ----- 

\( n = \text{the total sample size} = 50 \) 
\( n_1 = \text{the cubic of total sample size} = 125000 \)

----- VECTOR SCORE STATISTIC ----- 
\( \text{W(1)} = -0.76 \) 
\( \text{W(2)} = 0.76 \) 
\( \text{The sum of W} = \text{equal 0} \)

----- SIGMA MATRIX ----- 
\[
\begin{bmatrix}
0,0634 & 0,0634 \\
0,0634 & 0,0634
\end{bmatrix}
\]

----- GRAM-SCHMIDT ORTHOGONALIZATION VECTORS ----- 

1 
3,97251 
0,00000

----- TESTING HYPOTHESIS ----- 

\( S** = 21.751296 \) 
\( P(\text{CHI-SQUARE}(1) > 21.751296) = 0.000003 \)

\( S** = 24.494703 \) 
\( P(\text{CHI-SQUARE}(1) > 24.494703) = 0.000001 \)

\( S = 28.604469 \)
**I (1) = 26.404469, 0.000000**

**USING BONFERRONI INEQUALITY**

**REJECT 0.000000 IF 0.000000 < 0.05/6 = 0.008333**

S** as a lower bound to S**, S AND S** will be asymptotically equivalent statistics. S** is computationally simpler than S**, however, only S will be an asymptotically valid statistic under hypothesis.

**BONFERRONI CRITICAL VALUE**

(assumes alpha = 0.05, 6 pairwise comparisons)

Chi-square for 10F, (1 = 0.05/6) 100(TH)%

**P(Chi-square(1) <= 0.981694 ) = 0.991667**

---------------------------------------

----- 4 VS 6 ----- 

**NT** = THE TOTAL SAMPLE SIZE = 50
**N** = THE CUBIC OF TOTAL SAMPLE SIZE = 125000

----- VECTOR SCORE STATISTIC ----- 

w(1) = -546.6
w(2) = 546.6

THE SUM OF W(I) EQUAL 0.

----- SIGMA MATRIX ----- 

0.0622 -0.0622
-0.0622 0.0622

----- GRAM-SCHMIDT ORTHOGONALIZATION VECTORS ----- 

1

4.01067
--- TESTING HYPOTHESIS ---

S** = 28.619136
P(Chi-Square(1) >= 28.619136) = 0.00000

S* = 30.006055
P(Chi-Square(1) >= 30.006055) = 0.00000

S = 38.362630
P(Chi-Square(1) >= 38.362630) = 0.00000

USING BONFERRONI INEQUALITY
REJECT 0.000000 IF 0.000000 < 0.05/6 = 0.008333

S** AS A LOWER BOUND TO S*, S AND S** WILL BE
ASYMPTOTICALLY EQUIVALENT STATISTICS.
S* IS COMPUTATIONALLY SIMPLER THAN S,
HOWEVER, ONLY S WILL BE AN ASYMPTOTICALLY
VALID STATISTIC UNDER HYPOTHESIS.

BONFERRONI CRITICAL VALUE
(ASSUMES ALPHA = 0.05, 6 PAIRWISE COMPARISONS)
#Chi-Square for 1DF, (1 = 0.05/6) = 0.100(TH) =
P(Chi-Square(1) <= 0.981694) = 0.091667

******************************************************************************

******************************************************************************

----- 5 VS 7 -----  

NT = THE TOTAL SAMPLE SIZE = 50
NT1 = THE CUBIC OF TOTAL SAMPLE SIZE = 125000

----- VECTOR SCORE STATISTIC -----  

w(1) = 322,
w(2) = 322,
THE SUM OF \( w(i) \) EQUAL 0.

----- SIGMA MATRIX -----

\[
\begin{pmatrix}
0.0497 & 0.0497 \\
0.0497 & 0.0497 \\
\end{pmatrix}
\]

----- GRAM-SCHMIDT ORTHOGONALIZATION VECTORS -----

1

\[
\begin{pmatrix}
0.46688 \\
0.00000 \\
\end{pmatrix}
\]

----- TESTING HYPOTHESIS -----

S* = 9.953664

\[ \text{CHI-SQUARE}(1) = 9.953664 \Rightarrow 0.001605 \]

S = 10.058234

\[ \text{CHI-SQUARE}(1) = 10.058234 \Rightarrow 0.000061 \]

S = 16.698985

\[ \text{CHI-SQUARE}(1) = 16.698985 \Rightarrow 0.000044 \]

USING BONFERRONI INEQUALITY

REJECT 0.000044 IF 0.000044 < 0.05/6 = 0.008333

S AS A LOWER BOUND TO S*, S AND S* WILL BE
ASYMPTOTICALLY EQUIVALENT STATISTICS,
S* IS COMPUTATIONALLY SIMPLER THAN S,
HOWEVER, ONLY S WILL BE AN ASYMPTOTICALLY
VALID STATISTIC UNDER HYPOTHESIS.

BONFERRONI CRITICAL VALUE

(ASSUMES ALPH = 0.05, 6 PAIRWISE COMPARISONS)

\[ \text{CHI-SQUARE} \text{ FOR 10F, } (1 = 0.05/6) \text{100(TH)} \]

\[ \text{CHI-SQUARE}(1) = 6.981694 \Rightarrow 0.991667 \]

******************************************************************************
NT = THE TOTAL SAMPLE SIZE = 50
NT1 = THE CUBIC OF TOTAL SAMPLE SIZE = 125000

VECTOR SCORE STATISTIC

W(1) = -247,
W(2) = -247,
The sum of W(1) equal 0.

SIGMA MATRIX

<table>
<thead>
<tr>
<th></th>
<th>0.0447</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0447</td>
<td>0.0447</td>
</tr>
</tbody>
</table>

GRAM-SCHMIDT ORTHOGONALIZATION VECTORS

1
4.73000
0.00000

TESTING HYPOTHESIS

S* = 5.856864
P(Chi-Square(1) >= 5.856864) = 0.015516

S** = 10.724500
P(Chi-Square(1) >= 10.724500) = 0.001057

S* = 10.923724
P(Chi-Square(1) >= 10.923724) = 0.000949

Using Bonferroni inequality, reject 0.000949 if 0.000949 < 0.05/3 = 0.008333

S** as a lower bound to S*, S, and S will be asymptotically equivalent statistics.
S* is computationally simpler than S,
however, only $S$ will be an asymptotically
valid statistic under hypothesis.

Bonferroni critical value
(assumes alpha = 0.05, 6 pairwise comparisons)
$\chi^2 = $ square for 10f, $(1 = 0.05/6) 100(TH)%$

$p(\chi^2 = \text{square}(1) <= 0.981694) = 0.991667$

*******************************************

*******************************************

---- 8 vs 9 ----

$N_T$: the total sample size = 50
$N_T^1$: the cubic of total sample size = 125000

---- vector score statistic ----

$W(1) = 247$,
$W(2) = 247$,
the sum of $W(1)$ equal 0.

---- sigma matrix ----

0.0447 0.0447
0.0447 0.0447

---- gram-schmidt orthogonalization vectors ----

1
4.73090
0.00000

---- testing hypothesis ----

$S* = 5.856864$
$P(\chi^2(1) \geq 5.856864) = 0.015516$

$s = 10.724500$

$P(\chi^2(1) \geq 10.724500) = 0.001057$

$s = 10.923724$

$P(\chi^2(1) \geq 10.923724) = 0.000949$

Using Bonferroni inequality:

Reject $0.000949$ if $0.000949 < 0.05/6 = 0.008333$

$s^*$ as a lower bound to $s^*$, $s$ and $s^*$ will be asymptotically equivalent statistics.

$s^*$ is computationally simpler than $s$.

However, only $s$ will be an asymptotically valid statistic under hypothesis.

Bonferroni critical value

(Assumes $\alpha = 0.05$, 6 pairwise comparisons)

$\chi^2$ square for 1 df, $(1 - 0.05/6)100(\%)$

$P(\chi^2(1) \leq 6.981694) = 0.991667$

*******************************************************************************