ANALYSIS OF COVARIANCE BY MATCHING FOR THE K-SAMPLE PROBLEM

by

Imogene McCanless

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1423

January 1983
ANALYSIS OF COVARIANCE BY MATCHING

FOR THE K-SAMPLE PROBLEM

by

Imogene McCanless

A Dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics in the School of Public Health.

Chapel Hill

1982

Approved by:

Dana Quade
Adviser

Joe Son
Reader

Mary P. Koch
Reader
ABSTRACT

IMOGENE MCCANLESS. Analysis of Covariance by Matching for the K-Sample Problem (under the direction of DANA QUADE).

Two methods of analysis of covariance by matching are presented for comparing K populations with respect to a continuous response variable while controlling for a discrete (possibly multivariate) covariable.

The first method presented, called the AMP sign test for K-samples, is an extension of Schoenfelder's All Matched Pairs (AMP) sign test and the Kruskal-Wallis test, basing the test on a combination of sets of two-sample statistics. It is found to have Pitman asymptotic relative efficiency (ARE) of unity with respect to the Kruskal-Wallis test, when the covariable is independent of the response variable.

The second method advanced, called the All Matched K-tuple (AMK) sign test, extends Schoenfelder's AMP sign test and Bhapkar's V-test, and analyzes matched K-tuples. This is also found to have ARE of unity, with respect to Bhapkar's V-test, when the covariable is independent of the response variable.

The two new methods are also compared with each other with respect to ARE. The results suggest that if the conditional distribution of the response variable given the covariable is bounded below, then the AMK test tends to perform better than the AMP test; otherwise the AMP test is preferred. Efficiency comparisons are reported for certain conditional distributions.
ACKNOWLEDGEMENTS

First, I would like to express my deep gratitude to my adviser, Dr. Dana Quade, for suggesting this topic and guiding its development. His continual patience with me, gentle persuasion, and abundance of insights were invaluable to this work.

Next, I would like to thank the other members of my doctoral committee - Drs. P.K. Sen, Gary G. Koch, Ronald W. Helms, Gerardo Heiss, and John R. Schoenfelder. Their moral support and helpful suggestions while reviewing this manuscript are greatly appreciated.

A special thank you goes to Dr. Schoenfelder for his continued interest in the extensions of his work, and the attention he gave to this manuscript.

Thanks, too, go to my family and friends for their continued support and encouragement and for making my years in Chapel Hill most enjoyable.

Finally, many thanks, and a big sigh of relief, to Ernestine Bland for her very timely and proficient typing of this manuscript.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>A. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>B. Review of the Literature</td>
<td>5</td>
</tr>
<tr>
<td>1. Analysis of Covariance</td>
<td>7</td>
</tr>
<tr>
<td>2. Matching</td>
<td>18</td>
</tr>
<tr>
<td>C. Nonparametric Methods Ignoring Covariabales</td>
<td>23</td>
</tr>
<tr>
<td>D. The AMP Sign Test</td>
<td>32</td>
</tr>
<tr>
<td>II. THE EXTENSION BASED ON THE KRUSKAL-WALLIS TEST</td>
<td>36</td>
</tr>
<tr>
<td>A. The Basic Framework and Notation</td>
<td>36</td>
</tr>
<tr>
<td>B. Parameter Definitions</td>
<td>38</td>
</tr>
<tr>
<td>C. Number of Analysis Units</td>
<td>39</td>
</tr>
<tr>
<td>D. The AMP Sign Statistic for K-Samples</td>
<td>42</td>
</tr>
<tr>
<td>E. The Limiting Distribution of D* under Translation-Type Alternatives</td>
<td>58</td>
</tr>
<tr>
<td>F. The AMP Sign Statistic for K-Samples, A Special Case: n_1=n_2=...=n_k</td>
<td>63</td>
</tr>
<tr>
<td>G. Chapter Summary and Concluding Remarks</td>
<td>70</td>
</tr>
<tr>
<td>III. THE EXTENSION BASED ON BHAPKAR'S TEST</td>
<td>73</td>
</tr>
<tr>
<td>A. The Basic Framework and Notation</td>
<td>73</td>
</tr>
<tr>
<td>B. Parameter Definitions</td>
<td>73</td>
</tr>
<tr>
<td>C. Number of Analysis Units</td>
<td>75</td>
</tr>
<tr>
<td>D. A Class of Test Statistics</td>
<td>79</td>
</tr>
<tr>
<td>E. The AMK Sign Test Statistic</td>
<td>81</td>
</tr>
<tr>
<td>F. The Limiting Distribution of G* Under Translation-Type Alternatives</td>
<td>89</td>
</tr>
<tr>
<td>G. The AMK Sign Statistic, A Special Case: n_1=n_2=...=n_k</td>
<td>94</td>
</tr>
<tr>
<td>H. Chapter Summary and Concluding Remarks</td>
<td>100</td>
</tr>
<tr>
<td>IV. EFFICIENCY CONSIDERATIONS AND EXAMPLES</td>
<td>103</td>
</tr>
<tr>
<td>A. Efficiency Considerations Comparing the AMP Sign Statistic with the Kruskal-Wallis Test for the Special Case n_1=n_2=...=n_k</td>
<td>103</td>
</tr>
<tr>
<td>B. Efficiency Considerations Comparing the AMK Sign Statistic with Bhapkar's V-Test for the Special Case n_1=n_2=...=n_k</td>
<td>107</td>
</tr>
<tr>
<td>C. Comments on &quot;Handpicked&quot; Examples, and an Example with &quot;Real&quot; Data.</td>
<td>111</td>
</tr>
</tbody>
</table>
D. Efficiency Comparisons Between AMK Sign Statistic and AMP Sign Statistic for K-Samples for the Special Case $n_1 = n_2 = \ldots = n_K$ .................................................. 114

V. COMMENTS AND SUGGESTIONS FOR FUTURE RESEARCH .............. 120
   A. Comments on the AMP Sign Test for K-Samples ................. 120
   B. Comments on the AMK Sign Test ............................ 122
   C. General Comments and Suggestions for Future Research ........ 124

REFERENCES .............................................................................. 126
CHAPTER I

INTRODUCTION

A. Introduction

The issue of comparing two or more populations with respect to some response variable is of concern in many disciplines. The literature is rich with presentations and discussions of sampling techniques and methods of analyzing the data after selection for the multisample problem. When the response variable is affected only by variables under the direct control of the investigator, the analysis proceeds in a straightforward manner. If, however, there are variables that are not of primary interest but nevertheless affect the response variable, the analysis is more complicated. This dissertation proposes two new techniques for making comparisons among groups in the presence of extraneous variables.

As a simple example of the problem being addressed, consider the case where an investigator is interested in assessing the effectiveness of a certain treatment designed to lower blood pressure levels in humans. He selects two samples and subjects the individuals in one sample to the active treatment while giving those in the other sample a placebo (i.e., there are a treatment group and a control group). He wishes to compare the treatment group to the control
group with respect to blood pressure measurements in order to determine whether the treatment has the effect of lowering the blood pressure. If the data that are collected provide information only on the blood pressure level of the subjects after this experiment is completed, the usual one-way analysis of variance (ANOVA) techniques are the standard analysis procedures that are used to analyze these data. If the hypothesis of no difference between groups is rejected, the natural conclusion drawn is that the treatment effect is significant. However, if the treatment group had lower blood pressure levels than the control group before the experiment began, a detected difference could be due to the disparity of blood pressure levels prior to the experiment rather than due to the treatment. It is well known that the best predictor of future blood pressure is current blood pressure. Hence, the analysis would be "cleaner" if it would in some way account for initial blood pressure levels.

In general, when an investigator wishes to compare two or more population groups with respect to a response variable, there may be extraneous variables that affect the response variable in some, perhaps predictable, way. These variables should be considered, either at the design stage or at the analysis stage of the experiment, in order to minimize bias. If the relationship between the extraneous variables and the response variable is known, this information can be used to improve the precision of estimates, in addition to reducing bias.
Notation (and general framework)

The problem may be described in the following way: Define $Y_{ij}$ to be the response measurement of the $j$th subject from the $i$th sample ($i=1,\ldots,K; j=1,\ldots,n_i$), where there are $K$ population groups under consideration, and $n_i$ subjects are selected from the $i$th population. Let $X$ be a variable (possibly multivariate) known or suspected to influence $Y$. Then $X_{ij}$ is the measurement of this extraneous variable on the $j$th subject in the $i$th sample. Let $X$ be called a covariable. If $X$ has the same marginal distribution in each of the populations, then $X$ is said to be concomitant. The issue at hand is the following: It is desired to make comparisons among the $K$ population groups with respect to $Y$ based on the analysis of the selected samples while incorporating the information on the covariable $X$.

Reasons for Considering $X$

Returning to the blood pressure treatment example, suppose that the only data collected are the final blood pressure measurements ($Y_{ij}$). The standard one-way ANOVA would produce an estimate of the between-treatment group difference. However, if the initial blood pressure levels are different, this initial difference could cause the estimate to be biased. If the treatment group has lower initial blood pressure levels than the control group, then an observed post-treatment difference could be due to the pre-treatment difference, rather than the treatment; hence, an erroneous conclusion could be drawn. Or, if the treatment group has higher blood pressure levels than the control group, then a clinically significant treatment effect may not be detected due
to the initial disparity. It follows that ignoring the initial blood pressure measurements (or age, or body mass, or other well known predictors) could severely bias the estimated difference between the two groups.

In addition to introducing bias into estimates of population differences, the precision of the estimates will be affected if important covariables are ignored. Considering the above experiment where no covariable measurements are collected, ANOVA techniques may be used to analyze the data. In this analysis, the total variability of the data is partitioned into components. The interesting components are the variability attributed to treatment effects and the "within group" variability (i.e., variability due to error). Inevitably, if there are varying initial blood pressure levels (and differing age groups, body size groups, etc.) this will increase the total variability, which will increase the error variance. If covariables are accounted for (either in the design of the experiment or at the analysis stage) the error variance will be reduced, resulting in greater precision in estimating the treatment effects.

In general, if an experiment has a response variable that is influenced by some extraneous variable (covariable), ignoring the covariable introduces bias and inflates the estimate for error variance. Clearly, a "complete" analysis of the experiment will incorporate the covariable X. The method of including X must reduce the bias and increase the precision of the comparisons.
B. Review of the Literature

Methods of Accounting for $X$

Basically, two methods exist in the literature for incorporating a covariable into data analysis: structural or adjustment methods, and balancing or matching methods.

The classical (parametric) analysis of covariance (ANOCOVA) is the most familiar adjustment method. This technique combines the features of linear regression analysis and ANOVA. The usual parametric ANOVA procedures for testing differences in means (for $K$ population groups) are based on partitioning the total sum of squares into components; if the mean square for the means (treatment) is significantly large relative to the mean squared error, the hypothesis of equal means is rejected. The ANOCOVA procedure for testing differences in means is also based on partitioning the total sum of squares into components. But in this latter case, the test is for a difference in means of residuals, where the residuals are the differences between the actual observations and a quantity predicted by regression on $X$. Hence, the values of the response variable are adjusted for values of concomitant variables.

In general, adjustment methods of analysis postulate (or find) some relationship between $X$ and $Y$, and adjust the $Y_{ij}$ accordingly, yielding "residuals" on which usual ANOVA procedures are performed. ANOCOVA and other adjustment methods will be described in greater detail later.

Matching is a method of incorporating covariables in the analysis by equalizing the distribution of $X$ in each treatment group. There
are various methods of matching. For the two-sample case, pair-
matching is the most common. A matched pair is a pair of observa-
tions, one from each sample, which agree on X. If X is discrete,
the observations are said to agree on X if they have the same value
of X. If X is continuous, observations agree on X if their X values
are "nearly" the same. "Nearly the same" could mean that the X
values are within some tolerance, or within the same subgroup of
the continuum of X, or are "closest neighbors," etc. These and other
matching schemes will be discussed in more detail later. In general,
matching methods ensure group comparability with respect to the co-
variable. These methods are directed primarily at reducing (or
eliminating) the bias, but may also have the additional effect of
increasing the precision of the estimates.

Randomization is a method of design that eliminates bias due to
X. In a completely randomized design, the researcher randomly assigns
the experimental units to treatment groups. This has the consequence
of assigning the X effect to error. This is the technique that Wold
(1956) describes as the "controlled experiment," and is frequently
used in practice. The use of randomization protects against bias
arising from unmeasured or unanalyzed extraneous variables. It is
not, however, always feasible. In fact, it is sometimes unethical
to use randomization, especially in studies on human population groups.
In particular, as Cochran (1969) discussed, ethical considerations
often prohibit randomization in observational studies. In these
studies, the objective is to investigate possible cause-effect rela-
tionships. The analysis includes comparisons of groups subjected to
different treatments which are pre-assigned in a non-random manner. In these and similar studies, effects of extraneous variables are typically considered via adjustment procedures or matching.

**Analysis of Covariance**

In this section, adjustment methods of including $X$ in the analysis are presented. The classical parametric ANOCOVA is the most popular method for incorporating a covariable in analysis, and hence is presented first.

As Fisher (1934) has said, the usual parametric ANOCOVA "combines the advantages and reconciles the requirements of the two very widely applicable procedures known as regression and analysis of variance."

Consider the data $(X_{ij}, Y_{ij})$ $i=1,\ldots,K$ and $j=1,\ldots,n_i$. As before, $Y_{ij}$ are response measurements and $X_{ij}$ are measurements on a covariable for which adjustment is made. If $X$ is univariate and has the same mean in each sample, the one-way classification model considered may be expressed as:

$$Y_{ij} = \mu_i + \beta(X_{ij} - \bar{X}..) + \epsilon_{ij}$$

where

- $Y_{ij}$ = the response measurement on the $j$th observation in the $i$th sample,
- $X_{ij}$ = the measurement of the covariable on the $j$th observation in the $i$th sample,
- \( \bar{X}.. \) = the overall mean of the covariable,
- $\mu_i$ = the $i$th population mean,
- $\epsilon_{ij}$ = the error.
The \( \varepsilon_{ij} \) are assumed to be independent, identically distributed \( N(0, \sigma^2) \).

Under this model, assumptions are:

1) \( Y \) has a linear regression on \( X \).

2) The regression of \( Y \) on \( X \) has the same slope in all populations.

3) The covariable \( X \) is not affected by the "treatments."

In addition to these assumptions, either randomization or concomitance is required, or some assumption which permits extrapolation of the model beyond the available data.

If \( X \) is multivariate, extensions are clear. For example, if the data require adjustment for two covariables, the two-way classification model may be expressed

\[
Y_{ij} = \mu_i + \beta_1(X_{ij1} - \bar{X}_{..1}) + \beta_2(X_{ij2} - \bar{X}_{..2}) + \varepsilon_{ij}
\]

where

\( X_{ijm} \) = the measurement of covariable \( m \) (\( m=1,2 \)) on the \( j \)th observation in sample \( i \),

\( \bar{X}_{..m} \) = the corresponding sample mean for covariable \( m \) (\( m=1,2 \)),

and \( Y_{ij}, \mu_i \), and \( \varepsilon_{ij} \) are defined as before.

Assuming the one-way classification model, the normal equations imply the following (least squares) estimators for \( \beta \) and \( \mu_i \):

\[
\hat{\beta} = \frac{\sum_{i=1}^{K} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{..})(Y_{ij} - \bar{Y}_{..})}{\sum_{i=1}^{K} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{..})^2 - \sum_{i=1}^{K} n_i (X_{i.} - \bar{X}_{..})^2}
\]

\[
\hat{\mu}_i = \bar{Y}_{i.} - \hat{\beta}(\bar{X}_{i.} - \bar{X}_{..})
\]
where
\[
\bar{Y}_{..} = \frac{1}{n} \sum_{i=1}^{K} n_i \bar{Y}_{i.}
\]
\[
\bar{Y}_{..} = \frac{1}{n} \sum_{i=1}^{K} \sum_{j=1}^{n_i} Y_{ij}, \quad n = \sum_{i=1}^{K} n_i
\]
\[
\bar{x}_{.i} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}
\]
\[
\bar{x}_{..} = \frac{1}{n} \sum_{i=1}^{K} \sum_{j=1}^{n_i} x_{ij}
\]

Under this model the adjusted mean scores \( \hat{\beta}(x_{.i} - \bar{x}_{..}) \) are defined to be the predicted values obtained from the model by allowing \( X \) to assume its mean value for each of the \( K \) samples. That is, \( \hat{\beta}(x_{.i} - \bar{x}_{..}) \) is "attached" to \( \bar{Y}_{.i} \), the usual ANOVA estimator for the ith population mean. To test the hypothesis of no difference among the \( K \) populations (of equivalently the hypothesis of no treatment effects) the usual ANOVA is performed on these regression-adjusted estimates; that is, the residuals resulting from the regression of \( Y \) on \( X \). This analysis proposes to eliminate the effect due to \( X \) by adjusting the response measurements for it. In ANOVA, the treatment mean \( \bar{Y}_{.i} \) is used to estimate the mean response with the ith treatment. With ANOCOVA, an adjustment for the covariable (based on the regression) is made to make the \( \bar{Y}_{.i} \)'s comparable with respect to the concomitant variable. In the simplest case, the adjustment takes the form
\[
\bar{Y}_{.i} \text{(adj)} = \bar{Y}_{i.} - \hat{\beta}(x_{.i} - \bar{x}_{..})
\]
(i.e., \( X \) has the same mean in all samples). These adjusted treatment
means are simply estimates of the intercepts of the treatment regression lines. Comparisons among the populations may be made adjusted for X by making comparisons among the \( \bar{Y}_{i*.} \text{(adj)} \).

A more general ANOCOVA model may take the form

\[
Y_{ij} = \mu + \beta_1 X_{ij} + \beta_2 D_{ij} + \epsilon_{ij}
\]

where \( \mu \) is the overall mean, \( D_{ij} \) is a dummy variable distinguishing the samples, and the \( Y_{ij}, \beta_i, \) and \( \epsilon_{ij} \) are as previously defined. The assumptions under this model are the same as before, however, the co-variable is used more efficiently than under the previous model. Testing the significance of the \( \beta_2 \) in this model is equivalent to testing that there is no difference among the K populations with respect to \( Y \), having taken account of the relationship between \( X \) and \( Y \).

For this problem, the least squares estimator of the regression coefficient is derived using all the data. Some researchers suggest that only the controls should be used to derive a proper estimate for \( \beta_1 \) (Belson, 1956). Cochran (1969) provides conditions under which this is appropriate.

The relationship between \( X \) and \( Y \) may, of course, be something other than linear. However, as long as the form of the relationship is known (or postulated and tested), then the responses may be "adjusted" for \( X \), and the ANOVA performed on the residuals.

In addition to the standard parametric ANOCOVA procedures presented, many other adjustment techniques are possible, but few seem to have been developed in the literature. Following are some non-parametric adjustment methods which have been advanced.
Nonparametric ANOCOVA

Quade (1967) proposed a rank ANOCOVA. In his presentation he invokes a general principle which is applicable to the usual (parametric) technique as well:

If the hypothesis is true, then the populations are all identical and the samples can be pooled. So use the pooled sample to determine a relationship through which $Y$ can be predicted from $X$. Then compare each observed response $Y_{ij}$ with the value which would be predicted for it from the corresponding $X_{ij}$, and assign it a score $Z_{ij}$, positive if $Y_{ij}$ is greater than predicted and negative if smaller. Finally, compare the populations by performing an ordinary one-way analysis of variance of the scores.

The procedure proposed by Quade is applicable when a random sample of observations has been selected from each of the $K$ populations $(Y_{ij}, X_{ij})$ where $Y_{ij}$ is the univariate response and $X_{ij}$ is the (possibly multivariate) concomitant variable measurement ($i=1,\ldots,K$ and $j=1,\ldots,n_i$). The procedure is designed to test the hypothesis that the conditional distribution of $Y|X$ is the same for each population, against the alternative that at least one population is stochastically larger for all fixed values of $X$. The procedure is applied as follows:

1. Let $R_{ij}$ be the rank (from the smallest to largest) of $Y_{ij}$ among all $N = \Sigma n_i$ observed values of $Y$, so that $R_{ij}$ is the rank corrected for the mean. (Use average ranks in case of ties.)

2. Let $C_{ij}^{(u)}$ be the corrected rank of $X_{ij}^{(u)}$ among the observed values of $X^u$ (for $u=1,\ldots,p$).
3. Regress R on \( C(1), \ldots, C(p) \) using ordinary multiple linear regression, to form an equation whereby ranks of \( Y \) can be predicted from the ranks of \( X \).

4. Compute the predicted ranks \( \hat{R}_{ij} \) from the regression equation.

5. Define scores \( Z_{ij} \) to be the residuals from the regression of ranks. So \( Z_{ij} = R_{ij} - \hat{R}_{ij} \).

6. Test the hypothesis of identical conditional distributions of \( Y|X \) by comparing the variance ratio (VR) with the critical value for an \( F \) with \((K-1, n-K)\) degrees of freedom (\( n = \sum n_i \)) where

\[
VR = \frac{(n-K) \sum_{i=1}^{K} \left( \frac{n_i}{\sum_{j=1}^{n} Z_{ij}} \right)^2}{(K-1) \left( \sum_{i=1}^{K} \sum_{j=1}^{n_i} Z_{ij}^2 - \sum_{i=1}^{K} \left( \frac{n_i}{\sum_{j=1}^{n} Z_{ij}} \right)^2 \right) / n_i}
\]

This proposed test relaxes the usual assumptions of normality, linearity of regression, and homoscedasticity. It does, however, require that \( X \) be concomitant, where the standard procedure does not have this requirement. Hence, the standard procedure may be applicable in some situations where this procedure is not.

Puri and Sen (1969) advanced a generalization of this procedure in their proposal of ANOCOVA based on general rank scores. The situation for which this test is applicable is the same as for the above test. The procedure is outlined as follows:
1. Define the \((p+1)\times 1\) vector \(Z = (Y, X_1, \ldots, X_p)\) where \(p\) is the number of concomitant variates of interest. Then let \(Z_{ij}^j\) represent the \((p+1)\times 1\) vector of measurements on the \(j\)th observation from the \(i\)th sample.

2. Construct the matrix
\[
G = [Z_1^1 \ldots Z_{n_1}^1, Z_1^2 \ldots Z_{n_2}^2 \ldots Z_1^K \ldots Z_{n_K}^K].
\]

3. Rank each row of \(G\) in increasing order of magnitude, yielding a \((p+1)\times n\) matrix
\[
R_n = \begin{bmatrix}
R_{01}^{(1)} & \ldots & R_{0n_1}^{(1)} & \ldots & R_{01}^{(K)} & \ldots & R_{0n_K}^{(K)} \\
R_{11}^{(1)} & \ldots & R_{1n_1}^{(1)} & \ldots & R_{11}^{(K)} & \ldots & R_{1n_K}^{(K)} \\
& \vdots & & \vdots & & & \\
R_{p1}^{(1)} & \ldots & R_{pn_1}^{(1)} & \ldots & R_{p1}^{(K)} & \ldots & R_{pn_K}^{(K)}
\end{bmatrix}
\]

Note that each row of \(R\) is a permutation of the numbers
\(1, \ldots, N\).

4. Let \(\{E_{n, \alpha}^{(1)}, \alpha = 1, \ldots, n_j\}\) be a set of real (known) constants for \(i=0,1,\ldots,p\). Then replace the \(i\)th row of \(R_n\) by \(E_{n, \alpha}^{(i)}\), yielding the \((p+1)\times n\) matrix
\[
E_n = \begin{bmatrix}
E_n^{(0)} & \ldots & E_n^{(0)} & \ldots & E_n^{(0)} & \ldots & E_n^{(0)} \\
E_n^{(0)} & \ldots & E_n^{(0)} & \ldots & E_n^{(0)} & \ldots & E_n^{(0)} \\
& \vdots & & \vdots & & & \\
E_n^{(p)} & \ldots & E_n^{(p)} & \ldots & E_n^{(p)} & \ldots & E_n^{(p)}
\end{bmatrix}
\]
5. Define:

\[ T_{n,r}^{(i)} = \frac{1}{n_i} \sum_{\alpha=1}^{n_i} E_{n,\alpha}^{r}(i) \text{ for } r=0,\ldots,p, \ i=1,\ldots,K \]

where \( r=0 \) corresponds to the response measurement and \( r=1, \ldots,p \) to the respective variates of the covariable \( X \). There are \((p+1)K\) such random variables.

6. Define also:

\[ E_n^{(r)} = \frac{1}{n} \sum_{\alpha=1}^{n} E_{n,\alpha}^{(r)}, \quad \bar{E}_n = (E_n^{(0)}, E_n^{(1)}, \ldots, E_n^{(p)}). \]

7. Compute \( L_n = T_{n,r}^{(i)} - \bar{E}_n^{(r)}, r=0,\ldots,p \) and \( i=1,\ldots,K \).

8. The statistic \( L_n \) is a function of the residuals of the regression lines of \( T_{n,0}^{(i)} \) on \( T_{n,1}^{(i)}, T_{n,2}^{(i)}, \ldots, T_{n,p}^{(i)} \) \((i=1,\ldots,K)\) and, as such, is shown to have a limiting distribution which is chi-squared with one degree of freedom. Under the null hypothesis, the distribution is central chi-squared; under the alternative hypothesis, its distribution is non-central chi-squared.

According to Puri and Sen, this procedure can be used with the "usual" restrictions placed on the scores. When the scores for Puri and Sen's test are the ranks themselves, the technique is asymptotically the same as Quade's proposed method discussed earlier. Hence, Quade's test is asymptotically a special case of Puri and Sen's. Differences in power will exist, however, as each technique uses a different theoretical distribution for its asymptotic approximation.

Another nonparametric procedure is proposed by McSweeney and
Porter (1971). This method involves applying parametric ANOVA procedures to the mean-deviated ranks in the same way that Friedman's chi-squared test is parametric ANOVA applied to ranks. Monte Carlo work performed by McSweeney and Porter generated an empirical distribution for the variance ratio thus obtained, which appeared to be closely approximated by the F-distribution.

Concluding Remarks on ANOVA

Presented in the preceding discussion are some of the nonparametric ANOVA procedures that are advanced in the literature, including a brief discussion of the standard parametric ANOVA. These procedures are used primarily in experimental situations (where randomized layouts are frequently used) to increase the precision of estimates of population differences, or equivalently, treatment effects. (Note that these nonparametric procedures assume concomitance, which is guaranteed by randomization, while the standard ANOVA procedures require either concomitance or that some assumption be made which allows extrapolation of the model beyond the available data. Nevertheless, in experimental situations, problems of bias should not be of critical concern.) In a randomized experiment, the amount of precision increase depends on the correlation between X and Y. Cochran (1957) shows that when this correlation is smaller than 0.3 in absolute value, the increase in precision of ANOVA over ANOVA is negligible. In observational studies, ANOVA is used primarily to reduce bias caused by non-random assignment of X among the treatment groups, but also serves to increase precision.
ANOCOVA is a powerful statistical technique; however, it does not necessitate requirements, as discussed previously. Elashoff (1969) compiled results of various researchers who have studied the robustness of ANOCOVA to violations of its assumptions. In her paper, each assumption is considered separately with respect to the effect on the analysis when the assumption is not met. She notes that little investigation has been made of situations where more than one assumption is unsatisfied.

Elashoff concludes that the assumptions of normality, linearity, and homoscedasticity are mostly for statistical convenience, citing studies as following:

1. Atiquallah (1964) investigated robustness against non-normality for ANOCOVA concluding that the test is "... appreciably affected by non-normality even in balanced classifications. The degree of sensitivity to non-normality is determined by the distribution of the concomitant variables." He finds that the problem is not serious when the concomitant variable is normal. Also, in many cases transformations of the data may be useful in producing more nearly normal distributions.

2. Atiquallah (1964) investigated the case where the true regression of Y on X is quadratic. He concluded that results are seriously biased unless the coefficient of the quadratic term is quite small. As Elashoff points out, however, transformations of the data may help to produce the required linearity
(see Kruskal, 1968). If the linearity assumption is not satisfied, adjustments should be made based on a more accurate model of the relationship between X and Y.

3. Potthoff (1965) addresses the issues of unequal variances for a special case of two comparison groups, and concludes that the effect of unequal variances is minimized when $n_1 = n_2$ and the variance of X is the same in each group. If the variance of the errors depends on X in a uniform way across treatment groups, a transformation of the data may produce the necessary homoscedasticity.

Although Elashoff states that normality, linearity, and homoscedasticity are for statistical simplicity, she explains that other assumptions of standard ANOVA are crucial to its use:

1. Many researchers identify randomization as a requirement for standard ANOVA (Winer, 1962; Evans and Anastasio, 1968; Lord, 1963). Major issues cited include concern over not removing all the bias and concern over extrapolation. Concomitance is, however, sufficient for ANOVA. If neither randomization nor concomitance is satisfied, it is crucial that some other assumption be satisfied that allows for extrapolation.

2. A basic postulate underlying the use of ANOVA is that the covariable is unaffected by the treatment. If this is unsatisfied, the adjustment may remove part of the treatment
effect or produce a spurious treatment effect.

3. The assumption that the regression of Y on X has the same slope in all populations (i.e., there is no treatment-slope interaction) is also crucial to ANOCOVA use. If the assumption is not satisfied, the treatment which is best on the average may not be the one best at all levels of X.

It is, as discussed above, agreed by many that the random assignment of subjects to treatment groups is desirable to ANOCOVA use. ANOCOVA is, however, a widely used method of analysis in studies where the randomization is impossible or unethical. In such cases, the techniques are advantageous, but the results must be interpreted with prudence. Elashoff suggests that a more correct approach might be to perform the analysis on data for which the values of the covariable are fixed, rather than using adjustment techniques. The method of including a covariable in analysis without using adjustment methods is known as matching.

Matching Schemes

In the literature the term "matching" most frequently refers to pair-matching. In a pair-matching scheme each member of the treatment group is paired with one and only one member of the control group subject to the restriction that the members "agree" on X. Notationally, a pair of observations, one from each sample \((X_{1r}, Y_{1r})\) and \((X_{2s}, Y_{2s})\) are matched in this sense if \(X_{1r}\) and \(X_{2s}\) agree. If X is categorical, the pair of observations is matched if \(X_{1r} = X_{2s}\), i.e.,
if the covariable assumes exactly the same value in each observation. If \( X \) is continuous, there are several methods of defining a matched pair. Three schemes are presented:

1. The continuum of \( X \) is partitioned into well-defined subintervals. Then \((X_{1r}, Y_{1r})\) and \((X_{2s}, Y_{2s})\) are matched if \( X_{1r} \) and \( X_{2s} \) fall into the same subinterval. This method actually converts \( X \) from a continuous variable to a discrete one, and is then equivalent to the above categorical case.

2. A tolerance \( c > 0 \) is specified. Then \((X_{1r}, Y_{1r})\) and \((X_{2s}, Y_{2s})\) are matched if \(|X_{1r} - X_{2s}| \leq c\). This scheme is called caliper matching.

3. The observations in the treatment group are ordered in any well-defined manner. Then \((X_{11}, Y_{11})\) is paired with the control \((X_{2s}, Y_{2s})\) which minimizes \(|X_{11} - X_{2t}|\). Next \((X_{12}, Y_{12})\) is paired with the next available control \((X_{2t}, Y_{2t})\) which minimizes the absolute difference. The scheme is continued until all pair matches are defined. This scheme is called "nearest neighbor" matching.

There are other schemes presented in the literature. These schemes for pair-matching with respect to a continuous covariable are highly flexible, with refined definitions of matching left to the investigator (e.g., lengths of subintervals, tolerance definition, ordering specifications, etc.). The ramifications of various decisions are discussed in the literature.
Billewicz (1965) investigated methods of selecting subintervals for continuous \( X \) when using the first method discussed above. Cochran and Rubin (1973) investigated bias reduction relative to various ordering schemes for the third method of pair matching on continuous \( X \) discussed above. Other consequences of matching definition decisions are discussed in Nathan (1963), Billewicz (1963,1964), Raynor (1977), Yinger et al (1967), and many others. For a more detailed discussion of the literature on methods of matching, including advantages, disadvantages, criticisms, and appropriateness of matching, see Schoenfelder (1981). It must be pointed out, however, that matched samples are not random samples, and cannot be treated as such in the analysis stage.

As mentioned earlier, taking account of a covariable can occur at either the design stage or the analysis stage. To compare "random" analysis and matched sampling may be misleading. Schoenfelder addressed this issue, explaining that a fairer comparison is to contrast random sampling and its resulting analysis with matched sampling and its resulting analysis. These comparisons could provide a solid theoretical basis for the widespread use of matching.

Schoenfelder proposed that a form of "matched analysis" which could be performed on random samples could be devised. This technique could then reasonably be compared with the usual "random analysis" on random samples.

"Matched" Analysis

Frequently an investigator is confronted with two random
samples (of sizes \( n_1 \) and \( n_2 \)) from population groups (or treatment groups) and requested to analyze the data to make comparisons between the groups. Matching in the usual sense is not a viable option, since the data are already collected. In order to employ the usual technique of matching on \( X \), he must combine the two samples into one matched sample of size \( n \leq \min(n_1, n_2) \) independent matched pairs (IMP). Consequently, the \( n_1 + n_2 - 2n \) unmatched observations are then discarded. The obvious loss of information using this technique makes use of matching in this way very unattractive. Quade (1967) suggested an analysis based on all matched pairs (AMP) as opposed to using only IMP. To clarify the notion, consider \( n = 3 \) IMPs that have the following \( X \) values:

\[
X_{11} = X_{21} = 3 \\
X_{12} = X_{22} = 4 \\
X_{13} = X_{23} = 4
\]

Using IMPs only, \( Y_{11} \) is compared in \( Y_{21} \), since their associated \( X \) values match; \( Y_{12} \) is compared with \( Y_{22} \) and \( Y_{13} \) is compared with \( Y_{23} \). However, the associated \( X \) values of \( Y_{12} \) and \( Y_{23} \) match, as do those for \( Y_{13} \) and \( Y_{22} \). The usual IMP analysis ignores this, but the AMP incorporates these "cross matches" in the analysis.

Schoenfelder (1981) examines the efficiency of performing an AMP analysis when matched sampling has been performed relative to the usual IMP analysis that is standard for matched samples. In the cases studied, he finds that if the comparison is based on the optimally weighted AMP sign statistic, that statistic, in terms of
asymptotic relative efficiency, is always half again more efficient than the sign test based on IMPs. He also examines the relative efficiency of basing comparisons on the proposed matched ANCOVA (AMP analysis) using random samples relative to the usual ANCOVA test based on random samples. He finds this asymptotic relative efficiency to be 0.955. This result reveals that, when comparing two populations in the presence of a discrete covariable, if there are any doubts about the parametric ANCOVA assumptions, then Schoenfelder's matched ANCOVA based on the optimal AMP sign statistic should be used. In doing this, only a slight loss in efficiency results if the assumptions are indeed satisfied.

Concluding Remarks on Matching

The appropriateness, advantages, and disadvantages of matching are the subjects of discussion by many researchers. Matching is a desirable method of incorporating a covariable $X$ in the analysis due, primarily, to its simplicity and intuitive appeal, and to the relaxed condition that it is not necessary to know precisely how $X$ and $Y$ are related. In the ANCOVA, the relationship between $X$ and $Y$ must be postulated, or found, but this is not necessary for matching. However, when using matched sampling there is the added expenditure of resources, and increased risk of lost information due to mis-matches, attrition, inability to find matches, etc. If it is known that $X$ is related to $Y$ such that $X$ should be considered in the comparisons to be made, but the precise relationship between $X$ and $Y$ is not known, matched sampling provides the obvious alternative to using, and
possibly misusing ANOCOVA. If random samples have been selected, using AMP analysis results in a negligible loss in efficiency, but has the advantages of relaxed assumptions over ANOCOVA. The AMP analysis procedure for the two-sample problem where \( X \) is discrete is developed by Schoenfelder (1981). The generalization of this test (say to \( K \) samples) can serve as a powerful statistical tool with many applications.

C. Nonparametric Methods Ignoring Covariables

In this section, the test procedures forming the basis of the proposed tests are reviewed. The Mann-Whitney test is a nonparametric test for comparing two populations with respect to a response variable aimed at detecting shift differences. Schoenfelder's AMP sign test is based on the Mann-Whitney test, but incorporates a covariable in the analysis. The Kruskal-Wallis test is an extension of the Mann-Whitney test to a general \( K \)-sample problem, which derives the comparison by combining all possible two-sample statistics. Bhapkar's \( V \)-test extends the Mann-Whitney test to a general \( K \)-sample problem by analyzing \( K \)-tuples. The test procedures that are proposed in this dissertation are extensions of Schoenfelder's AMP sign test to the \( K \)-sample problem by generalizing the Kruskal-Wallis test (Chapter II) and Bhapkar's \( V \)-test (Chapter III).

The Mann-Whitney Statistic

The usual nonparametric test for comparing two samples of size \( n_1 \) and \( n_2 \) is based on the Mann-Whitney (Wilcoxon) statistic, which is
given by

\[ U_{n_1n_2} = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I(Y_{2j} > Y_{1i}) \]

where

\[ I(a > b) = \begin{cases} 
1, & a > b \\
0, & \text{else}
\end{cases} \]

The null hypothesis of interest is that the populations do not differ in distribution; the alternative is that population 2 is stochastically larger.

**Theorem 1:** (Lehmann, 1975). Let \( U_{n_1n_2} \) be the Mann-Whitney statistic as defined above. Then

a. \( E(U) = n_1n_2 \tau_{12} \)

b. \( \text{Var}(U) = n_1n_2\tau_{12} + n_1n_2(n_1-1)\tau_{112} + n_1n_2(n_2-1)\tau_{122} - n_1n_2(n_1+n_2-1)\tau_{12}^2 \)

where

\( \tau_{12} \) = the probability that of two observations, one from each sample, the one from the first is smaller.

\( \tau_{112} \) = the probability that of three observations, two from the first sample and one from the second, the one from the second is largest.

\( \tau_{122} \) = the probability that of three observations, two from the second and one from the first sample, the one from the first is smallest.

Note that under the null hypothesis, \( \tau_{12} = 1/2, \tau_{112} = \tau_{122} = 1/3 \); under the expressed alternative \( \tau_{12} > 1/2 \). Hence, the null hypothesis is rejected if the test statistic is too large. Mann and Whitney (1947)
showed that under the null hypothesis the statistic is asymptotically normal with mean \( n_1 n_2 / 2 \) and variance \( n_1 n_2 (n_1 + n_2 + 1)/12 \).

The Mann-Whitney test assumes:

1. Both samples are random samples from their respective populations.
2. There is mutual independence between the two samples.
3. The response measurements are continuous.

This test is used when two population groups are compared in the absence of, or ignoring, covariables.

The Kruskal-Wallis Test

Perhaps the most widely used nonparametric technique for addressing the K-sample location problem is the Kruskal-Wallis test. The appropriate framework is that of the usual one-way layout, where the hypothesis of interest asserts that the K populations are identical, and this is tested against the alternative that location differences exist. The strict validity of the Kruskal-Wallis test does not depend on the location model, but the power of the test is primarily aimed at detecting location differences. The only assumptions required for the test are: 1) the data are K independent, random samples from the K populations, of sizes \( n_i \), \( i = 1, 2, \ldots, K \), and 2) the response variable is univariate and continuous. The continuity assumption is imposed for statistical convenience, so that ties may be ignored.

The Test Statistic

Let the data be expressed as \( Y_{ij} \), \( i = 1, 2, \ldots, K \); \( j = 1, 2, \ldots, n_i \),
\[ n = \sum_{i}^{n_{i}}. \] Let \( R_{ij} \) be the rank of \( Y_{ij} \) among the \( n \) observations, let \[ R_{i} = \sum_{j=1}^{n_{i}} R_{ij} \] be the rank sum for sample \( i \), and let \( R_{i.} = \frac{1}{n_{i}} R_{i} \) be the average of the ranks associated with the \( i \)th sample observations. It can easily be demonstrated that \( E[R_{i.}|H_{0}] = \frac{n_{i}+1}{2} \). The Kruskal-Wallis test statistic is expressible as

\[ H = \frac{12}{n(n+1)} \sum_{i=1}^{K} n_{i}(R_{i.} - \frac{n+1}{2})^2. \]

Note that under \( H_{0} \), \( (R_{i.} - \frac{n+1}{2})^2 \) tends to be small. When shift differences exist, the \( R_{i.} \)'s tend to be more disparate, and thus (at least some of) the \( n_{i}(R_{i.} - \frac{n+1}{2})^2 \) values tend to be larger, yielding large values of \( H \).

Kruskal (1952) shows that \( H \) is asymptotically distributed as a chi-squared random variable with \( K-1 \) degrees of freedom. Some exact tables exist for small \( K \) and \( n_{i} \) (Kruskal and Wallis, 1952; Kraft and Van Eeden, 1968; Iman, Quade, and Alexander, 1975). The adequacy of the chi-squared approximation for small samples is investigated by Gabriel and Lachenbruch (1969). The asymptotic relative efficiency (ARE) with respect to the normal theory one-way layout is \( 3/\pi \), or approximately 0.955 (see Andrews, 1954). Kruskal (1952) and Kruskal and Wallis (1952) discuss the consistency; the test based on large values of \( H \) is consistent if and only if there is at least one population for which the limiting probability is not one-half (i.e., \( \tau_{it} \neq 1/2 \) for some \( i \neq t \)) that a random observation from this population is greater than an independent randomly selected member from the other sample observations. When \( K=2 \), the Kruskal-Wallis test is easily seen to be equivalent to the Wilcoxon-Mann-Whitney test. Let
\[ U_{it} = \sum_{j=1}^{n_i} \sum_{j=1}^{n_t} \text{sgn}(Y_{ij} - Y_{tj}) \]

be the "centered" Mann-Whitney statistic (with zero expectation under the null hypothesis) for comparing samples i and t. Then the Kruskal-Wallis statistic may be expressed as

\[ H = \frac{3}{n(n+1)} \sum_{i=1}^{k} \frac{1}{n_i} [\sum_{t} U_{it}]^2 \]

From this form of the statistic it is clear that the Kruskal-Wallis test makes comparisons among the K populations by combining all possible two-sample (Mann-Whitney) statistics and weighting these according to sample size.

**Bhapkar's Test**

Bhapkar (1961) proposed a nonparametric test, called the V-test, for the problem of several samples. The test is appropriate under these general conditions:

1. The data \( Y_{ij}, \ j = 1,2,\ldots,n_i \) are independent (real valued) observations from the ith population with c.d.f. \( F_i \), \( i = 1,2,\ldots,K \).
2. The K samples are independent.
3. The \( F_i \) are assumed to be continuous.
4. The general homogeneity hypothesis \( H_0: F_i = F \) for all \( i \) is tested against location alternatives. In particular, if the investigator assumes that the populations differ only by translation if at all, then the test is for equality of location parameters.

Variations of the V-test appear in the literature. Bhapkar (1964) presented a W statistic; Deshpande (1965a) proposed an L statistic;
Deshpande (1970) presented a general class of test statistics for the K-sample problem, of which the V-test, W-test, and L-test are members. A brief description of each of the named test procedures follows.

For each of these tests, the following procedure is required:

1. Construct K-tuples of observations by selecting one observation from each sample. Obviously the total number of K-tuples that can be thus formed is $\prod_{i=1}^{k} n_i$.

2. Rank the K observations within each K-tuple.

3. Define $v_{ij}$ to be the number of K-tuples in which the observation from sample i has rank j (and is hence larger than exactly j-1 observations and smaller than exactly k-j observations in that K-tuple). Note that the continuity assumption for the $F_i$ permits ties to be ignored.

4. Define $U_{ij} = v_{ij} / \prod_{i} n_i$. Observe that $U_{ij}$ will assume values in the closed interval $[0,1]$. Note that if the observations belonging to sample i are small in comparison with those belonging to the other samples, then $U_{ij}$ will tend to assume large values for small values of j and small values for relatively larger values of j. Note also that if the samples come from populations which do not differ in location, but population i has a greater dispersion than the other populations, then it is expected that $U_{ij}$ will be relatively large for extreme values of j (near either 0 or K) and relatively small values of j near the median rank.
The following numerical example makes the algebraic descriptions more lucid.

<table>
<thead>
<tr>
<th>Sample 1</th>
<th>Sample 2</th>
<th>Sample 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>7</td>
</tr>
</tbody>
</table>

Then 12 K-tuples (triples) can be formed:

- \((1,3,6)\)
- \((1,3,7)\)
- \((1,5,6)\)
- \((1,5,7)\)
- \((2,3,6)\)
- \((2,3,7)\)
- \((2,5,6)\)
- \((2,5,7)\)
- \((4,3,6)\)
- \((4,3,7)\)
- \((4,5,6)\)
- \((4,5,7)\)

\[
v_{11} = 10 \\
v_{12} = 2 \\
v_{13} = 0 \\
v_{21} = 2 \\
v_{22} = 10 \\
v_{23} = 0 \\
v_{31} = 0 \\
v_{32} = 0 \\
v_{33} = 12
\]

\[
u_{11} = \frac{5}{6} \\
u_{12} = \frac{1}{6} \\
u_{13} = 0 \\
u_{21} = \frac{1}{6} \\
u_{22} = \frac{5}{6} \\
u_{23} = 0 \\
u_{31} = 0 \\
u_{32} = 0 \\
u_{33} = 1
\]

The V-Test

In accordance with Bhapkar (1961), define

\[
V = n(2k-1)\left[\sum_{i=1}^{k} p_i(U_{11} - K^{-1})^2 - \left\{\sum_{i=1}^{k} p_i(U_{11} - K^{-1})\right\}^2\right]
\]

where \(p_i = n_i/n\). Under the null hypothesis, \(E[U_{11}|H_0] = K^{-1}\). Hence, \(V\) is a measure of deviation from the null hypothesis since, under \(H_0\), \(V\) tends to have small expected value, and under translation alternatives, \(V\) tends to be larger. Bhapkar shows that the asymptotic
distribution of $V$ is chi-squared with $K-1$ degrees of freedom, and
with a noncentrality parameter that is zero under the null hypothesis
and positive under translation alternatives.

Example:  $V = (7)(2.3-1) \left[ \frac{3}{7} \sum p_i (u_{i1} - \frac{1}{3})^2 - \left[ \frac{3}{7} \sum p_i (u_{i1} - \frac{1}{3}) \right]^2 \right]$

$= 35 \left[ \frac{3}{7} \left( \frac{5}{6} - \frac{1}{3} \right)^2 + \frac{2}{7} \left( \frac{1}{6} - \frac{1}{3} \right)^2 + \frac{2}{7} \left( 0 - \frac{1}{3} \right)^2 - \frac{3}{7} \left( \frac{5}{6} - \frac{1}{3} \right) \right.$

$+ \frac{2}{7} \left( \frac{1}{6} - \frac{1}{3} \right) + \frac{3}{7} \left( 0 - \frac{1}{3} \right)^2 \right]$

$= 35 \left[ \frac{250}{1764} \right] = 4.960$

$P = P_r \left\{ \chi^2(2) > 4.96 \right\} = 0.0837$

The $L$-Test

In accordance with Deshpande (1965a), define $\ell_i = -u_{i1} + u_{iK}$
for $i = 1, 2, ..., K$. Note that if the observations from sample $i$ are
smaller than other observations, $\ell_i = -1$, and if they are larger
than other observations, $\ell_i = +1$. Thus define the test statistic

$L = \frac{n(2K-1)(K-1)}{2K^2 \{ (2K-2) \choose K-1 \} - 1} \left[ \left\{ \sum_{i=1}^k p_i \ell_i \right\}^2 - \left\{ \sum_{i=1}^k p_i \ell_i \right\} \right]$

The $L$ statistic is sensitive to shift alternatives, but does not
appear to be sensitive to differences in scale parameters. Deshpande
(1965a) shows that the asymptotic null distribution of $L$ is chi-
squared with $K-1$ degrees of freedom. The null hypothesis is rejected
for sufficiently large values of L.

Example: \[ L = \frac{(7)(5)(4)(6)}{(2)(9)(5)} \cdot \left[ \frac{3}{7} \left( -\frac{5}{6} + 0 \right)^2 + \frac{2}{7} \left( -\frac{1}{6} + 0 \right)^2 + \frac{2}{7} \left( -0 + 1 \right)^2 \right. \\
\left. - \left\{ \frac{3}{7} \left( -\frac{5}{6} \right) + \frac{2}{7} \left( \frac{1}{6} \right) + \frac{2}{7} \left( 1 \right) \right\} \right] \]

\[ = 9 \frac{1}{3} \left[ \frac{1018}{1764} \right] = 5.386 \]

\[ P = P_i \left\{ X^2(2) > 5.386 \right\} = 0.0677 \]

The W-Test

Define \( w_i = \sum_{j=1}^{k} (j-1)u_{ij}, \) \( i = 1,2,...,K. \) Note that if sample \( i \) contains the smallest observations, \( w_i = 0; \) if sample \( i \) contains the largest observations, then \( w_i = -1. \) Then the test statistic proposed is

\[ W = \frac{12n}{K^2} \left[ \sum_{i=1}^{k} P_i w_i^2 - \left\{ \sum_{i=1}^{k} P_i w_i \right\}^2 \right] \]

It may be noted that if \( K=3, \) the L-test and the W-test are identical. Bhapkar shows that in the special case where \( n_i = N, \) say, the W-statistic is equivalent to the Kruskal-Wallis H. The null asymptotic distribution of \( W \) is chi-squared with \( K-1 \) degrees of freedom, and the null hypothesis is rejected for large values of \( W. \)

Example: \( W_1 = \frac{1}{6}; \) \( W_2 = \frac{5}{6}; \) \( W_3 = 2 \)

\[ W = \frac{12(7)}{9} \left[ \frac{3(1)^2}{7(6)} + \frac{2(5)^2}{7(6)} + \frac{2(2)^2}{7(6)} - \left\{ \frac{3(1)}{7(6)} + \frac{2(5)}{7(6)} + \frac{2(2)}{7(6)} \right\}^2 \right] \]

\[ = \frac{4(7)}{3} \left[ \frac{1018}{1764} \right] = 5.386, \text{ as expected (see L-test).} \]
**REMARKS:** Bhapkar (1964) discusses the asymptotic efficiencies of these tests. The asymptotic distribution of the \( W \)-test is the same as that of the Kruskal-Wallis \( H \)-test, and, under the same sequence of alternatives, the \( \text{ARE}(W, H) = 1 \), and hence, \( \text{ARE}(W, F) = 3/\pi \) (where \( F \) is the usual analysis of variance (ANOVA \( F \))), if the underlying distributions are normal. The \( V \)-test (i.e., the test based on the number of \( K \)-tuples for which the observations from the \( i \)th sample are least) is much more efficient for populations bounded below. The \( L \)-test (based on numbers of \( K \)-tuples with respect to both the smallest and largest observations) is much more efficient for distributions bounded above and below. The \( W \)-test appears to be more efficient for unbounded distributions.

D. The AMP Sign Test

Quade (1967) suggested an all matched pairs statistic, to use all matched pairs rather than just independent ones. Schoenfelder (1981) developed the AMP sign statistic

\[ T_{n_1n_2} = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I(Y_{2j} > Y_{1j}, X_{2j} = X_{1i}) \]

where

- \((X_{1i}, Y_{1i}); i = 1, \ldots, n_1\) is a random sample from population 1,
- \((X_{2j}, Y_{2j}); j = 1, \ldots, n_2\) is a random sample from population 2.

The statistic \( T_{n_1n_2} \) is the number of matched pairs where the observation from the second sample is larger on \( Y \). The statistic is a sum of Mann-Whitney statistics, and, as such, is a member of
a general class of weighted AMP sign statistics, each member of which is representable as

\[ Q_{n_1n_2} = \sum_{c=1}^{C} w_c U_c \]

where \( w_c \geq 0 \) are appropriately defined weights and \( U_c \) is the Mann-Whitney statistic defined for the subcollection of observations matched on \( X \) with \( X=\bar{X}_c \). Schoenfelder derives optimal weights and proves that under the null hypothesis of no difference \( Q_{n_1n_2}^* \) is asymptotically normal, where

\[ Q_{n_1n_2}^* = \sum_{c=1}^{C} p_c^{-1} U_c \]

and

\[ p_c = P(X = \bar{X}_c), \quad c=1,2,...,C \quad (\sum p_c = 1, \quad p_c > 0 \forall c) \]

is the probability function of \( X \), and, under homogeneity, Schoenfelder, p. 98, shows:

\[ E(Q_{n_1n_2}^*) = (1/2) \, n_1n_2 \]

\[ \text{Var}(Q_{n_1n_2}^*) = (1/2)n_1n_2k + (1/3)n_1n_2(n_1+n_2-2) - (1/4)n_1n_2(n_1+n_2-1). \]

The optimal weights were derived assuming that the \( p_c \)'s were known. If they are unknown, analysis would proceed using estimates of the optimal weights; hence, the statistic with estimated optimal weights is

\[ \hat{Q}_{n_1n_2}^* = \sum_{c=1}^{C} \left( \frac{n_c}{n} \right)^{-1} U_c. \]

Since \( \hat{p}_c = \frac{n_c}{n} \) is a consistent estimator for \( p_c \), it follows (from
Slutsky's theorem (1925)) that $\tilde{Q}_{n_1 n_2}$ and $Q_{n_1 n_2}$ have the same limiting power (see Schoenfelder, 1981, p. 127).

In this dissertation two extensions of Schoenfelder's AMP sign test are presented for comparing $K$ populations with respect to a continuous univariate response. The covariable is assumed to be discrete and is possibly multivariate.

The first extension is presented in Chapter II, and is called theAMP sign test for $K$ samples. It is based on the Kruskal-Wallis test, basing the test on a combination of sets of two-sample statistics. The null asymptotic distribution is found to be chi-squared with $K-1$ degrees of freedom; the asymptotic distribution under translation-type alternatives is shown to be noncentral chi-squared with $K-1$ degrees of freedom, and the noncentrality parameter is derived. The AMP sign test for $K$ samples is shown to have Pitman ARE of unity with respect to the Kruskal-Wallis test, when the covariable is independent of the response variable.

The second extension is presented in Chapter III, and is called the all matched $K$-tuple (AMK) sign test. This test is based on Bhapkar's $V$-test, and analyzes matched $k$-tuples. The asymptotic distribution of the test statistic is found to be central chi-squared ($K-1$ degrees of freedom) under the null hypothesis, and noncentral chi-squared under translation-type alternatives. (The noncentrality parameter is derived.) In the case where the covariable is independent of the response variable, the AMK sign test is shown to have ARE of unity relative to Bhapkar's $V$-test.

In Chapter IV, the ARE results given above are derived. Also,
the two new methods are compared with respect to each other via ARE results for certain conditional distributions. The results suggest that if the conditional distribution of the response variable given the covariable is bounded below, then the AMK methods tend to be superior to AMP methods; otherwise, AMP methods are preferred.

Finally, Chapter V summarizes the salient features of this research and discusses suggestions for future research.
CHAPTER II

THE EXTENSION BASED ON THE KRUSKAL-WALLIS TEST

In this chapter, a method of analysis of covariance by matching is presented that is an extension of Schoenfelder's All Matched Pairs (AMP) sign test to the general \( K \geq 2 \) sample problem that is based on the Kruskal-Wallis test. As such, the \( K \)-sample comparison is based on the combination of two-sample statistics.

A. The Basic Framework and Notation

The Data:

The data are assumed to be representable as vectors \((X_{ij}, Y_{ij})\) where \( i = 1,2,\ldots,K \) (indexing the populations) and \( j_1 = 1,2,\ldots,n_i \) (indexing the observations within populations); \( n_i \) is the number of observations from population \( i \), and \( n = \sum n_i \). The covariable \( X_i \) (possibly multivariate) which is incorporated into the analysis is assumed to be a discrete random variable with finite support, so that it takes on possible values \( V_c \), for \( c=1,2,\ldots,C \) (\( C \) is the number of possible values of \( X \)), with probabilities \( \varepsilon_c = \Pr[X_{ij} = V_c], \varepsilon_c > 0 \) for all \( C \), and \( \sum \varepsilon_c = 1 \). The response variable \( Y \) is univariate and continuous. The continuity assumption is made for statistical convenience so that the treatment of ties need not be considered.
The Hypotheses:

The general hypothesis of homogeneity for K populations is

\[ H_0: F_i(y) = F(y) \quad \text{for all } y=1,2,\ldots,K, \]

where \( F_i(y) \) denotes the cumulative distribution function for the \( i \)th population. The \( K \)-sample location model asserts that the populations are identical except possibly for a shift. This chapter presents a test based on the Kruskal-Wallis test which does not require the location model for its validity, however, is primarily sensitive to shift differences.

The issue of current interest is that of comparing \( K \) populations for the purpose of detecting shift differences in the presence of a covariable. Hence, the null hypothesis incorporating \( X \) may be expressed

\[ H_0: F_i(Y|X) = F(Y|X), \quad i=1,2,\ldots,K. \]

Note that nothing is assumed about the form of the distribution of \( X \) asserted to be common to the \( K \) populations, simply that the distributions are identical. The particular alternative considered is that the conditional distributions differ by shifts. It is recognized that there may be other distribution differences, but the power of the test presented in this chapter is aimed at detecting shift differences. Hence, its appropriateness for any given testing situation must be considered by the investigator.

The Assumptions:

As has already been stated, the general framework assumes that
the data are from K independent random samples, hence the observations (which are vectors) are independent within samples and among samples. The following three assumptions are made throughout this presentation:

A.1 The covariable X is concomitant and has finite support.
A.2 The response variable Y is continuous and univariate.
A.3 Conditional on the X's, the Y's are independent.

Note that Assumption A.2 is not required for the strict validity of the presented analysis technique, but is rather for statistical convenience so that ties on Y may be ignored.

B. Parameter Definitions

The parameters describing probabilities of certain rankings among observations must be defined, conditionally and unconditionally.

**Definition 2.1:** Let $\tau_{ii'}$ denote the [unconditional] probability that of two randomly selected observations, one from population i and one from population i', the one from population i' is larger on Y.

**Definition 2.2:** Let $\tau_{ii'i''i'''}$ denote the probability that of three [four] randomly selected observations, one each from populations i, i', i'', [i'''], (not necessarily distinct), their relative ranks are i, i', i'', [i'''],

**Definition 2.3:** To each $\tau$ there corresponds a $\theta$ (for example: $\theta_{ii'}(V_C)$ to $\tau_{ii'}$) which is the conditional probability of the event given the observations are matched on X with $X = V_C$. 

Definition 2.4: To each $\tau$ there corresponds a $\gamma$ (for example: $\gamma_{ij} \equiv \tau_{ij}$) which is the probability of the event given that all $X$ are matched; these are called matched probabilities.

Definition 2.5: Let $\beta_m$ denote the probability that $m > 1$ randomly selected observations match on $X$. For the special case $m = 2$, the subscript $K$ will be dropped, i.e., $\beta \equiv \beta_2$.

Proposition 2.1: Let the $\theta$'s and $\gamma$'s be the conditional probabilities and the matched probabilities defined above. Then

\[ \gamma_{ii'} = \sum_C \varepsilon^2 \theta_{ii'}(V_C)/\beta_2 \]
\[ \gamma_{ii'i''} = \sum_C \varepsilon^3 \theta_{ii'i''}(V_C)/\beta_3 \]
\[ \gamma_{ii'i''i'''} = \sum_C \varepsilon^4 \theta_{ii'i''i'''}(V_C)/\beta_4. \]

Proof: This is a straightforward application of the result of Schoenfelder (1981, p. 39).

C. Number of Analysis Units

Since the extension of Schoenfelder's AMP analysis based on the Kruskal-Wallis test is an analysis performed on all matched pairs (recall, all pairs $(X_{ij}, Y_{ij})$ and $(X_{i'j'}, Y_{i'j'})$ such that $i \neq i'$ and $X_{ij}$ and $X_{i'j'}$ match), then the number of units of analysis (the number of matched pairs) is a random variable, say $A$.

\[ A = \sum_{i < i', j' \neq j} \sum_{i \neq i'} \sum_{j \neq j'} M(X_{ij}, X_{i'j'}, X_{i'j'}) \]

where
\[ M(X_{ij}, X'_{i'j'}) = \begin{cases} 
1 & \text{if } X_{ij} \text{ and } X'_{i'j'} \text{ match} \\
0 & \text{else.} 
\end{cases} \]

If all pairs match, then clearly the number of analysis units is

\[ \sum_{i < i'} n_i n_{i'} = \frac{n^2 - \Sigma n_i^2}{2}. \]

In general, this is not expected. It is desirable to determine \( E(A) \) and \( \text{Var}(A) \).

**Theorem 2.1:** Let \( A_{ij} \) be the number of matched pairs from samples \( i \) and \( i' \), respectively, with sample sizes \( n_i \) and \( n_{i'} \), satisfying Assumption A.1. Then,

i) \( E(A_{ij}) = n_i n_{i'}, \beta \)

ii) \( \text{Var}(A_{ij}) = n_i n_{i'} [\beta + (n_i + n_{i'} - 2)\beta_3 - (n_i + n_{i'} - 1)\beta^2] \),

where \( \beta \) and \( \beta_3 \) are as defined above.

**Proof:** See Schoenfelder (1981).

It is worth noting that \( A_{ij}, n_i n_{i'} \) is a two-sample U-statistic for estimating \( \beta \).

**Theorem 2.2:** Let \( A \) be the number of matched pairs from \( K \) independent random samples of sizes \( n_i, i=1,2,...,K \), satisfying Assumption A.1. Then,

i) \( E(A) = \beta \left( \frac{n^2 - \Sigma n_i^2}{2} \right) \)

ii) \( \text{Var}(A) = (\beta_3 - \beta^2) \left( \frac{2}{3} n^3 + \frac{1}{3} \Sigma n_i^3 - n \Sigma n_i^2 \right) + (\beta - 2\beta_3 + \beta^2) \left( \frac{n^2 - \Sigma n_i^2}{2} \right) \).
Proof:

i) It is obvious that

\[ A = \sum_{i < i'} A_{i,i'}. \]

\[ E[A] = E[\sum_{i < i'} A_{i,i'}] = \sum_{i < i'} E(A_{i,i'}) = \sum_{i < i'} n_i n_{i'} \beta^2 = \beta \left( \frac{n_i^2 - \sum n_i^2}{2} \right) \]

ii) \( \text{Var}(A) = \text{Var}(\sum_{i < i'} A_{i,i'}) = \sum_{i < i'} \text{Var}(A_{i,i'}) \)

\[ + 2 \sum_{i < i'} \sum_{i < i'} \text{cov}(A_{i,i'}, A_{i',i'}) \]

\[ + 2 \sum_{i < i'} \sum_{i < i'} \text{cov}(A_{i,i'}, A_{i',i'}) \]

\[ + 2 \sum_{i < i'} \sum_{i < i'} \text{cov}(A_{i,i'}, A_{i',i'}). \]

\[ \text{Cov}(A_{i,i'}, A_{i',i'}) = E[A_{i,i'} A_{i',i'}] - E[A_{i,i'}] E[A_{i',i'}] \]

\[ = E \left[ \sum_{i=1}^{n_i} \sum_{j=1}^{n_i} M(X_{i,j}, X_{i',j'}) \cdot \sum_{k=1}^{n_i} \sum_{k'=1}^{n_i} M(X_{i,k}, X_{i',k'}) \right] \]

\[ - n_i^2 n_i n_i' \beta^2 \]

\[ = E \left\{ \sum_{k \neq j} \sum_{j \neq k} \sum_{k \neq j} M(X_{i,j}, X_{i',j'}) \cdot M(X_{i,k}, X_{i',k'}) \right\} - n_i^2 n_i n_i' \beta^2 \]

\[ + \left\{ \sum_{k \neq j} \sum_{j \neq k} M(X_{i,j}, X_{i',j'}) \cdot M(X_{i,k}, X_{i',k'}) \right\} - n_i^2 (n_i - 1) \beta^2 \]

\[ + n_i n_i' n_i'' (n_i - 2) \beta^2 \]

\[ = \left\{ n_i n_i n_i'' (n_i - 1) \beta^2 + n_i n_i' n_i'' \beta^2 \right\} - n_i^2 n_i n_i' \beta^2 \]

Similarly, \( \text{cov}(A_{i,i'}, A_{i',i''}) = n_i n_i n_i'' [(n_i - 1) \beta^2 + \beta^2] - n_i n_i^2 n_i' \beta^2. \)
So,
\[ \text{Cov}(A_{i,i'}^{(n)}, A_{i,i''}^{(n)}) = n_i n_{i'} n_{i''} (\beta_2 - \beta^2) \]
\[ \text{Cov}(A_{i,i'}^{(n)}, A_{i'',i''}^{(n)}) = n_i n_{i'} n_{i''} (\beta_3 - \beta^2) \]
\[ \text{Cov}(A_{i,i'}^{(n)}, A_{i'''}^{(n)}) = 0, \text{ clearly, when } i, i', i'', i''' \text{ are distinct.} \]

And
\[ \text{Var}(A) = \sum_{i<i'} \text{Var}(A_{i,i'}) + 4 \sum_{i<i'<i''} n_i n_{i'} n_{i''} (\beta_3 - \beta^2) \]
\[ \sum_{i<i'<i''} n_i n_{i'} n_{i''} = \frac{1}{6} n^3 + \frac{1}{3} \Sigma n_i^3 - \frac{1}{2} n \Sigma n_i^2 \]
\[ \sum_{i<i'} n_i n_{i'} (n_i + n_{i'}) = n \Sigma n_i^2 - \Sigma n_i^3 \]

\[ \text{Var}(A) = \sum_{i<i'} n_i n_{i'} \left[ \beta + (n_i + n_{i'} - 2) \beta_3 - (n_i + n_{i'} - 1) \beta^2 \right] \]
\[ + 4 \left[ \frac{1}{6} n^3 + \frac{1}{3} \Sigma n_i^3 - \frac{1}{2} n \Sigma n_i^2 \right] (\beta_3 - \beta^2) \]
\[ = \left( \beta - 2 \beta_3 + \beta^2 \right) \left( \frac{n^2 - \Sigma n_i}{2} \right) + \left[ \frac{2}{3} n + \frac{1}{3} \Sigma n_i^3 - n \Sigma n_i^2 \right] \left( \beta_3 - \beta^2 \right) \]

D. The AMP Sign Statistic for K Samples

Define,
\[ U_{C_{i,i'}}^{n} = \sum_{j_i=1}^{n_i} \sum_{j_{i'}=1}^{n_{i'}} \text{sgn}(Y_{i,i'}^{j_i,j_{i'}} - Y_{i,i'}^{j_{i'},j_{i'}}) I\{X_{i,i'}^{j_i,j_{i'}} = X_{i,i'}^{j_{i'},j_{i'}} = V_C \} \]
as the centered Mann-Whitney statistic for comparing samples \( i \) and \( i' \) using only those observations for which the covariable assumes the value \( V_C \). Then Schoenfelder's unweighted AMP sign statistic is
\[ T_n = \sum_c \sum_c U_{C_{i,i'}}^{n} \]
the general weighted AMP statistic is

\[ Q_n = \sum_c W_c U_{C12} \]

where the \( W_c \) are general weights,

\[ Q^*_n = \sum_c \rho_{c}^{-1} U_{C12} \]

is the optimally weighted AMP sign statistic, and

\[ \tilde{Q}^*_n = \sum_c \left( \frac{m_c}{n} \right)^{-1} U_{C12} \]

is the optimally weighted AMP sign statistic with estimated weights, where \( m_c \) is the number of observations available for which \( X = V_c \).

While it would be desirable to derive optimal weights for the AMP sign statistic for the K-sample problem, complications arise in trying to invert a complicated covariance matrix. Consequently, in an attempt to simplify the presentation, a particular weighted AMP sign test for K-samples will be developed and its properties (in terms of ARE) discussed later.

At this point, another assumption is imposed on the framework. In addition to Assumptions A.1 - A.3, assumption A.4 is now in force.

**Assumption A.4:** The K populations are parallel in pairwise probabilities.

**Definition 2.6:** If for every pair of populations, given one observation from each member of the pair, with the two \( X \)'s matched at \( V_c \), the expected ordering of the \( Y \)'s does not depend on \( c \), then the populations are said to be parallel in pairwise probabilities.
Corollary 2.1: If the populations are parallel in pairwise probabilities, then for all values $\gamma_c$ of $\gamma$,

$$\theta_{ii'}(\gamma_c) = \gamma_{ii'}.$$ 

Lemma 2.1: Let $\gamma_{ij}$, $j = 1, 2, ..., n_i$ for a fixed $i$, be independent (real or vector) random variables, identically distributed with c.d.f. $F_i$, $i = 1, 2, ..., K$. Further, let $n = \sum n_i$ and

$$U_n(r) = \left[ \prod_{i=1}^{k} \binom{n_i}{m_i(r)} \right]^{-1} \sum h(r) \left( \gamma_{1 \alpha_1}, ..., \gamma_{1 \alpha_{m_1}}, ..., \gamma_{2 \beta_1}, ..., \gamma_{2 \beta_{m_2}}, ..., \gamma_{k \delta_1}, ..., \gamma_{k \delta_{m_k}}(r) \right),$$

where each $h(r)$ is a function symmetric in each set of its arguments and $\sum$ denotes the sum over all combinations $(\alpha_1, ..., \alpha_{m_1}(r))$ of $m_1(r)$ integers chosen from $(1, 2, ..., n_1)$, and so on for the $\beta$'s and the $\delta$'s.

Assume that $E[h(r)] = 0(r)$ and $E[h(r)]^2 < \infty$. Then

i) $E[U_n(r)] = 0(r)$

ii) $\text{Cov}[U_n(r), U_n(s)] = \left[ \prod_{i=1}^{k} \binom{n_i}{m_i(s)} \right]^{-1} \sum_{d_i=0}^{m_i(r, s)} \left( \prod_{i=1}^{k} \binom{m_i(r)}{d_i} \right) \binom{n_i - m_i(r)}{m_i(s) - d_i} \xi_{d_i, d_2, ..., d_k, (r, s)},$

where $m_i(r, s) = \min(m_i(r), m_i(s))$ and

$$\xi_{d_1, d_2, ..., d_k, (r, s)} = E[h(r)(\gamma_{1 1}, ..., \gamma_{1 d_1, 1}, \gamma_{1 d_1+1}, ..., \gamma_{1 m_1(r)}; ...; \gamma_{k 1}, ..., \gamma_{k d_k, 1}, ..., \gamma_{k m_k(r)})]$$
\[ x h(s) Y_{11}, \ldots, Y_{\varphi_{11}}, \ldots, Y_{1m_1(r) + 1}, \ldots, Y_{1, m_1(r) + m_1(s) - d_1}; \ldots; \]

\[ Y_{k1}, \ldots, Y_{k\varphi_{kd_k}}, Y_{k, m_k(r) + 1}, \ldots, Y_{k, m_k(r) + m_k(s) - d_k} \] - \( \varphi(r) \varphi(s) \),

\( r, s = 1, 2, \ldots, g \).

Note that if \( r = s \), this produces \( \text{var}[U_n^{(r)}] \), and

iii) \( n^2 [U_n - \bar{\varphi}] \) is, in the limit as \( n \to \infty \) in such a way that \( n = na_i \), the \( a_i \)'s being fixed numbers such that \( \sum a_i = 1 \), normally distributed with zero mean vector and asymptotic covariance matrix.

\[ \Sigma = (\sigma_{rs}) \] given by

\[ \sigma_{rs} = \sum_{i=1}^{k} \frac{m_i(r) m_i(s)}{a_i} \epsilon_{0,0, \ldots, 0,1,0, \ldots, 0} \] \( (r, s) \),

\( (1 \text{ at the } i^{\text{th}} \text{ place}) \) \( r, s = 1, 2, \ldots, g \),

where

\[ U_n = (U_n^{(1)}, \ldots, U_n^{(g)}) \] and

\[ \varphi = (\varphi^{(1)}, \ldots, \varphi^{(g)}). \]

**Proof:** This lemma is given in the above version in Bhapkar (1961). The generalized U-statistics results are a straightforward extension of Hoeffding's theorem (1948), and the details are not supplied here.

Define the parameters

\[ \delta_i = \sum_{i' \neq 1}^{k} (Y_{ii'} - Y_{i'i}) \] for \( i = 1, 2, \ldots, K \)

and note that under \( H_0 \)
\[ \hat{\delta} = (\hat{\delta}_1, \hat{\delta}_2, \ldots, \hat{\delta}_K) = \hat{0}. \]

To estimate \( \hat{\delta}_i \) one may use the U-statistic
\[ T_i = (\prod n_i)^{-1} W_i \]
in which the numerator is
\[
W_i = \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \cdots \sum_{j_k=1}^{n_k} h^{(i)}[(X_{ij_1}, Y_{ij_1}, X_{ij_2}, Y_{ij_2}, \ldots, \ X_{ij_k}, Y_{ij_k})] \]
and the kernel is
\[
h^{(i)}[(X_{ij_1}, Y_{ij_1}, X_{ij_2}, Y_{ij_2}, \ldots, X_{ij_k}, Y_{ij_k})] = \sum_{i' = 1}^{k} \text{sgn}(Y_{i'j_i} - Y_{ij_i}) \sum_{c} \epsilon_c^{-1} I\{X_{ij_i} = X_{i'j_i} = Y_{ij_i} = Y_{i'j_i} = c\} \]
so that
\[
E[h^{(i)}(\cdot)] = \left[ \sum_{i' = 1}^{k} \text{sgn}(Y_{i'j_i} - Y_{ij_i}) \sum_{c} \epsilon_c^{-1} I\{X_{ij_i} = X_{i'j_i} = Y_{ij_i} = Y_{i'j_i} = c\} \right] \]
\[
= \sum_{c} \epsilon_c^{-1} \epsilon_c^{-1} \sum_{i' = 1}^{k} \sum_{i' \neq i} \left[ \theta_{i'j_i} - \theta_{i'j_i} \right] \]
Under pairwise parallelism,
\[
E[h^{(i)}(\cdot)] = \sum_{c} \epsilon_c \sum_{i' \neq i}^{k} (Y_{i'j_i} - Y_{ij_i}) = \delta_i. \]
Let
\[ T_n = (T_1, T_2, \ldots, T_K) \]
and \[ \hat{\Sigma} = (\sigma_{rs}) \], the asymptotic covariance matrix of \( \sqrt{n} (T_n - \hat{\delta}) \). An
The expression for $\sum$ is given in the following theorem.

**Theorem 2.3:** Let $T_n$ be the vector of $K$-sample $U$-statistics defined above, and $\delta$ the vector of corresponding expected values (under $H_0$: $\delta \equiv 0$). Then the asymptotic covariance matrix of $\sqrt{n} T_n$ (under $H_0$) is $\Sigma = (\Sigma_{rs})$, given by

$$
\sigma_{rr} = \frac{1}{3} (k-1)^2 a_r - 1 + \frac{1}{3} \sum_{i \neq r} \frac{k}{a_i - 1} = \frac{1}{3} \left[ \sum_{i = 1}^{k} \frac{1}{a_i - 1} + (k^2 - 2k) a_r - 1 \right]
$$

$$
\sigma_{rs} = \frac{1}{3} (k-1)[a_r - 1 + a_s - 1] + \frac{1}{3} \sum_{i \neq r} \frac{k}{a_i - 1} = \frac{1}{3} \left[ \sum_{i = 1}^{k} \frac{1}{a_i - 1} - k(a_r - 1 + a_s - 1) \right]
$$

where $n_i = n a_i$.

**Proof:** From Lemma 2.1 the asymptotic covariance matrix of $\sqrt{n} T_n$ is given by $(\Sigma_{rs})$, where

$$
\sigma_{rs} = \sum_{i = 1}^{k} a_i - 1 \xi_{0,0,1,0,\ldots,0}, \quad r, s = 1, 2, \ldots, K
$$

(1 in position $i$)

Hence the $\xi$'s are derived.

$$
(1) \xi_{1,0,0,\ldots,0} = E[h^{(1)}(Y_{1j_1}, Y_{2j_2}, \ldots, Y_{kj_k}) \cdot h^{(1)}(Y_{1j_1}, Y_{2j_2}, \ldots, Y_{kj_k})] - \left[ E[h^{(1)}(\cdot)] \right]^2
$$

$$
= E \left\{ \sum_{i' = 2}^{k} \text{sgn}(Y_{i',j_i'} - Y_{1j_1}) \sum_{c} \varepsilon_c^{-1} I \left\{ X_{i',j_i'} = X_{1j_1} = V_c \right\} \right. \\
\left. \cdot \sum_{i'' = 2}^{k} \text{sgn}(Y_{i''j_i''} - Y_{1j_1}) \sum_{c'} \varepsilon_{c'}^{-1} I \left\{ X_{i''j_i''} = X_{1j_1} = V_{c'} \right\} \right\}
$$
\[ E \left\{ \sum_{i=2}^{k} \sum_{i''=2}^{k} \text{sgn}(Y_{i',j_i}, -Y_{1,j_1}) \text{sgn}(Y_{i''},j_i'' - Y_{1,j_1}) \sum_{c} \varepsilon_c^{-1} \varepsilon_{c'}^{-1} \right\}
\]
\[ \cdot I \left\{ \frac{X_{i',j_i}}{z_{1,j_1}} = V_{c'} \right\} I \left\{ \frac{X_{i'',j_i}}{z_{1,j_1}} = V_{c} \right\} \]
\[ = \sum_{c} \varepsilon_c^{-2} \sum_{i=2}^{k} \sum_{i''=2}^{k} \left\{ \text{Pr}(Y_{1,j_1} < Y_{i',j_i}, < Y_{i'',j_i''} | X_{1,j_1} = X_{i',j_i}, = X_{i'',j_i''} = V_{c}) \right\} \]
\[ + \text{Pr}(Y_{1,j_1} < Y_{i'',j_i''} | X_{1,j_1} = X_{i',j_i}, = X_{i'',j_i''} = V_{c}) \]
\[ + \text{Pr}(Y_{i',j_i}, < Y_{i'',j_i''} | X_{1,j_1} = X_{i',j_i}, = X_{i'',j_i''} = V_{c}) \]
\[ + \text{Pr}(Y_{i'',j_i''} < Y_{i',j_i} | X_{1,j_1} = X_{i',j_i}, = X_{i'',j_i''} = V_{c}) \]
\[ - \text{Pr}(Y_{i',j_i} < Y_{i'',j_i''} | X_{1,j_1} = X_{i',j_i}, = X_{i'',j_i''} = V_{c}) \]
\[ - \text{Pr}(Y_{i'',j_i''} < Y_{i',j_i} | X_{1,j_1} = X_{i',j_i}, = X_{i'',j_i''} = V_{c}) \} \]
\[ = \sum_{c} \varepsilon_c \cdot \sum_{i=2}^{k} \sum_{i''=2}^{k} \left( \frac{4}{6} - \frac{2}{6} \right) \]
\[ = \frac{1}{3} (k-1)^2 \sum_{c} \varepsilon_c \]
\[ \varepsilon_{1,0, \ldots, 0} = \frac{1}{3} (k-1)^2 \]

Similarly,
\[ \varepsilon_{0, \ldots, 0, 1, 0, \ldots, 0} = \frac{1}{3} (k-1)^2 , \quad r = 1, 2, \ldots, k \]
\[ (1 \text{ in } i\text{th place}) \]
\[
\zeta_{0, 1, 0, \ldots, 0} = E[h^{(1)}(Y_{i_1}^1, Y_{2j_2}^2, \ldots, Y_{kj_k}^k) \cdot h^{(1)}(Y_{i_1}^1, Y_{2j_2}^2, Y_{3j_3}^3, \ldots, Y_{kj_k}^k)] - \{E[h^{(1)}(\cdot \cdot \cdot)]\}^2
\]

\[
= E\left[ \sum_{i_1^1=1}^{k} \text{sgn}(Y_{i_1}^{i_1^1}, - Y_{1j_1}) \sum_c c \cdot I\left\{X_{i_1}^{i_1^1} = X_{-1j_1} = V_c\right\} \cdot \sum_{i_1^2=2}^{k} \text{sgn}(Y_{i_1}^{i_1^2}, - Y_{1j_1}) \sum_c c \cdot I\left\{X_{i_1}^{i_1^2} = X_{-1j_1} = V_c\right\} \right]
\]

\[
= E\left[ \sum_{i_1^1=2}^{k} \sum_{i_1^2=2}^{k} \text{sgn}(Y_{i_1}^{i_1^1}, - Y_{1j_1}) \text{sgn}(Y_{i_1}^{i_1^2}, - Y_{1j_1}) \times \sum_c c \cdot c \cdot I\left\{X_{i_1}^{i_1^1} = X_{-1j_1} = V_c\right\} \times \sum_c c \cdot c \cdot I\left\{X_{i_1}^{i_1^2} = X_{-1j_1} = V_c\right\} \right]
\]

\[
= E\left[ \sum_{i_1^1=2}^{k} \sum_{i_1^2=2}^{k} \text{sgn}(Y_{i_1}^{i_1^1}, - Y_{1j_1}) \text{sgn}(Y_{i_1}^{i_1^2}, - Y_{1j_1}) \times \sum_c c \cdot c \cdot I\left\{X_{i_1}^{i_1^1} = X_{-1j_1} = V_c\right\} \times \sum_c c \cdot c \cdot I\left\{X_{i_1}^{i_1^2} = X_{-1j_1} = V_c\right\} \right]
\]

\[
+ \text{sgn}(Y_{2j_2} - Y_{1j_1}) \text{sgn}(Y_{2j_2} - Y_{1j_1}) \times \sum_c c \cdot c \cdot I\left\{X_{2j_2} = X_{-1j_1} = V_c\right\} \times \sum_c c \cdot c \cdot I\left\{X_{2j_2} = X_{-1j_1} = V_c\right\}
\]

\[
= \sum_c c \cdot c \cdot 4 \sum_{i_1^1=2}^{k} \sum_{i_1^2=2}^{k} 0 \times \sum_c c \cdot c \cdot \left[ \text{Pr}(Y_{2j_2} < Y_{1j_1} < Y_{1j_1} \mid X_{1j_1} = X_{2j_2} = V_c) \right]
\]

\[
+ \text{Pr}(Y_{2j_2} < Y_{1j_1} < Y_{1j_1} \mid X_{1j_1} = X_{2j_2} = V_c) \left\{ X_{2j_2} = X_{1j_1} = V_c \right\}
\]

\[
+ \text{Pr}(Y_{1j_1} < Y_{1j_1} < Y_{2j_2} \mid X_{1j_1} = X_{2j_2} = V_c)
\]
\[ + \Pr(Y_{1j_1} < Y_{1j_1} < Y_{2j_2} \mid X_{1j_1} = X_{1j_1} = X_{2j_2} = V_c) \]
\[ - \Pr(Y_{1j_1} < Y_{2j_2} < Y_{1j_1} \mid X_{1j_1} = X_{1j_1} = X_{2j_2} = V_c) \]
\[ - \Pr(Y_{1j_1} < Y_{2j_2} < Y_{1j_1} \mid X_{1j_1} = X_{2j_2} = X_{1j_1} = V_c) \]
\[ = \sum_{c} \sum_{c'} \frac{(4}{6} \cdot \frac{2}{6} = \frac{1}{3}. \]

\[ \zeta_{0,1,0,\ldots,0}^{(1)} = \frac{1}{3} \]

Similarly,
\[ \zeta_{0,0,\ldots,0,1,0,\ldots,0}^{(1)} = \frac{1}{3} \quad r = 1,2,\ldots,K \]

(1 at any place except r)

\[ \zeta_{1,0,\ldots,0}^{(1,2)} = \mathbb{E}\left[ h^{(1)}(Y_{1j_1}, Y_{2j_2}, \ldots, Y_{kj_k}) \cdot h^{(2)}(Y_{1j_1}, Y_{2j_2}, \ldots, Y_{kj_k}) \right] \]
\[ - \mathbb{E}(h^{(1)}(\cdot)) \cdot \mathbb{E}(h^{(2)}(\cdot)) \]

\[ = \mathbb{E}\left[ \sum_{i=2}^{k} (Y_{1j_1} - Y_{1j_1}) \sum_{c} \sum_{c'} \frac{1}{\hat{c}} I\{X_{1j_1} = X_{1j_1} = V_c\} \cdot \right. \]
\[ \cdot \left. \sum_{i''=1}^{k} (Y_{1j_1} - Y_{2j_2}) \sum_{c''} \frac{1}{\hat{c}''} I\{X_{2j_2} = X_{2j_2} = V_c''\} \right] \]
\[ = \mathbb{E}\left[ \sum_{i=2}^{k} \sum_{i''=1}^{k} (Y_{1j_1} - Y_{1j_1})(Y_{1j_1} - Y_{2j_2}) \times \right. \]
\[ \cdot \left. \sum_{c} \sum_{c'} \frac{1}{\hat{c}} \frac{1}{\hat{c}'} I\{X_{1j_1} = X_{1j_1} = V_c\} I\{X_{1j_1} = X_{2j_2} = V_c'\} \right] \]
\[
E \left[ \sum_{i' \geq 2} (Y_{1^i'j_{i'}} - Y_{1j_1})(Y_{1j_1} - Y_{2j_2}) \times \sum_{c} \epsilon_c^{-1} \epsilon_{c'}^{-1} I\{X_{1j_1} = X_{i'j_{i'}}, = \mu_c\} I\{X_{1j_1} = X_{2j_2}, = \mu_c\} \right] \\
+ \sum_{i' = 2}^{k} \sum_{i'' = 3}^{k} (Y_{i^i'j_{i'}} - Y_{1j_1})(Y_{i''j_{i''}} - Y_{2j_2}) \times \sum_{c} \epsilon_c^{-1} \epsilon_{c'}^{-1} I\{X_{1j_1} = X_{i'j_{i'}}, = \mu_c\} \cdot I\{X_{i''j_{i''}} = X_{2j_2}, = \mu_c\} \\
\cdot I\{X_{i''j_{i''}} = X_{2j_2}, = \mu_c\} \right]
\]

\[
(1,2) \sum_{i_0, \ldots, 0} \sum_{c} \epsilon_c^{-2} \epsilon_{c'}^{-3} \sum_{i' = 2}^{k} \left\{ \Pr(Y_{2j_2} < Y_{1j_1} \leq Y_{i'j_{i'}}, \left| X_{1j_1} = X_{2j_2}, = \mu_c\right) \right\} \\
+ \Pr(Y_{i''j_{i''}} < Y_{1j_1} < Y_{2j_2}, \left| X_{1j_1} = X_{2j_2}, = \mu_c\right) \\
- \Pr(Y_{1j_1} < Y_{i''j_{i''}} < Y_{2j_2}, \left| X_{1j_1} = X_{2j_2}, = \mu_c\right) \\
- \Pr(Y_{1j_1} < Y_{2j_2} < Y_{i''j_{i''}}, \left| X_{1j_1} = X_{2j_2}, = \mu_c\right) \\
- \Pr(Y_{i''j_{i''}} < Y_{2j_2} < Y_{1j_1}, \left| X_{1j_1} = X_{2j_2}, = \mu_c\right) \\
- \Pr(Y_{2j_2} < Y_{i''j_{i''}} < Y_{1j_1}, \left| X_{1j_1} = X_{2j_2}, = \mu_c\right) \right\} \\
+ \sum_{c} \epsilon_c^{-2} \epsilon_{c'}^{-4} \sum_{i' = 2}^{k} \sum_{i'' = 3}^{k} \cdot 0 \\
= \left( \sum_{c} \epsilon_c \right) \sum_{i' = 2}^{k} \left( \frac{2}{6} - \frac{4}{6} \right)
\]
\[
\zeta_{1,0,\ldots,0} = -\frac{1}{3} (k-1)
\]

Also, \(\zeta_{0,1,0,\ldots,0} = -\frac{1}{3} (k-1)\)

Similarly,

\[
\zeta_{0,0,\ldots,0,1,0,\ldots,0} = -\frac{1}{3} (k-1) \quad \text{for } r,s = 1,2,\ldots,k, \quad r \neq s
\]

\((1 \text{ in place } r \text{ or place } s)\)

\[
\zeta^{(1,2)}_{0,0,1,0,\ldots,0} = \mathbb{E} \left[ h^{(1)}(Y_{1j_1}, Y_{2j_2}, Y_{3j_3}, \ldots, Y_{kj_k}) \right] - \mathbb{E} \left[ h^{(1)}(\cdot) \right] \mathbb{E} \left[ h^{(2)}(\cdot) \right]
\]

\[
= \mathbb{E} \left[ \sum_{i=2}^{k} \text{sgn}(Y_{i} - Y_{1j_1}) \sum_{c} \varepsilon_c^{-1} \mathbb{I}(X_{1} = V_c) \right] \mathbb{I}(X_{i} = V_c) - \mathbb{E} \left[ h^{(1)}(\cdot) \right] \mathbb{E} \left[ h^{(2)}(\cdot) \right]
\]

\[
= \mathbb{E} \left[ \sum_{i'=2}^{k} \sum_{i''=1}^{k} \text{sgn}(Y_{i'} Y_{i''} - Y_{1j_1} Y_{2j_2}) \sum_{c} \varepsilon_c^{-1} \mathbb{I}(X_{i'} = V_c) \right] \mathbb{I}(X_{i} = V_c) - \mathbb{E} \left[ h^{(1)}(\cdot) \right] \mathbb{E} \left[ h^{(2)}(\cdot) \right]
\]

\[
= \mathbb{E} \left[ \sum_{i'=2}^{k} \sum_{i''=1}^{k} \sum_{c} \varepsilon_c^{-1} \varepsilon_c^{-1} \mathbb{I}(X_{i'} = V_c) \mathbb{I}(X_{i''} = V_c) \right]
\]

\[
\mathbb{I}(X_{1} = V_c) \mathbb{I}(X_{i} = V_c) \mathbb{I}(X_{i'} = V_c) \mathbb{I}(X_{i''} = V_c)
\]
\[
\begin{align*}
&= \mathbb{E} \left[ \text{sgn}(Y_{3j_3} - Y_{1j_1}) \cdot \text{sgn}(Y_{3j_3} - Y_{2j_2}) \cdot \sum_{c} \sum_{c'} \varepsilon_c^{-1} \varepsilon_{c'}^{-1} \cdot \\
&\quad \cdot I\{X_{1j_1} = X_{3j_3} = V_c\} \cdot I\{X_{2j_2} = X_{3j_3} = V_{c'}\} \right] \\
&\quad + \frac{1}{k} \sum_{i''=1}^{k} \sum_{i'=2}^{i''-1} \text{sgn}(Y_{i''j_{i''}} - Y_{1j_1}) \cdot \text{sgn}(Y_{i''j_{i''}} - Y_{2j_2}) \cdot \sum_{c} \sum_{c'} \varepsilon_c^{-1} \varepsilon_{c'}^{-1} \cdot \\
&\quad \cdot I\{X_{i''j_{i''}} = X_{1j_1} = V_c\} \cdot I\{X_{i''j_{i''}} = X_{2j_2} = V_{c'}\} \right]\end{align*}
\]

\[
\zeta_{0,0,1,0,\ldots,0} = \sum_{c} \varepsilon_c^{-2} \varepsilon_{c'}^{-3} \cdot \left[ \text{Pr}(Y_{1j_1} < Y_{2j_2} < Y_{3j_3} \mid X_{1j_1} = X_{2j_2} = X_{3j_3} = V_c) \right. \\
\left. + \text{Pr}(Y_{2j_2} < Y_{1j_1} < Y_{3j_3} \mid X_{1j_1} = X_{2j_2} = X_{3j_3} = V_c) \right] \\
\left. + \text{Pr}(Y_{3j_3} < Y_{1j_1} < Y_{2j_2} \mid X_{1j_1} = X_{2j_2} = X_{3j_3} = V_c) \right]
\]

\[
+ \sum_{c} \varepsilon_c^{-2} \varepsilon_{c'}^{-4} \cdot \sum_{i''=1}^{k} \sum_{i'=2}^{i''-1} \cdot 0 \\
\left( \text{not}(i''=i') \right) \\
\left( \text{not}(i''=3) \right)
\]

\[
\zeta_{0,0,1,0,\ldots,0} = \frac{1}{3}
\]
Similarly,

\[ \xi_0, \ldots, 0, 1, 0, \ldots, 0 = \frac{1}{3}, \quad r, s = 1, 2, \ldots, K, \quad r \neq s. \]

(1 in any place except \( r \) and \( s \))

\[ \sigma_{rr} = \sum_{i=1}^{k} a_i^{-1} \xi_0, 0, \ldots, 1, 0, \ldots, 0 \quad (r) \]

\[ \sigma_{rr} = a_r^{-1} \xi_0, 0, \ldots, 1, \ldots, 0 \quad (r) + \frac{1}{3} \sum_{i \neq r} a_i^{-1} \xi_0, 0, \ldots, 0, 1, 0, \ldots, 0 \quad (1 \text{ at } r \text{th place}) \]

\[ \sigma_{rr} = a_r^{-1} \left( \frac{1}{3} (K-1)^2 \right) + \frac{1}{3} \sum_{i=1}^{k} a_i^{-1} \frac{1}{3} \left[ \sum_{i=1}^{k} a_i^{-1} + (K^2 - 2K) a_r^{-1} \right] \]

\[ \sigma_{rs} = a_r^{-1} \xi_0, 0, \ldots, 0, 1, 0, \ldots, 0 \quad (r, s) \]

\[ \sigma_{rs} = a_r^{-1} \xi_0, 0, \ldots, 0, 1, 0, \ldots, 0 \quad (1 \text{ at } r \text{th place}) \]

\[ + \frac{1}{3} \sum_{i \neq 1} a_i^{-1} \xi_0, 0, \ldots, 0, 1, 0, \ldots, 0 \quad (1 \text{ at } \text{ith place}) \]

\[ \sigma_{rs} = a_r^{-1} + a_s^{-1} \left( - \frac{1}{3} (K-1) \right) + \frac{1}{3} \sum_{i \neq r, s} a_i^{-1} \]

\[ = \frac{1}{3} \left[ \sum_{i=1}^{k} a_i^{-1} - K(a_r^{-1} + a_s^{-1}) \right] \]

**Corollary 2.2:** Let \( T_n \) be the vector of \( K \)-sample U-statistics defined above. Then the asymptotic distribution of \( \sqrt{n} T_n \) has, under the null hypothesis, a limiting normal distribution with zero mean vector and
covariance matrix $\Sigma$, given in the previous theorem.

**Proof:** The result follows directly from Lemma 2.1.

Let $\bar{T} = \sum_{i=1}^{k} T_i / K$. Conditional on $\bar{T}$, the $T_i$'s are subject to a single linear constraint, and their distribution is singular, so their asymptotic distribution is singular, and $\Sigma$ is singular. In fact, $\text{Rank}(\Sigma) = K-1$.

Note that,

$$
\sigma_{rr} + \sum_{s=1}^{K} \sigma_{rs} = \frac{1}{3} \left[ \sum_{i=1}^{K} \sigma_{ii}^{-1} + (K^2 - 2K) \sigma_{rr}^{-1} \right] + \sum_{s=1}^{K} \left[ \frac{1}{3} \left( \sum_{i=1}^{K} \sigma_{ii}^{-1} - K(\sigma_{rr}^{-1} + \sigma_{ss}^{-1}) \right) \right]
$$

$$
= \frac{1}{3} \left\{ \sum_{i=1}^{K} \sigma_{ii}^{-1} + K^2 \sigma_{rr}^{-1} - 2K \sigma_{rr}^{-1} + (K-1) \sum_{i=1}^{K} \sigma_{ii}^{-1} - (K-1)K \sigma_{rr}^{-1} \right\}
$$

$$
= 0.
$$

Hence, clearly $\Sigma \Sigma = 0$.

Consequently, any $K-1$ of the $T_i$'s may be selected as the non-singular set. Therefore, let

$$
\mathcal{L}_0 = \sqrt{n} (T_1, T_2, \ldots, T_{K-1})
$$

and let $\Sigma_0$ be the corresponding covariance matrix.
Lemma 2.2: If the vector \( \mathbf{x} \) has a normal distribution with mean vector \( \mu \) and nonsingular variance-covariance matrix \( \Lambda \), then the quadratic form \( \mathbf{x}' \Lambda^{-1} \mathbf{x} \) has a \( \chi^2_r(\lambda) \) distribution, where \( r \) is the rank of \( \Lambda \) and \( \lambda = \mathbf{m}' \Lambda^{-1} \mathbf{m} \).


Corollary 2.3: Let \( \mathbf{z}_0 = \sqrt{n} \left( T_1, T_2, \ldots, T_{K-1} \right) \) and \( \Sigma_0 = \left( (\sigma_{rs}) \right) \), \( r, s = 1, 2, \ldots, K-1 \), \( \neq s \) be the corresponding covariance matrix for \( \sqrt{n} \mathbf{z}_0 \). Then
\[
\mathbf{z}_0' \Sigma_0^{-1} \mathbf{z}_0
\]
has, under the null hypothesis, asymptotically a chi-squared distribution with \( K-1 \) degrees of freedom.

Proof: The result follows from Corollary 2.2 and Lemma 2.2.

Theorem 2.4: Let \( \mathbf{z}_0 = \sqrt{n} \left( T_1 - \bar{T}, T_2 - \bar{T}, \ldots, T_{K-1} - \bar{T} \right) \). Then, conditional on \( \bar{T} \),
\[
\mathbf{z}_0' \Sigma_0^{-1} \mathbf{z}_0
\]
has, under \( H_0 \), asymptotically a chi-squared distribution with \( K-1 \) degrees of freedom.

Proof: \( E[ E[ \mathbf{z}_0' | \bar{T} ]] = E[ \mathbf{z}_0' - \bar{T} ] = 0. \)

The covariance matrix is unaltered by adding \( \bar{T} \) to each component of \( \mathbf{z}_0 \).

It has been shown that conditional on \( \bar{T} \), \( \mathbf{z}_0' \Sigma_0^{-1} \mathbf{z}_0 \) is asymptotically chi-squared with \( K-1 \) degrees of freedom under \( H_0 \). It can be
easily shown that, since the distribution is chi-squared (i.e., depends on the degrees of freedom and nothing more), it does not depend on the conditioning factor, and is, hence, the same as the unconditional distribution.

**Corollary 2.4:** Let \( t'_{\xi_0} = \sqrt{\bar{n}} (T_1 - T, T_2 - T, \ldots, T_{K-1} - T) \), and \( \Sigma_0 \) be as defined above. Then

\[
D^* = t'_{\xi_0} \Sigma^{-1}_{\xi_0} t_{\xi_0}
\]

has asymptotically a null distribution which is \( \chi^2(K-1) \).

**Proof:**

Conditional on \( \bar{T} \), \( D^* \) is asymptotically \( \chi^2(K-1) \), by the previous theorem. Hence,

\[
\text{the characteristic function of } D^* \sim (1 - 2it)^{-\frac{(K-1)}{2}}.
\]

So

\[
E[e^{itD^*} | \bar{T}] \sim (1 - 2it)^{-\frac{(K-1)}{2}}
\]

and

\[
E[E[e^{itD^*} | \bar{T}]] \sim (1 - 2it)^{-\frac{(K-1)}{2}},
\]

and the result follows.

It has just been demonstrated that \( D^* = t'_{\xi_0} \Sigma^{-1}_{\xi_0} t_{\xi_0} \) has asymptotically a null distribution which is chi-squared with \( K-1 \) degrees of freedom. To express \( D^* = t'_{\xi_0} \Sigma^{-1}_{\xi_0} t_{\xi_0} \), the AMP sign statistic for \( K \)-samples, explicitly, it is necessary to invert \( \Sigma_0 \). This will be considered in a later section. First, the limiting distribution for \( D^* \) under certain types of alternatives is considered.
E. The Limiting Distribution of \( D^* \) Under Translation-Type Alternatives

In this section, the asymptotic distribution of \( D^* \) is investigated, assuming a sequence of translation-type alternative hypotheses, \( H_{n_i} \), for \( n = 1, 2, \ldots \). The hypothesis \( H_{n_i} \) specifies that

\[
F_i(y|x) = F(y - n^{-1} \theta_i - \theta_i), \quad i=1, 2, \ldots, K,
\]

where \( \theta_i \neq \theta_i' \), for some \( i \neq i' \). The limiting distribution is then found as \( n \to \infty \).

**Theorem 2.5:** For each index \( n \), assume \( \lim_{n \to \infty} \frac{n_i}{n} = a_i \) exists and \( 0 < a_i < 1 \), and the truth of \( H_{n_i} \). If \( F \) possesses a continuous derivation \( f \), and there exists a function \( g \) such that

\[
\left| \frac{[f(y+h|x) - f(y|x)]}{h} \right| \leq g(y|x)
\]

and

\[
\int_{-\infty}^{\infty} g(y|x) f(y|x) \, dy < \infty,
\]

then, for \( n \to \infty \), the statistic \( D^* \) has a limiting noncentral \( \chi^2 \) distribution with \( K-1 \) degrees of freedom, and noncentrality parameter

\[
\lambda_{D^*} = \left( \sum a_i \right) \cdot 4 \cdot \delta' \delta^{-1} \delta \left[ \int_{-\infty}^{\infty} F'(y|x) \, dF(y|x) \right]^2.
\]

(Note: Although \( \lambda_{D^*} \) appears to depend on \( x \), by Assumption A.4, it is invariant to the choice of any particular value of \( x \), and hence does not depend on \( x \).)
Proof:

Let \( \gamma_n^{(i)} = E[T_i(\cdot)|H_n] \)

\[
= E\left\{ \sum_{i=1}^{k} \text{sgn}(Y_i, j_i - Y_{ij_1}, j_1) \sum_{c} \epsilon_c^{-1} I\{X_{i}^{(i)}, j_i = X_{ij_1}^{(i)}, j_1 = Y_{c}\} \right\}
\]

\[
\gamma_n^{(i)} = \sum_{c} \epsilon_c^{-1} \sum_{i', \neq i}^{2} \left\{ \text{pr}(Y_i, j_i < Y_{ij_1}, j_1, X_{i}^{(i)}, j_i = X_{ij_1}^{(i)}, j_1) - \text{pr}(Y_i, j_i > Y_{ij_1}, j_1, X_{i}^{(i)}, j_i = X_{ij_1}^{(i)}, j_1) \right\}
\]

\[
\gamma_n^{(i)} = \sum_{i' = 1}^{k} \left\{ \text{pr}(Y_i, j_i < Y_{ij_1}, j_1, X_{i}^{(i)}, j_i = X_{ij_1}^{(i)}, j_1) - \text{pr}(Y_i, j_i > Y_{ij_1}, j_1, X_{i}^{(i)}, j_i = X_{ij_1}^{(i)}, j_1) \right\}
\]

\[
= 2 \left\{ \sum_{i' = 1}^{k} \text{pr}(Y_i, j_i < Y_{ij_1}, j_1, X_{i}^{(i)}, j_i = X_{ij_1}^{(i)}, j_1) \right\} - (K-1)
\]

\[
\psi_i = 2 \left\{ \text{pr}(Y_1, j_1 < Y_{ij_1}, j_1, X_{1}^{(i)}, j_1 = X_{ij_1}^{(i)}, j_1) + \text{pr}(Y_2, j_2 < Y_{ij_1}, j_1, X_{2}^{(i)}, j_1 = X_{ij_1}^{(i)}, j_1) + \ldots + \text{pr}(Y_{Kj_k}, j_k < Y_{ij_1}, j_1, X_{Kj_k}^{(i)}, j_1 = X_{ij_1}^{(i)}, j_1) \right\} - (K-1)
\]
By Assumption A.4, \( \Pr(Y_{ij} < Y_{i',j'} \mid X_{ij} = X_{i',j'}) \) is the same for all \( X \), hence, does not depend on \( X \). Choose \( X = \bar{x} \), arbitrary but fixed.

\[
\psi_i = 2 \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_1(y_{\bar{x}}) f_i(y_{\bar{x}}) \, dy_{\bar{x}} \, dy_i \right\} + \\
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_2(y_{\bar{x}}) f_i(y_{\bar{x}}) \, dy_{\bar{x}} \, dy_i + \ldots + \\
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_k(y_{\bar{x}}) f_i(y_{\bar{x}}) \, dy_{\bar{x}} \, dy_i \right\} - (K-1)
\]

\[
= 2 \left\{ \int_{-\infty}^{\infty} F_1(y_{\bar{x}} \mid x) f_i(y_{\bar{x}} \mid x) \, dy_i \right\} + \int_{-\infty}^{\infty} F_2(y_{\bar{x}} \mid x) f_i(y_{\bar{x}} \mid x) \, dy_i + \ldots + \\
\int_{-\infty}^{\infty} F_k(y_{\bar{x}} \mid x) f_i(y_{\bar{x}} \mid x) \, dy_i \right\} - (K-1).
\]

Under \( H_n \),

\[
F_i(y_{\bar{x}} \mid x) = F(y - \theta_i / \sqrt{n} \mid x).
\]

Thus,

\[
\psi_i = 2 \left\{ \int_{-\infty}^{\infty} F(y_{\bar{x}} - \theta_i / \sqrt{n} \mid x) \, dF(y_{\bar{x}}) + \int_{-\infty}^{\infty} F(y_{\bar{x}} - \theta_2 / \sqrt{n} \mid x) \, dF(y_{\bar{x}}) \\
+ \ldots + \int_{-\infty}^{\infty} F(y_{\bar{x}} - \theta_k / \sqrt{n} \mid x) \, dF(y_{\bar{x}}) \right\} - (K-1).
\]

Change of variables: Let \( y = y_{\bar{x}} - \theta_i / \sqrt{n} \)

\[
dy = dy_{\bar{x}}
\]
\[ \psi_i = 2 \cdot \left\{ \int_{-\infty}^{\infty} F(y + \theta_i/\sqrt{n} - \theta_\infty/\sqrt{n} \mid x) \, dF(y \mid x) + \int_{-\infty}^{\infty} F(y + \theta_i/\sqrt{n} - \theta_2/\sqrt{n} \mid x) \, dF(y \mid x) + \ldots + \int_{-\infty}^{\infty} F(y + \theta_i/\sqrt{n} - \theta_k/\sqrt{n} \mid x) \, dF(y \mid x) \right\} - (K-1). \]

Write:

\[ \psi_i(t) = 2 \cdot \left\{ \int_{-\infty}^{\infty} F(y + (\theta_i-\theta_\infty) \cdot t \mid x) \, dF(y \mid x) + \int_{-\infty}^{\infty} F(y + (\theta_i-\theta_2) \cdot t \mid x) \, dF(y \mid x) + \ldots + \int_{-\infty}^{\infty} F(y + (\theta_i-\theta_k) \cdot t \mid x) \, dF(y \mid x) \right\} - (K-1). \]

Expand \( \psi_i(t) \) in a Maclaurin series in \( t \). The constant term is clearly 0 (since this corresponds to \( \psi_i^{(0)} = \int_{-\infty}^{\infty} F(y \mid x) \, dF(y \mid x) + \ldots + \int_{-\infty}^{\infty} F(y \mid x) \, dF(y \mid x) \) and \( F_\infty(y \mid x) = F(y \mid x)^{\infty}_- \) for all \( i \), \( \Pr[Y_i < Y_i \mid x] = \frac{1}{2} \).

For the linear term:

\[ \frac{1}{2} \frac{d}{dt} \psi_i(t) = \int_{-\infty}^{\infty} \frac{d}{dt} F(y + (\theta_i-\theta_\infty) \cdot t \mid x) \, dF(y \mid x) + \int_{-\infty}^{\infty} \frac{d}{dt} F(y + (\theta_i-\theta_2) \cdot t \mid x) \, dF(y \mid x) + \ldots + \int_{-\infty}^{\infty} \frac{d}{dt} F(y + (\theta_i-\theta_k) \cdot t \mid x) \, dF(y \mid x) + 0. \]

(The above derivative exists, and "differentiation under the integral sign" is permitted by Assumptions 2.5.1 and 2.5.2.)
\[ \frac{1}{2} \frac{d}{dt} \psi_i(t) = \int_{-\infty}^{\infty} F'(y + (\theta_i - \theta_1) t | x) \psi_i(t | x) \ dF(y | x) \]
\[ + \int_{-\infty}^{\infty} F'(y + (\theta_i - \theta_2) t | x) (\theta_i - \theta_2) \ dF(y | x) + \ldots + \]
\[ + \int_{-\infty}^{\infty} F'(y + (\theta_i - \theta_k) t | x) (\theta_i - \theta_k) \ dF(y | x) \]

and
\[ \frac{1}{2} \frac{d}{dt} \psi_i(t) \bigg|_{t=0} = \sum_{i'=1}^{k} (\theta_i - \theta_{i'}) \int_{-\infty}^{\infty} F'(y | x) \ dF(y | x). \]

Let
\[ \lambda = \int_{-\infty}^{\infty} F'(y | x) \ dF(y | x) \]
and
\[ \delta_i = \sum_{i'=1}^{k} (\theta_i - \theta_{i'}). \]

Then
\[ \psi_i(t) = 2\delta_i \lambda t + O(t). \]

But
\[ \psi_i \left( \frac{1}{\sqrt{n}} \right) = \psi_i \]
\[ = E[T_i(\cdot) | H_n]; \]

thus, under \( H_n \)
\[ E[T_i] = \frac{2 \delta_i \lambda}{\sqrt{n}} + O \left( \frac{1}{\sqrt{n}} \right). \]

By a similar argument
\[ \sum_{\sim m} = \sum_{\sim} + O(n^{-3/2}) \]
where \( \sum_{\sim m} \) is the covariance matrix of \( \sqrt{n} (T - \gamma_n) \) as \( n \to \infty \), and \( O(n^{-3/2}) \) is a matrix whose elements are \( O(n^{-3/2}) \). It follows from Lemma 2.2 that,
under hypothesis $H_n$,

$$ \sqrt{n} \left( \overline{T} - \overline{\gamma}_n \right) $$

for large $n$

has zero mean vector and covariance matrix $\Sigma$ (asymptotically), and

is asymptotically normally distributed. Consequently,

$$ \sqrt{n} \left( \overline{T} - \overline{T}_n \right) $$

has a limiting normal distribution with mean vector

$$ \sqrt{n} \left[ \frac{2}{\sqrt{n}} \delta \lambda \right] $$

where $\delta' = (\delta_1, \delta_2, \ldots, \delta_k)$. The mean vector is hence

$$ 2 \lambda \delta \lambda $$

Note that $\sum_i \delta_i = 0$. And let $\delta_0' = (\delta_1, \delta_2, \ldots, \delta_{k-1})$. Then from Lemma 2.2, $\sqrt{n} \left( \overline{T}_n - \overline{T}_n \right)$ under hypothesis $H_n$ is asymptotically distributed as a chi-square with $K-1$ degrees of freedom, and noncentrality parameter

$$ \lambda_{D^*} = 4 \lambda^2 \delta_0' \Sigma_0^{-1} \delta_0 $$

$\lambda_{D^*}$ may be simplified if $\Sigma_0^{-1}$ is explicitly derived.

F. The AMP Sign Statistic for K-samples

A Special Case: $n_1=n_2=\ldots=n_k$

In this section, the AMP sign statistic for K-samples is presented for the special case when the sample sizes are equal. The AMP sign statistic for K-samples is given by

$$ D^* = \overline{t}_0' \Sigma_0^{-1} \overline{t}_0 $$
Note that $n_1=n_2=\ldots=n_K$ only if $a_1=a_2=\ldots=a_K=\frac{1}{K}$. The following lemma will be useful:

**Lemma 2.3:** Let the $m\times m$ matrix $C$ be defined by

$$ C = (q-t) I + tJ. $$

Then $C$ has an inverse if and only if $q\neq t$ and $q\neq -(m-1)t$. If $C^{-1}$ exists, then it is given by

$$ C^{-1} = \frac{1}{q-t} \left[ I - \frac{t}{q + (m-1)t} J \right]. $$

**Proof:** See Graybill (1967, pp. 173, 174).

1. $\sum_{i=1}^{K} a_i^{-1}$ is derived.

Recall, $\sum_{i=1}^{K} a_i^{-1} = (\sigma_{rs})$, where

$$
\sigma_{rr} = \frac{1}{3} \left[ \sum_{i=1}^{K} a_i^{-1} + (K^2-2K) a_r^{-1} \right], \quad r=1,2,\ldots,K-1
$$

$$
\sigma_{rs} = \frac{1}{3} \left[ \sum_{i=1}^{K} a_i^{-1} - K(a_r^{-1} + a_s^{-1}) \right], \quad r,s=1,2,\ldots,K-1; \quad r \neq s
$$

So, in this case, $a_i^{-1} = K$ for all $i$.

$$
\sigma_{rr} = \frac{1}{3} \left[ \sum_{i=1}^{K} K + (K^2-2K)K \right]
$$

$$
= \frac{1}{3} [K^2 + K^3 - 2K^2]
$$

$$
\sigma_{rr} = \frac{1}{3} K^2(K-1), \quad r=1,2,\ldots,K-1
$$

and
\[ \sigma_{rs} = \frac{1}{3} \left[ \sum_{i=1}^{k} K - K(K^2) \right] \]
\[ = \frac{1}{3} [K^2 - 2K^2] \]
\[ \sigma_{rs} = -\frac{1}{3} K^2 , \quad r=1,2,\ldots,K-1 \]

Hence, \[ \sum_{o} = \frac{1}{3} K^2 \left[ KI_{K-1} - \frac{1}{2}J_{K-1} \right] . \]

It follows from Lemma 2.3 that
\[ \sum_{o}^{-1} = \frac{1}{3} K^2 \left[ \frac{1}{K} I_{K-1} - \frac{1}{(K-1)} \frac{1}{2} J_{K-1} \right] \]
\[ \sum_{o}^{-1} = \frac{3}{K^3} \left[ \frac{1}{K} I_{K-1} + \frac{1}{2} J_{K-1} \right] \]

2. Statistic \( D^* \) is expressed explicitly.

Expanding \( t_o^t \sum_o^{-1} t_o \), where
\[ t_o = \sqrt{n} (T_1 - \bar{T}, T_2 - \bar{T}, \ldots, T_{K-1} - \bar{T}) \]
yields
\[ t_o^t \sum_o^{-1} t_o = \frac{3n}{K^3} \left\{ \sum_{i=1}^{k-1} (T_i - \bar{T})^2 + \left[ \sum_{i=1}^{k-1} (T_i - \bar{T}) \right]^2 \right\} \]
\[ = \frac{3n}{K^3} \left\{ \sum_{i=1}^{k-1} (T_i - \bar{T})^2 + \left[ \sum_{i=1}^{k-1} (T_i - \bar{T}) \right]^2 \right\} \]
\[ = \frac{3n}{K^3} \left\{ \sum_{i=1}^{k-1} (T_i - \bar{T})^2 + (-T_{K-1} + \bar{T})^2 \right\} \]
\[ t_o^t \sum_o^{-1} t_o = \frac{3n}{K^3} \sum_{i=1}^{k} (T_i - \bar{T})^2 . \]
3. The limiting distribution of $D^*$ under the null hypothesis.

The following result has been established.

**Corollary 2.5:** Let $D^*$ be the AMP sign statistic for $K$-samples with

$n_1 = n_2 = \ldots = n_k$. Then,

$$D^* = \frac{3n}{K^3} \sum_{i=1}^{k} (T_i - \bar{T})^2.$$

Under the hypothesis of [conditional] homogeneity, $D^*$ is asymptotically
chi-squared with $K-1$ degrees of freedom.

4. The limiting distribution of $D^*$ under translation-type alternatives.

As in Section E, assume a sequence of translation-type alternative
hypotheses, $H_n$, for $n=1,2,\ldots$, where

$$H_n: F_i(y|x) = F(y-\theta_i/\sqrt{n} | x), \quad i=1,2,\ldots,K$$

**Corollary 2.6:** Assume the truth of $H_n$. If $F$ possesses a continuous

derivative $F'$, and there exists a function $g$ such that

$$\left| \frac{F'(y+h|x) - F'(y|x)}{h} \right| \leq g(y|x)$$

and

$$\int_{-\infty}^{\infty} g(y|x) \, dF(y|x) < \infty,$$

then for $n \to \infty$, the statistic $D^*$ has a limiting noncentral chi-squared
distribution with $K-1$ degrees of freedom, and noncentrality parameter

$$\lambda_{D^*} = 12 \cdot \frac{1}{K} \sum_{i=1}^{k} (\theta_i - \bar{\theta})^2 \left\{ \int_{-\infty}^{\infty} F'(y|x) \, dF(y|x) \right\}^2$$

Proof:

The result follows directly from Theorem 4.2, simplifying the noncentrality
parameter.

\[
\delta_0' \delta_0^{-1} \delta_0 = \frac{3}{K^3} \left[ \sum_{i=1}^{K-1} \delta_i^2 + \left[ \sum_{i=1}^{K-1} \delta_i \right]^2 \right] \\
= \frac{3}{K^3} \left[ \sum_{i=1}^{K-1} \delta_i^2 + [-\delta_K]^2 \right], \text{ since } \sum \delta_i = 0 \\
= \frac{3}{K^3} \sum_{i=1}^{K} \delta_i^2.
\]

Let \( \bar{\theta} = \frac{1}{K} \sum_{i=1}^{K} \theta_i \).

Then \( \sum_{i=1}^{K} \delta_i^2 = \sum_{i=1}^{K} \left( K \theta_i - \sum_{i=1}^{K} \theta_i \right)^2 = K^2 \sum_{i=1}^{K} (\theta_i - \bar{\theta})^2 \).

Hence

\[
\lambda_{D^*} = 4 \cdot \lambda^2 \left\{ \frac{3}{K} \cdot K^2 \sum_{i=1}^{K} (\theta_i - \bar{\theta})^2 \right\} \\
= 4 \cdot \lambda^2 \left\{ \frac{3}{K} \sum_{i=1}^{K} (\theta_i - \bar{\theta})^2 \right\} \\
\lambda_{D^*} = 12 \cdot \frac{1}{K} \cdot \sum_{i=1}^{K} (\theta_i - \bar{\theta})^2 \left\{ \int_{-\infty}^{\infty} F'(y|x) \, dF(y|x) \right\}
\]

5. The AMP sign statistic with estimated weights.

Assuming that the \( \varepsilon_c \)'s are known, the AMP analysis for K-samples

is based on

\[
T_i = (\Pi_{i=1}^{n_i})^{-1} \left[ \sum_{j_1=1}^{n_1} \ldots \sum_{j_k=1}^{n_k} \left\{ \sum_{i'=1}^{k} \text{sgn}(Y_{i',j_i}, -Y_{i,j_i}) \sum_{c=1}^{\varepsilon_c} I\{X_{i,j_i} = X_{i',j_i} = 0\} \right\} \right]
\]

Let \( w_c = \varepsilon_c^{-1} \). Then \( w_c \)'s are weights. If the \( \varepsilon_c \)'s are unknown, then \( w_c \)'s

are estimated by \( \hat{w}_c = \hat{\varepsilon}_c^{-1} \) where \( \hat{\varepsilon}_c = m_c/n \), and
\begin{align*}
m_c &= \sum_{i=1}^{k} n_i c = \sum_{i=1}^{k} \sum_{j=1}^{n_i} I\{X_{i,j} = V_c\}.
\end{align*}

Note
\begin{align*}
T_i = \sum_{c=1}^{C} n_c \left( \prod_{i=1}^{n_i} \right)^{-1} \left\{ \sum_{j=1}^{n_i} \sum_{j=1}^{n_k} \left( \sum_{i=1}^{k} \text{sgn}(Y_i, j_i, -Y_{i,j}) I\{X_i, j_i = V_c\} \right) \right\}.
\end{align*}

The quantity in brackets is a K-sample U-statistic, say $T'_c$.

And
\begin{align*}
T_i = \sum_{c=1}^{C} w_c T'_c, i.
\end{align*}

Define
\begin{align*}
\tilde{T}_i = \sum_{c=1}^{C} \hat{w}_c T'_c, i.
\end{align*}

We may invoke Slutsky's theorem (1925) to show that the statistic $\tilde{T}_i$ has the same limiting power as $T_i$. Consider
\begin{align*}
J_o = \sum_{c=1}^{C} \varepsilon_c^{-1} T'_c, \sqrt{n}
\end{align*}

By the central limit theorem for U-statistics, $J_o$ is asymptotically normal with mean zero and finite variance, $\sigma^2$. The test based on $T_i$ rejects if $J_o$ exceeds $\sigma Z^*$, where $Z^*$ is the appropriate normal critical value. If $\hat{\varepsilon}_c$ is substituted for $\varepsilon_c$ in $J_o$,
\begin{align*}
\tilde{J}_o = \sqrt{n} \cdot \tilde{T}_i,
\end{align*}

and the test based on $\tilde{T}_i$ rejects if $\tilde{J}_o$ exceeds $\sigma Z^*$. Now consider
\begin{align*}
\tilde{J}_o - J_o &= \sqrt{n} (\tilde{T}_i - T_i)
\end{align*}

\begin{align*}
= \sum_{c=1}^{C} \sqrt{n} \left( \frac{1}{\hat{\varepsilon}_c} - \frac{1}{\varepsilon_c} \right) \cdot T'_c, i.
\end{align*}
As $n \to \infty$, $C_{nc}$ is asymptotically normal with mean zero and variance $\epsilon_c(1-\epsilon_c)$. But as $n \to \infty$, the sequence of random variables $d_{nc}$ converges to $\psi_i$ under the fixed alternative or to $\frac{1}{K}$ under the contiguous alternative $\psi_i(n) = \frac{1}{K} + \sigma/\sqrt{n}$. Hence $\sum_{c=1}^{C} C_{nc} d_{nc}/\psi_i$ has the same limiting distribution as $\sum_{c=1}^{C} C_{nc}$ by Slutsky's theorem (1925). Note that $\sum_{c=1}^{C} C_{nc}$ is identically zero. Hence the tests based on $\tilde{T}_i$ and $\tilde{T}_j$ have the same limiting power.


For the special case where $n_1 = n_2 = \ldots = n_K = \frac{n}{K}$, if all pairs are matched (i.e., if $c=1$), it is easy to demonstrate that the AMP sign test is equivalent to the Kruskal-Wallis test.

Recall from Chapter I that a useful expression for the Kruskal-Wallis statistic is

$$H = \frac{1}{n(n+1)} \sum_{i=1}^{k} \frac{1}{n_i} \left( \sum_{t} U_{it} \right)^2$$

where

$$U_{it} = \sum_{j_t=1}^{n_i} \sum_{j_i=1}^{n_t} \text{sgn}(Y_{itj_t} - Y_{ij_i}).$$

Since $n_i = \frac{n}{K}$,

$$H = \frac{3K}{n^2(n+1)} \sum_{i=1}^{k} \left( \sum_{t} U_{it} \right)^2.$$
If \( C = 1 \), then \( \varepsilon = 1 \), and

\[
W_i = \sum_{j_1 = 1}^{n_1} \sum_{j_2 = 1}^{n_2} \ldots \sum_{j_k = 1}^{n_k} \sum_{i' = 1}^{k} \text{sgn}(Y_{i'j_1}, Y_{i'j_2}, \ldots, Y_{i'j_k})
\]

and

\[
T_i = (\Pi n_i)^{-1} W_i
\]

\[
T_i = \sum_i \frac{U_{ii'}}{n_i n_i'} = \sum_i \frac{U_{ii'}}{(n/k)^2} = \frac{k^2}{n^2} \sum_i U_{ii'}
\]

Clearly \( \sum T_i = 0 \) (for this case).

Hence, the AMP sign statistic is

\[
D^* = \frac{3n}{K^3} \sum_{i=1}^{k} (T_i)^2
\]

\[
= \frac{3n}{K^3} \sum_{i=1}^{k} \left( \frac{k^2}{n^2} \sum_i U_{ii'} \right)^2
\]

\[
= \frac{3K}{n^3} \sum_{i'=1}^{k} \sum_i U_{ii'}
\]

\[
\frac{n}{n+1} D^* = H, \text{ and they are asymptotically equal.}
\]

G. Chapter Summary and Concluding Remarks

In this chapter, a new method of matched analysis of covariance is advanced for the K-sample problem, called the All Matched Pairs Sign Test for K-samples. The test statistic is

\[
D^* = t_0' \Sigma^{-1} t_0,
\]
where
\[
 t_0 = \sqrt{n} \left( T - \overline{T} \right), \\
 T_i = W_i / \prod_{i} n_i \\
 W_i = \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \ldots \sum_{j_k=1}^{n_k} h(i) \left[ (x_{1j_1}, y_{1j_1}), (x_{2j_2}, y_{2j_2}), \ldots, (x_{kj_k}, y_{kj_k}) \right] \\
 h(i)[\cdot] = \sum_{i'=1}^{k} \text{sgn}(y_{i',j_i'} - y_{i'j_k}) \sum_{c=1}^{c-1} I \left\{ x_{ij_1} = x_{2j_2} = \ldots = x_{kj_k} \right\}, \\
 \overline{T} = K^{-1} \sum_{i'=1}^{k} T_i ,
\]
and \( \Sigma_{0} \) is the covariance matrix corresponding to \( t_0 \), and is given by

\[
 \left( \sigma_{rs} \right) \\
 \sigma_{rr} = \frac{1}{3} \left[ \sum_{i=1}^{k} a_i^{-1} + (k^2 - 2k) a_r^{-1} \right], \quad r = 1, 2, \ldots, K-1 \\
 \sigma_{rs} = \frac{1}{3} \left[ \sum_{i=1}^{k} a_i^{-1} - k(a_r^{-1} + a_s^{-1}) \right], \quad r, s = 1, 2, \ldots, K-1, \quad r \neq s
\]
where \( n_i = n a_i \).

It is demonstrated that \( D^* \) is asymptotically distributed as a chi-squared random variable under the null hypothesis and translation-type alternatives. The noncentrality parameter under a sequence of alternatives

\[
 H_n: F_i(y|\overline{\theta}) = F(y-n^{-1/2} \theta |\overline{\theta}), \quad i=1, 2, \ldots, K
\]

is given by

\[
 \lambda_{D^*} = 4 \delta_0 \sum_{0}^{\Sigma_0} \delta_0 \left[ \int_{-\infty}^{\infty} F'(y|\overline{\theta}) dF(y|\overline{\theta}) \right]^2
\]
where
\[ \delta'_0 = (\delta_1, \delta_2, \ldots, \delta_{K-1}), \]
for
\[ \delta'_i = \sum_{i' \neq i}^k (\theta_i - \theta_{i'}), \]

In the special case where the sample sizes are equal, \( D^* \) is simplified to
\[ D^* = \frac{3n}{K^2} \sum_{i=1}^k (T_i - \bar{T})^2, \]
and the noncentrality parameter is
\[ \lambda_{D^*} = 12 \cdot \frac{1}{K^2} \sum_{i' \neq i}^k (\theta_i - \bar{\theta})^2 \left\{ \int_{-\infty}^{\infty} F'(y|x) \, dF(y|x) \right\}^2. \]

The form of the test statistic for general sample sizes is not expressed explicitly due to difficulties involved in inverting a rather complicated covariance matrix. However, if numerical methods are permitted, given data, the test statistic and noncentrality parameter can be derived.

Although the demonstration and discussion are postponed until Chapter IV, it may be noted here that the ARE of the AMP sign test for \( K \)-samples and the Kruskal-Wallis test is unity, when \( Y \) is independent of \( X \).

Examples and efficiency results are reported in Chapter IV.
CHAPTER III

THE EXTENSION BASED ON BHAPKAR'S TEST

In this chapter a method of Analysis of Covariance (ANOCOVA) by matching is presented that is an extension of Schoenfelder's All Matched Pairs (AMP) (two-sample) sign test and Bhapkar's V-test (non-parametric analysis of variance). The test to be proposed, called the All Matched K-tuples (AMK) sign test, requires the analysis of matched K-tuples.

Definition 3.1: A matched k-tuple is a vector of responses \( (y_{1j_1}, y_{2j_2}, ..., y_{kj_k}) \) formed by selecting one observation from each sample such that the corresponding covariate vectors are matched.

A. The Basic Framework and Notation

The basic framework and notation for this chapter are the same as for Chapter II. The reader is referred back to Chapter II, Section A, for Assumptions A.1 - A.3.

B. Parameter Definitions

The parameters describing probabilities of certain rankings among observations must be defined conditionally and unconditionally.
Definition 3.2: Let $\phi_i$ denote the [unconditional] probability that of $K$ randomly selected observations, one from each sample, the observation from population $i$ is least on $Y$. (Note $\Sigma \phi_i = 1$.)

Definition 3.3: Let $\eta_i(V_c)$ denote the conditional probability that, of $K$ randomly selected observations, one from each sample, that are matched so that $X = V_c$ for each observation, the observation from population $i$ is least on $Y$.

Definition 3.4: Let $\psi_i$ denote the [matched] probability, that of $K$ randomly selected observations, one from each population, that are matched on $X$, the observation from population $i$ is least on $Y$.

Definition 3.5: Let $\beta_m$ denote the probability that $m > 1$ randomly selected observations match on $X$, as in Chapter II. In this chapter, for the special case when $m = K$, the subscript $K$ will be dropped: $\beta \equiv \beta_K$.

Proposition 3.1: Let $X$ be a discrete random variable with possible values $V_c$, $c=1,2,\ldots,C$ with $\varepsilon_c = \Pr(X = V_c)$, $\varepsilon_c = 0$ for all $C$, $\Sigma \varepsilon_c = 1$. Then for $m > 1$ independent observations on $X$, $\beta_m = \Sigma_{c}^{m} \varepsilon_c$.

Proof: Obvious by inspection.

Proposition 3.2: Let $\eta_i(V_c)$ and $\psi_i$ be as defined above. Then

$$\psi_i = \sum_{c}^{k} \varepsilon_{c_i} \eta_i(V_c) / \beta$$

Proof: Let $(X_{1j_1}, Y_{1j_1}), (X_{2j_2}, Y_{2j_2}), \ldots, (X_{kj_k}, Y_{kj_k})$ be sample observations. Then from Proposition 1,
\[ P(\text{match of the } K \text{ observations}) = \beta \]

Then note
\[
P(\text{match at } V_c | \text{match}) = P(X_{1j_1} = X_{2j_2} = \ldots = X_{kj_k} = V_c | \text{match})
\]
\[
= P(X_{1j_1} = X_{2j_2} = \ldots = X_{kj_k} = V_c) / P(X_{1j_1} = X_{2j_2} = \ldots = X_{kj_k})
\]
\[
= (\varepsilon_c)^k / \beta.
\]

Hence, it follows that
\[
\psi_i = \Pr\{y_{i}^{*} < y_i, \text{ for } i' = 1, 2, \ldots, k, i' \neq i \mid X_{1j_1} = X_{2j_2} = \ldots = X_{kj_k} \}
\]
\[
= \sum_{c} \Pr\{y_{i}^{*} < y_i, \text{ for } i' \neq i \mid X_{1j_1} = \ldots = X_{kj_k} = V_c \}
\]
\[
\cdot \Pr\{X_{1j_1} = \ldots = X_{kj_k} = V_c \mid X_{1j_1} = \ldots = X_{kj_k} \}
\]
\[
= \sum_{c} n_i(V_c) \varepsilon_c^k / \beta.
\]

C. Number of Analysis Units

The extension of Schoenfelder's AMP sign test presented in this chapter is an analysis of AMK, i.e.,

\[
\text{all } (X_{1j_1}, X_{2j_2}, \ldots, X_{kj_k}) \text{ where } X_{ij_i} = (X_{1j_1}^*, Y_{ij_i}^*), \text{ such that } X_{1j_1}^*, X_{2j_2}^*, \ldots, X_{kj_k}^* \text{ match. Consequently, the number of units of analysis (matched K-tuples) is a random variable, say}
\]
\[
A = \sum_{j_k = 1}^{n_k} \sum_{j_{k-1} = 1}^{n_{k-1}} \ldots \sum_{j_2 = 1}^{n_2} \sum_{j_1 = 1}^{n_1} M(X_{1j_1}^*, X_{2j_2}^*, \ldots, X_{k-1j_{(k-1)}^*}, X_{kj_k}^*)
\]
\[
(4.2)
\]

where
\[ M(X_{1j_1}, \ldots, X_{kj_k}) = \begin{cases} 1 & \text{if } X_{1j_1}, \ldots, X_{kj_k} \text{ match} \\ 0 & \text{else} \end{cases} \]

If all \( K \)-tuples of observations match, then obviously the number of analysis units is \( \prod_{i=1}^{k} n_i \). In general, however, this is not expected. Hence, it is desirable to derive \( \text{E}(A) \) and \( \text{Var}(A) \).

**Theorem 3.1:** Let \( A \) be the number of matched \( K \)-tuples from \( K \) independent random samples of sizes \( n_i \), \( n = \sum n_i \), which satisfy Assumptions A.1 - A.3. Then

\[ \text{i) } \text{E}(A) = \beta \prod_{i=1}^{k} n_i \]

\[ \text{ii) } \text{Var}(A) = \left( \prod_{i=1}^{k} n_i \right) \left[ \sum_{i'=1}^{k} \left( \prod_{i=1}^{k} (n_i - 1) \left\{ \beta_{2k-1} - \beta^2 \right\} + \sum_{i''=i}^{k} \left( n_i - 1 \right) \left\{ \beta_{2k-2} - \beta^2 \right\} + \ldots \right. \right] \]

**Proof:**

\[ \text{E}(A) = E \left[ \sum_{j_k=1}^{n_k} \ldots \sum_{j_1=1}^{n_1} M(X_{1j_1}, \ldots, X_{kj_k}) \right] = \sum_{j_k=1}^{n_k} \ldots \sum_{j_1=1}^{n_1} E \left[ M(X_{1j_1}, \ldots, X_{kj_k}) \right] \]

\[ = n_k \cdot \ldots \cdot n_1 \cdot E(M(\cdot, \ldots, \cdot)) \]

\[ = \beta \prod_{i=1}^{k} n_i. \]
Let $B = \left( \prod_{i=1}^{k} n_i \right)^{-1} A$. Then note that $B$ is a $K$-sample $U$-statistic with kernel $M(\cdot, \ldots, \cdot)$ for estimating the parameter $\beta$, the probability that $K$ randomly selected observations match on $X$. To invoke Lemma 1, it is required to demonstrate that $E(M(\cdot, \ldots, \cdot))^2 < \infty$. However, since $M$ is bounded between 0 and 1, it is obvious that $E[M(\cdot, \ldots, \cdot)]^2 < \infty$.

The conditions of Lemma 1 are met, and it is now invoked:

$g=1$, $m_1 = m_2 = \ldots = m_k = 1$.

$$\text{Var}(B) = \left( \prod_{i=1}^{k} n_i \right)^{-1} \left[ \sum_{d_1=0}^{1} \ldots \sum_{d_k=0}^{1} \prod_{i=1}^{k} \left( n_i - 1 \right) \zeta_{d_1, d_2, \ldots, d_k} \right]$$

Hence, derive the $\zeta_{d_0, d_2, \ldots, d_k}$.

$$\zeta_{1, 0, 0, \ldots, 0} = E \left[ \frac{M(X_{1j_1}, X_{2j_2}, \ldots, X_{kj_k}) \cdot M(X_{1j_1}, X_{2j_2}, \ldots, X_{kj_k})}{2} \right] - \left\{ E[M(\cdot, \ldots, \cdot)] \right\}$$

$$= E \left[ I \left\{ X_{1j_1} = X_{2j_2} = X_{3j_3} = \ldots = X_{kj_k} \right\} \cdot I \left\{ X_{1j_1} = X_{2j_2} = \ldots = X_{kj_k} \right\} \right] - \beta^2$$

$$= E \left[ I \left\{ X_{1j_1} = X_{2j_2} = X_{3j_3} = \ldots = X_{kj_k} \right\} \right] - \beta^2$$

$$= \Pr \left\{ 2K-1 \text{ randomly selected observations match on } X \right\} - \beta^2$$

$$= \beta_{2K-1} - \beta^2.$$

Similarly, $\zeta_{0, 0, \ldots, 0, 1, 0, \ldots, 0} = \beta_{2K-1} - \beta^2$ for all $i=1, 2, \ldots, K$ (1 at $i$th position)
\[ \zeta_{1,1,0,\ldots,0} = \mathbb{E}\left[M(\tau_{1j_1}, \tau_{2j_2}, \tau_{3j_3}, \ldots, \tau_{kj_k})M(\tau_{1j_1}, \tau_{2j_2}, \tau_{3j_3}, \ldots, \tau_{kj_k})\right] - \mathbb{E}(M(\cdot, \ldots, \cdot))^2 \]
\[ = \beta_{2K-2} - \beta^2 \]

Similarly, \( \zeta_{0,\ldots,0,1,0,\ldots,0,1,0,\ldots,0} = \beta_{2K-2} - \beta^2 \) for all \( i \neq i' \), (l's in positions \( i \) and \( i' \), \( i, i'=1,2,\ldots,K \), \( i \neq i' \))

\[ \zeta_{1,1,\ldots,1,0,1,\ldots,1} = \beta_{K+1} - \beta^2 \quad \text{for all } i=1,2,\ldots,K \]
(0 in \( i \)-th position)

\[ \zeta_{1,1,\ldots,1} = \beta - \beta^2 . \]

Hence, \( \zeta_{d_1,d_2,\ldots,d_k} = \beta - \frac{k}{2K} \sum_{i=1}^{k} d_i - \beta^2 \), for at least one \( d_i \neq 0 \).

(Note that when \( d_1=d_2=\ldots=d_k=0 \), \( \zeta_{d_1,d_2,\ldots,d_k} = 0 \).)

Then, from Lemma 2.1

\[
\text{Var}(B) = \left( \prod_{i=1}^{k} \frac{n_i}{n_i} \right)^{-1} \left[ \sum_{d_1=0}^{1} \sum_{d_k=0}^{1} \prod_{i=1}^{k} \left( 1-d_i \right) \zeta_{d_1,d_2,\ldots,d_k} \right]
\]
\[
= \left( \prod_{i=1}^{k} \frac{n_i}{n_i} \right)^{-1} \left[ \prod_{i=1}^{k-1} \left( n_i-1 \right) \zeta_{0,0,\ldots,0,1} + \prod_{i=1}^{k} (n_i-1) \zeta_{0,0,\ldots,0,1,0} \right]
\]
\[+ \prod_{i=1}^{k-1} (n_i-1) \zeta_{0,0,\ldots,1,0,0} + \prod_{i=1}^{k} (n_i-1) \zeta_{0,1,0,\ldots,0,0} \]
\[+ \prod_{i=2}^{k} (n_i-1) \zeta_{1,0,\ldots,0} + \prod_{i=1}^{k-2} (n_i-1) \zeta_{0,0,\ldots,0,1,1} \]
\[
\sum_{i=1}^{k} \left( \prod_{i \neq k-1} (n_i-1) \right) \zeta_{0,0,\ldots,1,0} + \sum_{i' \neq i''} \left( \prod_{i \neq k-2} (n_i-1) \right) \zeta_{1,1,\ldots,1,0} + \ldots + (n_1-1) \zeta_{0,1,\ldots,1,1,0} + \ldots
\]

\[
= \left( \prod_{i=1}^{k} \frac{n_i}{n_i} \right)^{-1} \left[ \sum_{i' \neq i''} \sum_{i=1}^{k} \prod_{i' \neq i''} (n_i-1) \left[ \beta_{2K-1} - \beta^2 \right] \right] + \sum_{i' \neq i''} \sum_{i=1}^{k} \prod_{i \neq i''} (n_i-1) \left[ \beta_{2K-2} - \beta^2 \right] + \ldots + \sum_{i=1}^{k} \prod_{i \neq i''} (n_i-1) \left[ \beta_{K+1} - \beta^2 \right] + \left[ \beta - \beta^2 \right] .
\]

Hence,

\[
\text{Var}(A) = \left[ \prod_{i=1}^{k} \frac{n_i}{n_i} \right]^{-1} \left[ \sum_{i' \neq i''} \sum_{i=1}^{k} \prod_{i' \neq i''} (n_i-1) \left[ \beta_{2K-1} - \beta^2 \right] \right] + \sum_{i' \neq i''} \sum_{i=1}^{k} \prod_{i \neq i''} (n_i-1) \left[ \beta_{2K-2} - \beta^2 \right] + \ldots + \sum_{i=1}^{k} \prod_{i \neq i''} (n_i-1) \left[ \beta_{K+1} - \beta^2 \right] + \left[ \beta - \beta^2 \right] .
\]

D. A Class of Test Statistics

Let

\[
U_i^* = \left( \prod_{i=1}^{k} \frac{n_i}{n_i} \right)^{-1} V_i^*
\]

where

\[
V_i^* = \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \ldots \sum_{j_k=1}^{n_k} \prod_{i'=1}^{k} \left\{ \left( Y_{ij_1} < Y_{i'j_1} \right) \right\}
\]

\[
= \sum_{j_1=1}^{n_1} \prod_{r \neq i} \sum_{r=1}^{n_r} \prod_{i'=1}^{k} \left\{ \left( Y_{ij_1} < Y_{rs} \right) \right\} .
\]

Then \( U_i^* \) is a K-sample U-statistic for estimating \( \phi_i \), with kernel

\[
h(Y_{1j_1}, Y_{2j_2}, \ldots, Y_{kj_k}) = \prod_{i'=1}^{k} \left\{ \left( Y_{ij_1} < Y_{i'j_1} \right) \right\} .
\]
and
\[
E[h(\cdot)] = E[Y_{ij_i} < Y_{i'j_i'}, \text{ for all } i' = 1, 2, \ldots, K, i' \neq i] = \phi_i.
\]

Bhapkar's V-test statistic is
\[
V = n(2K-1) \left[ \sum_{i=1}^{K} a_i (U_i^* - K^{-1})^2 - \left\{ \sum_{i=1}^{K} a_i (U_i^* - K^{-1}) \right\}^2 \right]
\]

where \( a_i = n_i/n \), as defined before.

A class of test statistics is advanced that is motivated by Bhapkar's V-test, but that incorporates \( X \). Let
\[
V_{i}^{(c)} = \sum_{j_i=1}^{n_i} \prod_{r \neq i} \left\{ \sum_{s=1}^{n_r} I\{Y_{ij_i} < Y_{rs}\} I\{X_{rs} = X_{ij_i} = X_{ij_i}^{(c)}\} \right\}.
\]

Then \( V_{i}^{(c)} \) represents the number of \( K \)-tuples for which the observation from sample \( i \) is least, calculated only for those observations whose covariables assume the value \( X_{ij_i}^{(c)} \). Let
\[
V_i = \sum_{c=1}^{C} b_c V_{i}^{(c)}
\]
\[
U_i = V_i / (\prod_{i} n_i)
\]

where the \( b_c \)'s are arbitrary weights, presumably functions of the \( e_c \)'s. Then consider a class of test statistics \( G \) whose members are representable as
\[
G = \sum_{i=1}^{K} a_i (U_i - \bar{U})^2 - \left[ \sum_{i=1}^{K} a_i (U_i - \bar{U}) \right]^2
\]

where
\[
\bar{U} = \frac{1}{K} \sum_{i} U_i.
\]

Members of the class will be considered relative to how the properties
of the test (i.e., efficiency) depend on the weights (the \( b_c \)'s). The AMK sign statistic is a member of \( G \), multiplied by an appropriate constant so that its asymptotic distribution is chi-squared.

\[
G^* = (\text{constant}) \left[ \sum_{i=1}^{k} a_i (U_i - \bar{U})^2 - \left\{ \sum_{i=1}^{k} a_i (U_i - \bar{U}) \right\}^2 \right].
\]

E. The AMK Sign Test Statistic

At this point, another assumption is imposed on the framework. In addition to Assumptions A.1-A.3, Assumption A.4 is now made.

Assumption A.4: The \( K \) populations are parallel in Bhapkar probability.

**Definition 3.6:** The \( K \) populations are said to be parallel in Bhapkar probability if the conditional probability \( \eta_i(V_c) \) is independent of the value of the covariable; i.e., \( \eta_i(V_c) \) is independent of \( c \).

It follows that, with Assumption A.4 imposed, the conditional probabilities are identical to the matched probabilities.

**Definition 3.7:** Let \( \gamma_i \equiv \sum c b_c \epsilon_c^{k} \eta_i(V_c) \). Then under the hypothesis of (conditional) homogeneity,

\[
\gamma_i = \frac{1}{K} \sum \epsilon c b_c^{k}, \quad i = 1, 2, \ldots, K.
\]

With the above defined parameter, the null hypothesis of interest may be expressed as

\[
H_0: \gamma_1 = \gamma_2 = \ldots = \gamma_K.
\]

Note that \( V_i \) may be expressed as
\[ V_i = \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \cdots \sum_{j_k=1}^{n_k} g^{(i)}(Y_{1j_1}, Y_{2j_2}, \ldots, Y_{kj_k}) \times \sum_c b_c I \left\{ x_{1j_1} = x_{2j_2} = \cdots = x_{kj_k} = V_c \right\} \]

where

\[ g^{(i)}(Y_{1j_1}, \ldots, Y_{kj_k}) = \begin{cases} 1 & \text{if } y_{ij_1} < y_{i'j_1}, \text{ for all } i'=1,2,\ldots,K, \\ 0 & \text{else} \end{cases} \]

(i.e., \( g^{(i)}(Y_{1j_1}, \ldots, Y_{kj_k}) = I\{y_{ij_1} = \min(y_{1j_1}, \ldots, y_{kj_k})\} \))

So,

\[ V_i = \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \cdots \sum_{j_k=1}^{n_k} h^{(i)}(x_{1j_1}, x_{2j_2}, \ldots, x_{kj_k}) \times \sum_c b_c I \left\{ x_{1j_1} = \cdots = x_{kj_k} = V_c \right\} \]

where \( h^{(i)}[\cdot] = g^{(i)}(Y_{1j_1}, \ldots, Y_{kj_k}) \sum_c b_c I \left\{ x_{1j_1} = \cdots = x_{kj_k} = V_c \right\} \).

Then \( U_i \) is a K-sample U-statistic of degree \((1,1,\ldots,1)\) for estimating \( \gamma_i \).

Let \( U^i = (U_1, U_2, \ldots, U_K) \)

and \( \gamma^i = (\gamma_1, \gamma_2, \ldots, \gamma_K) \).

(Under \( H_0, \ \gamma = 1 \cdot \frac{1}{K} (\sum_c b_c \epsilon_c) \))

And let \( \sigma_{rs} \) be the asymptotic covariance matrix of \( \sqrt{n}(U_n - \gamma) \).

An expression for the asymptotic null covariance matrix of \( \sqrt{n}(U_n - \gamma) \) is required, and is given in the following theorem.
Theorem 3.2: Let \( U_n \) be the vector of \( K \)-sample U-statistics and \( \gamma \) the vector of expected values, which is \( \frac{1}{K} (\sum c_c \epsilon_c^k) \) under the null hypothesis. Then, the asymptotic covariance matrix of \( \sqrt{n}(U_n - \gamma) \) is
\[
\Sigma = (\sigma_{rs}) \text{ given by:}
\]
\[
\sigma_{rr} = \sum b_c^2 \epsilon_c^{2K-1} \frac{1}{2K-1} \left[ a_r^{-1} + \frac{2}{K} \sum_{i=1}^{k} a_i^{-1} \right] - \left( \frac{1}{K} \sum b_c \epsilon_c^k \right)^2 \sum_{i=1}^{k} a_i^{-1}
\]
\[
\sigma_{rs} = \sum b_c^2 \epsilon_c^{2K-2} \frac{1}{K(2K-1)} \left[ a_r^{-1} + a_s^{-1} + 2 \sum_{i=1}^{k} a_i^{-1} \right] - \left( \frac{1}{K} \sum b_c \epsilon_c^k \right)^2 \sum_{i=1}^{k} a_i^{-1}
\]
\[
\sigma_{rr} > 0 \forall r, \sigma_{rs} < 0, r \neq s.
\]

Proof:

From Lemma 2.1, the asymptotic covariance matrix of \( \sqrt{n}(U_n - \gamma) \) is given by \( (\sigma_{rs}) \) where
\[
\sigma_{rs} = \sum_{i=1}^{k} a_i^{-1} \epsilon_i, (r, s) = 0, 0, \ldots, 1, 0, \ldots, 0 \quad \text{at ith position}, \quad r, s = 1, 2, \ldots, K.
\]
Hence, the \( \epsilon_{0,0}, \ldots, 1, \ldots, 0 \) (r,s) are derived.
\[ \xi_{1,0,0,\ldots,0}^{(1)} = E[h(1)(\gamma_{j_1}, \gamma_{j_2}, \ldots, \gamma_{k_j}) \cdot h(1)(\gamma_{j_1}, \gamma_{j_2}, \ldots, \gamma_{k_j})] - E[h(1)(\cdot)]^2 \]

\[ = E\left[ g^{(1)}(\gamma_{j_1}, \gamma_{j_2}, \ldots, \gamma_{k_j}) \sum_c b_c \ I \left\{ \xi_{1j_1} = \xi_{2j_2} = \ldots = \xi_{kj_k} = \gamma_c \right\} \right] \]

\[ \cdot g^{(2)}(\gamma_{j_1}, \gamma_{j_2}, \ldots, \gamma_{k_j}) \sum_c b_c \ I \left\{ \xi_{1j_1} = \xi_{2j_2} = \ldots = \xi_{kj_k} = \gamma_c \right\} \] \[ - \gamma_1^2 \]

\[ = \sum_c b_c^2 \epsilon_c^{2K-1} \cdot \frac{1}{2K-1} - \left( \frac{1}{K} \sum_c b_c \epsilon_c \right)^2 \]

Similarly,

\[ \xi_{0,0,\ldots,0,1,0,\ldots,0}^{(r)} = \sum_c b_c^2 \epsilon_c^{2K-1} \cdot \frac{1}{2K-1} - \left( \frac{1}{K} \sum_c b_c \epsilon_c \right)^2 \]

(1 at position \( r \))

\[ \xi_{0,1,0,\ldots,0}^{(1)} = E[h^{(1)}(\gamma_{j_1}, \gamma_{3j_3}, \ldots, \gamma_{k_j}) \cdot h^{(1)}(\gamma_{j_1}, \gamma_{3j_3}, \ldots, \gamma_{k_j})] - [E(h'(\cdot))]^2 \]

\[ = E\left[ g^{(1)}(\gamma_{j_1}, \gamma_{j_2}, \gamma_{3j_3}, \ldots, \gamma_{k_j}) \sum_c b_c \ I \left\{ \xi_{1j_1} = \xi_{2j_2} = \xi_{3j_3} = \ldots = \xi_{kj_k} = \gamma_c \right\} \right] \]

\[ \cdot g^{(1)}(\gamma_{j_1}, \gamma_{j_2}, \gamma_{3j_3}, \ldots, \gamma_{k_j}) \sum_c b_c \ I \left\{ \xi_{1j_1} = \xi_{2j_2} = \xi_{3j_3} = \ldots = \xi_{kj_k} = \gamma_c \right\} \]

\[ - [\gamma_1]^2 \]

\[ = \sum_c b_c^2 \epsilon_c^{2K-1} \cdot \frac{2}{K(2K-1)} - \left( \frac{1}{K} \sum_c b_c \epsilon_c \right)^2 \]
Similarly,

\[ \xi_{0,0,\ldots,1,\ldots,0}^{(r)} = \sum_c b_c^2 \varepsilon_c^{2K-1} \cdot \frac{2}{K(2K-1)} - \left( \frac{1}{K} \sum_c b_c \varepsilon_c \right)^2 \]

(1 at any position except r)

\[ \xi_{1,0,\ldots,0}^{(1,2)} = E[h^{(1)}(Y_{1j_1}, Y_{2j_2}, \ldots, Y_{k_j}) \cdot h^{(2)}(Y_{1j_1}, Y_{2j_2}, \ldots, Y_{k_j})] - \gamma_1 \gamma_2 \]

\[ = E \left[ g^{(1)}(Y_{1j_1}, Y_{2j_2}, \ldots, Y_{k_j}) \sum_c b_c I \left\{ X_{1j_1} = X_{2j_2} = \ldots = X_{k_j} = V_c \right\} \right] \cdot \]

\[ \cdot g^{(2)}(Y_{1j_1}, Y_{2j_2}, \ldots, Y_{k_j}) \sum_c b_c I \left\{ X_{1j_1} = X_{2j_2} = \ldots = X_{k_j} = V_c \right\} ] - \gamma_1 \gamma_2 \]

\[ = \sum_c b_c^2 \varepsilon_c^{2K-1} \cdot \frac{1}{K(2K-1)} - \left( \frac{1}{K} \sum_c b_c \varepsilon_c \right)^2 \]

Similarly,

\[ \xi_{0,0,\ldots,1,0,\ldots,0}^{(r,s)} = \sum_c b_c^2 \varepsilon_c^{2K-1} \cdot \frac{1}{K(2K-1)} - \left( \frac{1}{K} \sum_c b_c \varepsilon_c \right)^2 \]

(1 at position r or s)

\[ \xi_{0,0,1,0,\ldots,0}^{(1,2)} = E[h^{(1)}(Y_{1j_1}, Y_{2j_2}, Y_{3j_3}, Y_{4j_4}, \ldots, Y_{k_j}) \cdot \]

\[ \cdot h^{(2)}(Y_{1j_1}, Y_{2j_2}, Y_{3j_3}, Y_{4j_4}, \ldots, Y_{k_j})] - \gamma_1 \gamma_2 \]

\[ = E \left[ g^{(1)}(Y_{1j_1}, Y_{2j_2}, Y_{3j_3}, Y_{4j_4}, \ldots, Y_{k_j}) \sum_c b_c I \left\{ X_{1j_1} = X_{2j_2} = X_{3j_3} = X_{4j_4} = \ldots = X_{k_j} = V_c \right\} \right] \cdot \]

\[ \sum_c b_c I \left\{ X_{1j_1} = X_{2j_2} = X_{3j_3} = X_{4j_4} = \ldots = X_{k_j} = V_c \right\} \]
\[ g^{(2)}(Y_{1j_1}, Y_{2j_2}, \ldots, Y_{kj_k}) \times \]
\[ \sum_{c} b_c \epsilon_c \{ x_{1j_1} = x_{2j_2} = \ldots = x_{kj_k} = \gamma_c \} \] - \gamma_1 \gamma_2
\]
\[ = \sum_{c} b_c^2 \epsilon_c^{2K-1} \cdot \frac{2}{K(2K-1)} - \left( \frac{1}{K} \sum_{c} b_c \epsilon_c^k \right)^2 \]

Similarly,
\[ (r, s) \]
\[ \xi_{0,0,\ldots,0,1,0,\ldots,0} = \sum_{c} b_c^2 \epsilon_c^{2K-1} \cdot \frac{2}{K(2K-1)} - \left( \frac{1}{K} \sum_{c} b_c \epsilon_c^k \right)^2 \]
(1 at any position except r or s)

\[ \sigma_{rr} = \sum_{i=1}^{k} a_i^{-1} \xi_{0,0,\ldots,1,0,\ldots,0} \]
(r)
(1 at rth place)

So \[ \sigma_{rr} = a_r^{-1} \cdot \xi_{0,0,\ldots,1,0,\ldots,0} + \sum_{i=1}^{k} a_i^{-1} \xi_{0,0,\ldots,1,0,\ldots,0} \]
(1 at rth place) \[ i \neq r \]
(1 at ith place)

\[ = a_r^{-1} \left[ \sum_{c} b_c^2 \epsilon_c^{2K-1} \cdot \frac{1}{2K-1} - \left( \frac{1}{K} \sum_{c} b_c \epsilon_c^k \right)^2 \right] \]

\[ + \sum_{i=1}^{k} a_i^{-1} \left[ \sum_{c} b_c^2 \epsilon_c^{2K-1} \cdot \frac{2}{K(2K-1)} - \left( \frac{1}{K} \sum_{c} b_c \epsilon_c^k \right)^2 \right] \]

\[ i \neq r \]

\[ \sigma_{rr} = \sum_{c} b_c^2 \epsilon_c^{2K-1} \cdot \frac{1}{2K-1} \left[ a_r^{-1} + \frac{2}{K} \sum_{i=1}^{k} a_i^{-1} \right] - \left( \frac{1}{K} \sum_{c} b_c \epsilon_c^k \right)^2 \]

\[ \sum_{i=1}^{k} a_i^{-1} \]

and
\[ \sigma_{rs} = \sum_{i=1}^{k} a_i^{-1} \xi_{0,0,\ldots,1,0,\ldots,0} \]
(r, s)
(1 at ith place)
\[
\sigma_{rs} = \sum_{c} b_{c}^{2} e_{c}^{2K-1} \left( \frac{1}{K(2K-1)} \left[ (a_{r}^{-1} + a_{s}^{-1}) + 2 \sum_{i=1}^{k} a_{i}^{-1} \right] - \left( \frac{1}{K} \sum_{c} b_{c} e_{c}^{k} \right)^{2} \right)_{i \neq r, s}
\]

It is easy to see that conditional on \( \bar{U} \), the \( U_{i} \)'s are subject to a single linear constraint, and their distribution is singular, so their asymptotic distribution is also singular, and \( \sum_{i} \) is singular. In fact, \( \text{Rank}(\sum) = K-1 \).

A judicious restriction is now imposed on the \( b_{c} \)'s. Henceforth, weights are arbitrary subject to the condition that

\[
\sum_{c} b_{c}^{2} e_{c}^{2K-1} = \left( \sum_{c} b_{c} e_{c}^{k} \right)^{2}.
\]

With this restriction on the \( b_{c} \)'s, the following result may be observed:

\[
\sigma_{rr} + \sum_{s=1}^{k} \sigma_{rs} = \sum_{c} b_{c}^{2} e_{c}^{2K-1} \left\{ \frac{1}{K(2K-1)} \left[ 2 \sum_{i=1}^{k} a_{i}^{-1} + (K-2)a_{r}^{-1} \right] - \frac{1}{K} \sum_{i=1}^{k} a_{i}^{-1} \right\}
\]

\[
= \sum_{c} b_{c}^{2} e_{c}^{2K-1} \left\{ \frac{1}{K(2K-1)} \left[ 2 \sum_{i=1}^{k} a_{i}^{-1} - a_{r}^{-1} - a_{s}^{-1} \right] - \frac{1}{K} \sum_{i=1}^{k} a_{i}^{-1} \right\}
\]

\[
= \sum_{c} b_{c}^{2} e_{c}^{2K-1} \left[ \frac{1}{(K-1)} \left[ 2 \sum_{i=1}^{k} a_{i}^{-1} + (K-2)a_{r}^{-1} + (K-1) \cdot 2 \sum_{i'=1}^{k} a_{i'}^{-1} - (K-1)a_{r}^{-1} \right] \right]
\]
\[- \frac{1}{s} \sum_{s=1}^{s \neq r} a_s^{-1} \] \[- \frac{1}{2} K \cdot \frac{1}{K} \sum_{i=1}^{k} a_i^{-1} \] 

\[= \sum_{c} b_c^2 \epsilon_c^{2K-1} \left\{ \frac{1}{(2K-1)} \sum_{i=1}^{k} a_i^{-1} - \frac{1}{K} \sum_{i=1}^{k} a_i^{-1} \right\} 
= \sum_{c} b_c^2 \epsilon_c^{2K-1} \left\{ \frac{1}{(2K-1)} \sum_{i=1}^{k} a_i^{-1} - \frac{1}{K} \sum_{i=1}^{k} a_i^{-1} \right\} 
= 0.

Hence, subject to \[\sum_{c} b_c^2 \epsilon_c^{2K-1} = \left( \sum_{c} b_c \epsilon_c \right)^k, \]

\[\sum_{c} b_c \epsilon_c = 0. \]

1xK

Consequently, any K-1 of the \(U_i\)'s may be selected for a nonsingular vector.

\[d'_0 = \sqrt{n} (U_1 - \bar{U}, U_2 - \bar{U}, \ldots, U_{K-1} - \bar{U})\]

and let \(\sum_0\) be the corresponding covariance matrix. Then \(\sum_0\) is of full rank, and its inverse exists.

**Corollary 3.2:** Let \(d'_0\) be as defined above, and let \(\sum_0\) be the corresponding asymptotic covariance matrix. Then

\[G^* \equiv d'_0 \sum_0^{-1} d_0\]

has asymptotically a null distribution which is \(\chi^2_{(K-1)}\).
Proof:

Using arguments parallel to those employed in Chapter II to derive the asymptotic null distribution of $D^*$, it is easy to show that $G^*$ is also $\chi^2_{(K-1)}$. The details are omitted.

It is established that $G^* = d' \sum_{i=0}^{+} d_i$ has a null asymptotic distribution that is $\chi^2_{(K-1)}$. To demonstrate that this quadratic form is indeed the AMK test statistic requires that $\sum_{i=0}^{+}$ be inverted. In general, this is a very difficult task, hence, one special case will be considered: the case of equal sample sizes. However, prior to considering the special case, the limiting distribution of $G^*$ under certain alternative hypotheses is considered.

F. The Limiting Distribution of $G^*$ Under Translation-Type Alternatives

The purpose of this section is to investigate the distribution of $G^*$ assuming a sequence of translation-type alternative hypotheses, $H_n$, for $n=1,2,\ldots,\ldots$. The hypothesis $H_n$ specifies that

$$F_i(y|x) = F(y-n^{-1/2} \theta_i|x), \; i=1,2,\ldots,K,$$

where $\theta_i \neq \theta_{i'}$, for some $i' \neq i$. The limiting distribution is, then, found as $n \to \infty$.

**Theorem 3.3:** For each index $n$ assume $\lim_{n \to \infty} \frac{n_i}{n} = a_i$ exists and $0 < a_i < 1$, and the truth of $H_n$. If $F$ possesses a continuous derivative $f$, and there exists a function $g$ such that

$$|[f(y+h|x) - f(y|x)/h] \leq g(y|x)$$

3.3.1
and
\[ \int_{-\infty}^{\infty} g(y|x) f(y|x) \, dy|x < \infty, \] 3.3.2

then for \( n \to \infty \), the statistic \( G^* \) has a limiting noncentral \( \chi^2 \) distribution with \( K-1 \) degrees of freedom, and noncentrality parameter

\[ \lambda_{G^*} = \left( \sum_{c} b_c c_{c} \right)^2 \delta_{-0} \sum_{0} \left[ \int_{-\infty}^{\infty} [1-F(y|x)] K-2 \, f^2(y|x) \, dy|x \right] \]

(Note: Although \( \lambda_{G^*} \) appears to depend on \( \tilde{x} \), by Assumption A.4 it is invariant to the choice of any particular value of \( \tilde{x} \), and hence does not depend on \( \tilde{x} \).)

Proof:

Let \( \gamma_n^{(i)} = E[U_i(\cdot)|H_n] \)

\begin{align*}
&= E[U_i[(Y_{\tilde{x}1j_1}, X_{\tilde{x}1j_1}), (Y_{\tilde{x}2j_2}, X_{\tilde{x}2j_2}), \ldots, (Y_{k\tilde{x}k_j}, X_{k\tilde{x}k_j})]|H_n] \\
&= E[g^{(i)}(Y_{\tilde{x}1j_1}, Y_{\tilde{x}2j_2}, \ldots, Y_{k\tilde{x}k_j}) \sum_{c} b_c I\{X_{\tilde{x}1j_1} = X_{\tilde{x}2j_2} = \ldots = X_{k\tilde{x}k_j} = \tilde{c}\}] \\
\end{align*}

where (recall) \( g^{(i)}(Y_{\tilde{x}1j_1}, \ldots, Y_{k\tilde{x}k_j}) = \begin{cases} 
1 & \text{if } Y_{i\tilde{x}i} < Y_{i'\tilde{x}i'}, \ i', \ i'' = 1, 2, \ldots, K, \\
0 & \text{else} 
\end{cases} \)

\begin{align*}
&= \sum_{c} b_c c_{c} \cdot \Pr\{Y_{i\tilde{x}i} < Y_{i'\tilde{x}i'}, \ i'' = 1, 2, \ldots, K, \ i' \neq i \mid X_{\tilde{x}1j_1} = X_{\tilde{x}2j_2} = \ldots = X_{k\tilde{x}k_j}\} \\
&= \sum_{c} b_c c_{c} \cdot \psi_i \\
\end{align*}

Then,

\[ \psi_i = \Pr[ Y_{i\tilde{x}i} < Y_{\tilde{x}1j_1}, Y_{i\tilde{x}i} < Y_{\tilde{x}2j_2}, \ldots, Y_{i\tilde{x}i} < Y_{k\tilde{x}k_j} \mid X = \tilde{x}] \]

all except \( i \)
By Assumption A.4, \( \psi_1 \) is the same for all \( \sim \), hence arbitrarily choose \( \sim = x \), arbitrary but fixed.

\[
\psi_1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_1(y_1|x) f_2(y_2|x) \cdots f_k(y_k|x) \cdot f_i(y_i|\sim) \, dy_1|\sim \, dy_2|\sim \cdots \\
\cdots dy_k|\sim \\
= \int_{-\infty}^{\infty} [1-F_1(y_1|x)] [1-F_2(y_2|x)] \cdots [1-F_k(y_k|x)] f_i(y_i|\sim) \, dy_i|\sim
\]

Under \( H_n \),

\[
F_i(y|x) = F(y - \theta_i / \sqrt{n}|x).
\]

Thus,

\[
\psi_1 = \int_{-\infty}^{\infty} [1-F(y_1 - \theta_1 / \sqrt{n}|\sim)] [1-F(y_1 - \theta_2 / \sqrt{n}|\sim)] \cdots \\
\cdots [1-F(y_1 - \theta_k / \sqrt{n}|\sim)] \, dF(X_1 - \theta_i / \sqrt{n})
\]

Change of variables: Let \( y = y_i - \theta_i / \sqrt{n} \)

\[
dy = dy_i
\]

\[
\psi_1 = \int_{-\infty}^{\infty} [1-F(y + \theta_1 / \sqrt{n} - \theta_1 / \sqrt{n}|\sim)] [1-F(y + \theta_1 / \sqrt{n} - \theta_2 / \sqrt{n}|\sim)] \cdots \\
\cdots [1-F(y + \theta_1 / \sqrt{n} - \theta_k / \sqrt{n}|\sim)] \, f(y|x) \, dy|\sim
\]

Write

\[
\psi_1(t) = \int_{-\infty}^{\infty} [1-F(y + (\theta_1 - \theta_1)t|x)] [1-F(y + (\theta_1 - \theta_2)t|x)] \cdots \\
\cdots [1-F(y + (\theta_1 - \theta_k)t|x)] \, f(y|x) \, dy|x
\]

Expand \( \psi_1(t) \) in a Maclaurin series in \( t \).
The constant term is clearly \( \frac{1}{K} \) (since this corresponds to \( \psi_i(0) = \int_{-\infty}^{\infty} [1-F(y|x)]^{K-1} f(y|x) dy \), i.e., \( F_i(Y|x) = F(y|x) \) for all \( i \), and \( \psi_i(0), 1/K \)).

Find the linear term:

\[
\frac{d}{dt} \psi_i(t) = \int_{-\infty}^{\infty} \frac{d}{dt} \left\{ [1-F(y+(\theta_i-\theta_1)t|x)] [1-F(y+(\theta_i-\theta_2)t|x)] \ldots \cdot [1-F(y+(\theta_i-\theta_k)t|x)] f(y|x) \right\} dy|x
\]

(This derivative exists, and differentiation under the integral sign is permitted by Assumptions 3.3.1 and 3.3.2.)

\[
\frac{d}{dt} \psi_i(t) = [-f(y+(\theta_i-\theta_1)t|x) \cdot (\theta_i-\theta_1)] [1-F(y+(\theta_i-\theta_2)t|x)] \ldots [1-F(y+(\theta_i-\theta_k)t|x)] f(y|x)
\]

\[+ \left[ 1-F(y+(\theta_i-\theta_1)t|x) \right] [-f(y+(\theta_i-\theta_2)t|x) \cdot (\theta_i-\theta_2)] [1-F(y+(\theta_i-\theta_3)t|x)] \ldots \cdot [1-F(y+(\theta_i-\theta_k)t|x)] f(y|x) + \ldots
\]

\[+ 1-F(y+(\theta_i-\theta_1)t|x) [1-F(y+(\theta_i-\theta_2)t|x)] \ldots [1-F(y+(\theta_i-\theta_k)t|x)] f(y|x) + [1-F(y+(\theta_i-\theta_1)t|x)] \ldots [1-F(y+(\theta_i-\theta_k)t|x)] \frac{d}{dt} f(y|x)
\]

And

\[
\frac{d}{dt} \psi_i(t) \bigg|_{t=0} = -\sum_{i'=1}^{k} (\theta_i-\theta_{i'}) \int_{-\infty}^{\infty} \left\{ \left[1-F(y|x)\right]^{K-2} [1-F(y|x)]^{K-2} \right\} dy|x
\]

Let

\[
\lambda = \int_{-\infty}^{\infty} \left[1-F(y|x)\right]^{K-2} f^2(y|x) dy|x
\]

and

\[
\delta_i = \sum_{i'=1}^{k} (\theta_i-\theta_{i'}).\]
Then
\[ \psi_i(t) = \frac{1}{K} - \delta_i \lambda_t + 0(t). \]

But
\[ \psi_i\left(\frac{1}{\sqrt{n}}\right) = \psi_i \]
\[ = E[U_i(\cdot) | H_n]; \]

thus, under \( H_n \)
\[ E[U_i] = \sum_c b_c \varepsilon_c^k \left[ \frac{1}{K} - \frac{\delta_i}{\sqrt{n}} \lambda + O(\frac{1}{n}) \right] \]

By a similar argument
\[ \bar{\gamma}_n = \bar{\gamma} + O(n^{-3/2}), \]

where \( \bar{\gamma}_n \) is the covariance matrix of \( \sqrt{n} (\bar{\gamma}_n - \gamma_n) \) as \( n \to \infty \), and \( O(n^{-3/2}) \) is a matrix whose elements are \( O(n^{-3/2}) \). It follows from Lemma 2.1 that, under hypothesis \( H_n \),
\[ \sqrt{n} (\bar{\gamma}_n - \gamma_n), \text{ for large } n, \]

has zero mean vector and covariance matrix \( \bar{\gamma} \) (asymptotically), and is asymptotically normally distributed. Consequently
\[ \sqrt{n} (\bar{\gamma} - \tilde{\gamma}) \]

has a limiting normal distribution with mean vector
\[ - \sqrt{n} \left( \sum_c b_c^k \varepsilon_c^k \right) + \sqrt{n} \left( \sum_c b_c^k \varepsilon_c^k \cdot \left( \frac{1}{K} - \frac{\delta_i}{\sqrt{n}} \lambda \right) \right) = \sqrt{n} \left( \sum_c b_c^k \varepsilon_c^k \cdot \left[ - \frac{\delta_i}{\sqrt{n}} \right] \right). \]

The mean vector is, hence, \( \bar{\delta}' = (\delta_1, \delta_2, \ldots, \delta_K) \).

Note that \( \bar{\delta}_i = 0 \). And let \( \bar{\delta}'_0 = (\delta_1, \delta_2, \ldots, \delta_{K-1}) \). Then from Lemma 2.2, \( \sqrt{n} (\bar{\gamma} - \tilde{\gamma}), \) under hypothesis \( H_n \), is asymptotically distributed as a \( \chi^2 \) with \( K-1 \) degrees of freedom, and noncentrality parameter.
\[ \lambda_G^* = \left( \sum_{c} b_c \varepsilon_c^k \right)^2 \lambda^* \delta_0 \sum_{i=0}^{n-1} \delta_0. \]

\( \delta_0 \sum_{i=0}^{n-1} \delta_0 \) may be simplified if \( \sum_{i=0}^{n-1} \) is explicitly derived.

G. The AMK Sign Statistic: A Special Case: \( n_1 = n_2 = \ldots = n_K \)

In this section, the AMK sign statistic is presented for the K-sample problem when the sample sizes are equal. Recall that, in general,

\[ G = \sum_{i=1}^{k} a_i (U_i - \bar{U})^2 - \left[ \sum_{i=1}^{k} a_i (U_i - \bar{U}) \right]^2 \]

for

\[ U_i = V_i / \ln n_i, \]

\[ V_i = \sum_{j_k=1}^{n_k} \ldots \sum_{j_1=1}^{n_1} g(i)(Y_{j_k \ldots j_1}, \ldots, Y_{j_k \ldots j_1}) \sum_{c} b_c I \left\{ X_{j_k \ldots j_1} = \ldots = X_{j_k \ldots j_1} = V_c \right\} \]

where

\[ g(i)(Y_{j_k \ldots j_1}, \ldots, Y_{j_k \ldots j_1}) = \begin{cases} 1 & \text{if } Y_{j_i \ldots j_i} < Y_{i' j_i} \text{, } i' = 1, 2, \ldots, K, i' = i \\ 0 & \text{else} \end{cases} \]

and the \( n_i \)'s are determined so that

\[ n_i = n a_i \]

Note that \( n_1 = n_2 = \ldots = n_K \) only if \( a_1 = a_2 = \ldots = a_K = \frac{1}{K} \).

1. \( \sum_{i=0}^{n-1} \delta_0 \) is derived.

Recall, \( \sum_{i=0}^{n-1} = (\sigma_{rs}) \), where

\[ \sigma_{rr} = \sum_{c} b_c^2 \varepsilon_c^{2K-1} \cdot \frac{1}{K(2K-1)} \left[ (K-2) a_r - 2 \sum_{i=1}^{k} a_i^{-1} \right] - \left( \frac{1}{K} \sum_{c} b_c \varepsilon_c^k \right)^2 \sum_{i=1}^{k} a_i^{-1} \]

\[ r = 1, 2, \ldots, K-1, \]
\[ \sigma_{rs} = \sum_c b_c^2 \varepsilon_c^{2K-1} \cdot \frac{1}{K(2K-1)} \left[ 2 \sum_{i=1}^k a_i^{-1} - (a_r^{-1} + a_s^{-1}) \right] - \left( \frac{1}{K} \sum_c b_c \varepsilon_c^k \right)^2 \sum_{i=1}^k a_i^{-1} \]

\[ r, s = 1, 2, \ldots, K-1; \ r \neq s. \]

Since, in this case, \( a_i^{-1} = K \) for all \( i \),

\[ \sigma_{rr} = \sum_c b_c^2 \varepsilon_c^{2K-1} \cdot \frac{1}{K(2K-1)} \left[ (K-2)K + 2 \sum_{i=1}^k \right] - \left( \frac{1}{K} \sum_c b_c \varepsilon_c^k \right)^2 \sum_{i=1}^k K \]

\[ = \sum_c b_c^2 \varepsilon_c^{2K-1} \cdot \frac{1}{K(2K-1)} \left[ K(3K-2) - \left( \sum_c b_c \varepsilon_c^k \right)^2 \right] \]

\[ \sigma_{rr} \equiv g , \quad r = 1, 2, \ldots, K-1. \]

\[ \sigma_{rs} = \sum_c b_c^2 \varepsilon_c^{2K-1} \cdot \frac{1}{K(2K-1)} \left[ 2 \sum_{i=1}^k K - (K+K) \right] - \left( \frac{1}{K} \sum_c b_c \varepsilon_c^k \right)^2 \cdot \sum_{i=1}^k K \]

\[ = \sum_c b_c^2 \varepsilon_c^{2K-1} \cdot \frac{1}{K(2K-1)} \left[ 2K^2 - 2K \right] - \left( \frac{1}{K} \sum_c b_c \varepsilon_c^k \right)^2 \cdot K^2 \]

\[ = \sum_c b_c^2 \varepsilon_c^{2K-1} \cdot \frac{2(K-1)}{2K-1} - \left( \sum_c b_c \varepsilon_c^k \right)^2 \]

\[ \sigma_{rs} \equiv h , \quad r, s = 1, 2, \ldots, K-1; \ r \neq s. \]

Hence \( \sum_{i=0}^K = g I_{K-1} + h (J_{K-1} - I_{K-1}) \), which is a standard form that can be easily inverted. It follows from Lemma 2.3 that

\[ \sum_{i=0}^{-1} = \left( \sum_c b_c^2 \varepsilon_c^{2K-1} \cdot \frac{K}{2K-1} \right)^{-1} \left[ I - \frac{\left[ \sum_c b_c^2 \varepsilon_c^{2K-1} \cdot \frac{2K-2}{2K-1} - \left( \sum_c b_c \varepsilon_c^k \right)^2 \right]}{\left[ \sum_c b_c^2 \varepsilon_c^{2K-1} \cdot \frac{2K^2-3K+2}{2K-1} - (K-1) \left( \sum_c b_c \varepsilon_c^k \right)^2 \right]} \right] \]
\[
\left\{ \text{since } g + (K-2)h = \sum \mathbf{c} \cdot \mathbf{b}^2 \mathbf{e}^{2K-1} \left( \frac{3K-2+(K-2)(2K-2)}{2K-1} \right) - \left( \sum \mathbf{c} \cdot \mathbf{e}^k \right)^2 (K-1) \right\}
\]

= \left[ \sum b_c^2 e_c^{2K-1} \left( \frac{2K^2 - 3K+2}{2K-1} \right) - (K-1) \left( \sum b_c e_c^k \right)^2 \right].

Subject to the restriction that
\[
\left[ \sum b_c^2 e_c^{2K-1} \right] = \left( \sum b_c e_c^k \right)^2
\]

\[
\sum_{o}^{-1} = \left[ \sum b_c^2 e_c^{2K-1} \cdot \frac{K}{2K-1} \right]^{-1} \left[ I - \frac{2K-2}{2K-1} - 1 \right]
\]

And
\[
\frac{2K-2}{2K-1} - 1 = \frac{2K-2 - 2K+1}{2K^2 - 3K+2 - (K-1) \left( 2K^2 - 3K+1 \right)} = -1
\]

So
\[
\sum_{o}^{-1} = \left[ \sum b_c^2 e_c^{2K-1} \cdot \frac{K}{2K-1} \right]^{-1} \left[ I + J \right]
\]

2. Statistic \( G^* = d_0^t \sum_{o}^{-1} d_0 \) as expressed explicitly.

Note, then, that \( G^* \) may be expressed as
\[
d_0^t \sum_{o}^{-1} d_0 = \left[ \sum b_c^2 e_c^{2K-1} \cdot \frac{K}{2K-1} \right]^{-1} \left[ \sum_{i=1}^{K-1} n(u_i - \bar{u})^2 + \left[ \sum_{i=1}^{K-1} \sqrt{n} \left( u_i - \bar{u} \right) \right]^2 \right]
\]

\[
= \left[ \sum b_c^2 e_c^{2K-1} \right]^{-1} \cdot \frac{n(2K-1)}{K} \left\{ \sum_{i=1}^{K-1} (u_i - \bar{u})^2 + \left[ \sum_{i=1}^{K-1} (u_i - \bar{u}) \right]^2 \right\}
\]

\[
= \left[ \sum b_c^2 e_c^{2K-1} \right]^{-1} \cdot \frac{n(2K-1)}{K} \left\{ \sum_{i=1}^{K-1} (u_i - \bar{u})^2 + \left[ \bar{u} u_K - \frac{K-1}{K} \bar{u} \right]^2 \right\}
\]
\[ d_0^{K-1} d_0 = \left( \sum \limits_c b_c^2 \epsilon_c^{2K-1} \right)^{-1} n_{(2K-1)} \left( \frac{K-1}{K} \sum \limits_{i=1}^k (U_i - \bar{U})^2 + (-U_K + \bar{U})^2 \right) \]

\[ G^* = \left( \sum \limits_c b_c^2 \epsilon_c^{2K-1} \right)^{-1} n_{(2K-1)} \left( \frac{k}{K} \sum \limits_{i=1}^k (U_i - \bar{U})^2 \right) \]

3. **The limiting distribution of \( G^* \) under the null hypothesis.**

The following result has been established.

**Corollary 3.2:** Let \( G^* \) be the AMK sign statistic with \( n_1 = n_2 = \ldots = n_K \).

\[ G^* = \left( \sum \limits_c b_c^2 \epsilon_c^{2K-1} \right)^{-1} n_{(2K-1)} \left( \frac{k}{K} \sum \limits_{i=1}^k (U_i - \bar{U})^2 \right) \]

where the \( b_c \)'s are arbitrary weights subject to the restriction that

\[ \sum \limits_c b_c^2 \epsilon_c^{2K-1} = \left( \sum \limits_c b_c \epsilon_c^{K} \right)^2 \]

and the \( \epsilon_c, U_i, \bar{U} \) are as previously defined.

Then, \( G^* \) is asymptotically distributed as a chi-squared random variable with \( K-1 \) degrees of freedom, under the hypothesis of [conditional] homogeneity.

4. **The limiting distribution of \( G^* \) under translation-type alternatives when \( n_1 = n_2 = \ldots = n_K \).**

As in Section F, assume a sequence of translation-type alternative hypotheses, \( H_n \), for \( n=1,2,\ldots \), that approach the null hypothesis as \( n \to \infty \),

\[ F_i(y|x) = F(y-n^{-\frac{1}{2}} \theta_i|x) \quad \text{for all } i. \]

**Corollary:** For each index \( n \), assume the truth of \( H_n \) if \( F \) possesses a continuous derivative \( f \), and there exists a function \( g \) such that

\[ |[f(y+h|x) - f(y|x)]| \leq g(y|x) \]
and
\[ \int_{-\infty}^{\infty} g(y|x) f(y|x) \, dy \mid x < \infty, \]

then for \( n \to \infty \), the statistic \( G^* \) has a limiting noncentral \( \chi^2 \) distribution with \( K-1 \) degrees of freedom and noncentrality parameter

\[ \lambda_{G^*} = K \left( 2K-1 \right) \sum_{i=1}^{k} \left( \bar{\theta}_i - \bar{\theta} \right)^2 \left[ \int_{-\infty}^{\infty} \left[ 1 - F(y|x) \right]^{k-2} f^2(y|x) \, dy \mid x \right]^2. \]

**Proof:**

The result follows directly from Theorem 3.3 of Section F, with simplifying the noncentrality parameter.

\[
\delta_{\sum_{i=1}^{k} \delta_i} = \left[ \sum_{i=1}^{k} b_c e^c \frac{2K-1}{2K-1} \right]^{-1} \left\{ \sum_{i=1}^{k} \delta_i^2 + \left( \sum_{i=1}^{K-1} \delta_i \right)^2 \right\} \\
= \left[ \sum_{i=1}^{k} b_c e^c \frac{2K-1}{2K-1} \right]^{-1} \left\{ \sum_{i=1}^{k} \delta_i^2 + \left( \sum_{i=1}^{K-1} \delta_i \right)^2 \right\}, \text{ since } \sum_{i=1}^{k} \delta_i = 0 \\
= \left[ \sum_{i=1}^{k} b_c e^c \frac{2K-1}{2K-1} \right]^{-1} \sum_{i=1}^{k} \delta_i^2.
\]

Let \( \bar{\theta} = \frac{1}{K} \sum_{i=1}^{k} \theta_i \).

Then
\[
\sum_{i=1}^{k} \delta_i^2 = \sum_{i=1}^{k} \left( K\theta_i - \bar{\theta} \right)^2 = \sum_{i=1}^{k} \left( K\theta_i - K\bar{\theta} \right)^2 = K^2 \sum_{i=1}^{k} \left( \theta_i - \bar{\theta} \right)^2.
\]

Hence
\[
\lambda_{G^*} = \left( \sum_{i=1}^{k} b_c e^c \right)^2 \lambda^2 \cdot \left[ \sum_{i=1}^{k} b_c e^c \frac{2K-1}{2K-1} \right]^{-1} \frac{2K-1}{K} \lambda^2 \sum_{i=1}^{k} \left( \theta_i - \bar{\theta} \right)^2 \\
= \frac{\left( \sum_{i=1}^{k} b_c e^c \right)^2}{\sum_{i=1}^{k} b_c e^c \frac{2K-1}{2K-1}} \cdot (2K-1)K \cdot \sum_{i=1}^{k} \left( \theta_i - \bar{\theta} \right)^2 \lambda^2.
\]
5. Determining explicit weights for G* when $n_1 = \ldots = n_K$.

It is desirable for weights of most statistics to have certain optimum properties. A component of this research has been to investigate the properties of the AMK sign statistic, particularly as efficiency is related to the weights. A restriction was imposed on the weights at the end of Section E,

$$
\sum_c b_c^2 \varepsilon_c^{2K-1} = \left( \sum_c b_c \varepsilon_c^k \right)^2.
$$

For the equal sample size problem, the noncentrality parameter is expressed explicitly, and does not depend on the weights. Hence, the efficacy of the test is invariant to the choice of weights subject to

$$
\sum_c b_c^2 \varepsilon_c^{2K-1} = \left( \sum_c b_c \varepsilon_c^k \right)^2.
$$

A convenient choice is

$$
b_c = \varepsilon_c^{-K+1}.
$$

Then

$$
\sum_c b_c^2 \varepsilon_c^{2K-1} = \sum_c \left( \varepsilon_c^{-K+1} \right)^2 \varepsilon_c^{2K-1} = \sum_c \varepsilon_c = 1;
\text{ and } \left( \sum_c b_c \varepsilon_c^k \right)^2 = \left( \sum_c \varepsilon_c^{-K+1} \varepsilon_c^k \right)^2 = 1.
$$

Hence, for the equal sample size case,

$$
G^* = \frac{n(2K-1)}{K} \sum_{i=1}^k (U_i - \bar{U})^2,
$$

where

$$
U_i = V_i / \prod_i n_i,
$$

$$
V_i = \sum_{i=1}^n \sum g^{(i)}(y_{1j_1}, \ldots, y_{kj_k}) \sum_c \varepsilon_c^{-K+1} \prod_{j_1}^k \prod c_j \varepsilon_c
$$

$$
g^{(i)}(\cdot) = \mathbb{I}\{i\text{th observation is least}\}.
$$

Efficiency considerations will be discussed in Chapter IV. However,
it is noted here that, when \( \tilde{X} \) is independent of \( Y \), \( i.e., F_{1}(Y|\tilde{X}) = F_{1}(Y) \), the ARE of \( G^{*} \) relative to Bhapkar's \( V \)-test is unity. Hence, asymptotically, \( G^{*} \) performs as well as Bhapkar's \( V \)-test when Bhapkar's test is appropriate.

6. The AMK sign statistic for equal sample sizes with estimated weights.

If the \( \varepsilon_{c} \)'s are unknown, then the AMK analysis is based on

\[
\tilde{G} = \frac{n(2K-1)}{K} \sum_{i=1}^{k} (U_{i} - \bar{U})^{2}
\]

where

\[
U_{i} = \left( \prod_{i} n_{i} \right)^{-1} \sum_{j_{1}} \ldots \sum_{j_{k}} g(i) (y_{1j_{1}}, \ldots, y_{kj_{k}}) \sum_{c} \hat{w}_{c} I \{ x_{1j_{1}} = \ldots = x_{kj_{k}} = C \},
\]

where, for each \( c \), the selected weights \( \hat{w}_{c} = \varepsilon_{c}^{-K+1} \) are estimated by

\[
\hat{w}_{c} = \varepsilon_{c}^{-K+1}
\]

where \( \varepsilon_{c} = m_{c} / n \), where \( m_{c} = \sum_{i=1}^{k} \sum_{j_{1}=1}^{n_{i}} I \{ \tilde{X}_{i} = C \} \).

An argument parallel to that of Chapter II, Section 5, can be used to show that the distribution of \( \tilde{G} \) has the same limiting power as \( G^{*} \). In this case, \( \sqrt{\sum_{c} n_{c}} (\varepsilon_{c}^{-K+1} - \varepsilon_{c}^{-K+1}) \) and \( \sum_{c} n_{c} \) converges to zero in law.

H. Chapter Summary and Concluding Remarks

This chapter has presented a method of matched ANOCOVA for the \( K \)-sample problem, called the All Matched \( K \)-tuples Sign Test. The test statistic is

\[
G^{*} = d_{0}^{1} \sum_{d_{0}}^{-1} d_{0}
\]

where
\[ d_0 = \sqrt{n} (U_1 - \bar{U}, \ldots, U_{K-1} - \bar{U}), \]

\[ U_i = V_i / \prod_{j_i} n_i, \]

\[ V_i = \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \cdots \sum_{j_k=1}^{n_k} g(i)(Y_{1j_1}, Y_{2j_2}, \ldots, Y_{kj_k}) \times \]

\[ \sum_{c} e^{-K+1} I\{X_{1j_1} = X_{2j_2} = \cdots = X_{kj_k}\} \]

\[ g(i)(Y_{1j_1}, \ldots, Y_{kj_k}) = \begin{cases} 
1 & \text{if } Y_{ij_i} < Y_{ij_i}^{*}, \text{ for all } i \neq i' \\
0 & \text{else},
\end{cases} \]

and \( \Sigma_0 \) is the corresponding covariance matrix, given by \((\sigma_{rs})\).

\[ \sigma_{rr} = \frac{1}{2K-1} \left[ a_r^{-1} + \frac{2}{K} \sum_{i=1}^{k} a_i^{-1} \right] - \left( \frac{1}{K} \right)^2 \sum_{i=1}^{k} a_i^{-1}, \quad r=1,2,\ldots,K-1 \]

\[ \sigma_{rs} = \frac{1}{K(2K-1)} \left[ a_r^{-1} + a_s^{-1} + 2 \sum_{i=1}^{k} a_i^{-1} \right] - \left( \frac{1}{K} \right)^2 \sum_{i=1}^{k} a_i^{-1}, \quad r,s=1,2,\ldots,K-1; \]

\[ i \neq r, s. \]

\( G^* \) is asymptotically distributed as a chi-squared random variable under the null hypothesis and translation-type alternatives. The noncentrality parameter under a sequence of alternatives

\[ H_n: F_i(y|x) = F(y-n^{-\frac{1}{2}} \theta_i|x) , \quad i=1,2,\ldots,K \]

is given by

\[ \lambda_{G^*} = \delta_0 \sum_{i=1}^{K} \delta_0 \left[ \int_{-\infty}^{\infty} [1-F(y|x)]^{k-2} f^2(y|x) dy|x \right]^2, \]

where
\[ \delta_0' = (\delta_1', \delta_2', \ldots, \delta_{K-1}') \]

for \( \delta_i = K\bar{\theta}_i - \sum_{i'=1}^k \bar{\theta}_{i'} \).

In the special case where the sample sizes are all equal, then \( G^* \) is simplified to

\[ G^* = \frac{n(2K-1)}{K} \sum_{i=1}^k (U_i - \bar{U})^2, \]

and the noncentrality parameter is

\[ \lambda_{G^*} = K (2K-1) \sum_{i=1}^k (\bar{\theta}_i - \bar{\theta})^2 \left[ \int_{-\infty}^{\infty} \left[ 1 - F(y|x) \right]^{K-2} \frac{r^2(y|x)}{y|x} dy|x \right]^2. \]

In the general case (unequal sample sizes) a simplified expression is not derived due to difficulties in inverting the covariance matrix. However, given data, clearly the matrix can be inverted (perhaps using numerical methods), and the statistic calculated. The noncentrality parameter can also be found explicitly.

Although the demonstration is postponed until Chapter IV, it may be noted here that the ARE of the AMK sign test relative to Bhapkar's V-test is unity for the equal sample size case, whenever the covariable is independent of the response (i.e., when Bhapkar's V-test is appropriate). In fact, it is easily demonstrated that in the trivial case where all \( K \)-tuples are matched (\( C=1 \Rightarrow e_3 \equiv 0 \)), the AMK test is identically Bhapkar's V-test.

Examples and efficiency results are reported in Chapter IV.
CHAPTER IV

EFFICIENCY CONSIDERATIONS AND EXAMPLES

A. Efficiency Considerations Comparing the AMP Sign Statistic with the Kruskal Wallis Test for the Special Case \( n_1=\ldots=n_K \).

1. ARE of the AMP Sign Statistic to the Kruskal-Wallis Test:

It was demonstrated in Chapter II, Section F.4, that under translation-type alternatives, \( D^* \) has a limiting noncentral chi-squared distribution with \( K-1 \) degrees of freedom and noncentrality parameter

\[
\lambda_{D^*} = 12 \frac{1}{K} \sum_{i=1}^{K} (\theta_i - \bar{\theta})^2 \left\{ \int_{-\infty}^{\infty} F'(y|x) \, dF(y|x) \right\}^2
\]

Andrews (1954) has shown that the Kruskal-Wallis H-test is (under translation-type alternatives) distributed as a noncentral chi-squared random variable with \( K-1 \) degrees of freedom and noncentrality parameter (assuming equal sample sizes)

\[
\lambda_H = 12 \left\{ \int_{-\infty}^{\infty} F'(y) \, dF(y) \right\}^2 \frac{1}{K} \sum_{i=1}^{K} (\theta_i - \bar{\theta})^2
\]

A well-known result [Andrews (1954), Hannan (1956)] is that the asymptotic efficiency of one statistic to another (when their limiting distributions are chi-squared with the same number of degrees of freedom) is equal to the ratio of their noncentrality parameters.
Hence, it follows that the ARE (D*, H) is as follows:

\[
\text{ARE(D*, H)} = \frac{\lambda_{D*}}{\lambda_H} = \frac{12 \cdot \frac{1}{K} \sum_{i=1}^{k} (\theta_i - \bar{\theta})^2 \left\{ \int_{-\infty}^{\infty} F'(y|x) \, dF(y|x) \right\}^2}{12 \left\{ \int_{-\infty}^{\infty} F'(y) \, dF(y) \right\}^2 \cdot \frac{1}{K} \sum_{i=1}^{k} (\theta_i - \bar{\theta})^2}
\]

It follows that, if the covariable is independent of the response, then the ARE(D*, H) = 1. For large n, then, no loss in efficiency results by controlling for an irrelevant covariable. However, in finite samples some loss in efficiency may be expected.

2. Example 1:

The following illustrative example is presented.

<table>
<thead>
<tr>
<th>Sample I</th>
<th>Sample II</th>
<th>Sample III</th>
</tr>
</thead>
<tbody>
<tr>
<td>X = 1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>X = 2</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>X = 3</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>29</td>
</tr>
</tbody>
</table>
\[ \hat{e}_{-1}^{1} = \frac{24}{7} \]
\[ \hat{e}_{-1}^{2} = \frac{24}{8} \]
\[ \hat{e}_{-1}^{3} = \frac{24}{9} \]

\[ W_{1} = \frac{24}{7} (64) + \frac{24}{8} (120) + \frac{24}{9} (128) = 920.7619 \]
\[ W_{2} = \frac{24}{7} (16) + \frac{24}{8} (0) + \frac{24}{16} (16) = 97.5238 \]
\[ W_{3} = \frac{24}{9} (-80) + \frac{24}{8} (-120) + \frac{24}{9} (-144) = 957.3333 \]

\[ T_{1} = 1.7984 \]
\[ T_{2} = 0.1905 \]
\[ T_{3} = -1.8698 \]
\[ \bar{T} = 0.0397 \]

\[ D^{*} = \frac{3(24)}{27} [3.0929 + 0.0227 + 3.6462] \]
\[ = 2.67 [6.7618] \]
\[ D^{*} = 18.0310 \]

The Kruskal-Wallis Test

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>16</td>
<td>18</td>
<td>22</td>
</tr>
<tr>
<td>17</td>
<td>20</td>
<td>23</td>
</tr>
<tr>
<td>19</td>
<td>21</td>
<td>24</td>
</tr>
</tbody>
</table>

\[ R_{1}=83 \quad R_{2}=93 \quad R_{3}=124 \]
\[
H = \frac{12}{(24)(25)} \left[ \frac{83^2}{8} + \frac{93^2}{8} + \frac{124^2}{8} \right] - 3(25)
= \frac{1}{50} [3864.25] - 75
= 2.285
\]

Example 1 illustrates that for this case \((n_i=8, K=3)\) where the data indicate a location difference, the AMP sign test for \(K\)-samples detected the difference (rather dramatically) while the Kruskal-Wallis test did not.

3. Example 2:

Another illustrative example is presented.

<table>
<thead>
<tr>
<th>Sample I</th>
<th>Sample II</th>
<th>Sample III</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X = 1)</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>(X = 2)</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>(X = 3)</td>
<td>22</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>27</td>
</tr>
</tbody>
</table>

\[
W_1 = \frac{24}{7}(0) + \frac{24}{8}(-8) + \frac{24}{9}(-16) = -66.6667
\]

\[
W_2 = \frac{24}{7}(-16) + \frac{24}{8}(16) + \frac{24}{9}(16) = 35.8095
\]

\[
W_3 = \frac{24}{7}(16) + \frac{24}{8}(-8) + \frac{24}{9}(0) = 30.8571
\]
\[ T_1 = -0.1302 \]
\[ T_2 = 0.0699 \]
\[ T_3 = 0.0603 \]
\[ \bar{T} = 0.0000 \]

\[ D^* = \frac{3(24)}{27} [0.0169 + 0.00489 + 0.0036] \]
\[ = 2.667 [0.02547] \]

\[ \hat{D}^* = 0.0679 \]

B. Efficiency Considerations Comparing the AMK Sign Statistic with Bhapkar's V-test for the Special Case \( n_1 = n_2 = \cdots = n_K \):

1. ARE of AMK Sign Test to Bhapkar's V-test:

It was demonstrated in Chapter III, Section G.4, that under translation-type alternatives, \( G^* \) has a limiting noncentral chi-squared distribution with \( K-1 \) degrees of freedom and noncentrality parameter

\[ \lambda_{G^*} = K (2K-1) \sum_{i=1}^{k} (\theta_i - \bar{\theta})^2 \left[ \int_{-\infty}^{\infty} [1-F(y|x)]^{k-2} f^2(y|x) dy \right]^2. \]

Bhapkar (1961) shows that the V-test, under translation-type alternatives, is asymptotically distributed as a chi-squared with \( K-1 \) degrees of freedom, and noncentrality parameter (with \( n_1 = \cdots = n_K \))

\[ \lambda_V = K (2K-1) \sum_{i=1}^{k} (\theta_i - \bar{\theta})^2 \left[ \int_{-\infty}^{\infty} [1-F(y)]^{k-2} f^2(y) dy \right]^2. \]
Hence the $\text{ARE}(G^*,V)$ is as follows:

$$
\text{ARE}(G^*,V) = \frac{\lambda_{G^*}}{\lambda_V}
$$

$$
= \frac{K (2K-1) \sum_{i=1}^{k} (\theta_i - \bar{\theta})^2 \left[ \int_{-\infty}^{\infty} [1-F(y|x)]^{K-2} f^2(y|x) dy \right]^{2}}{K (2K-1) \sum_{i=1}^{k} (\theta_i - \bar{\theta})^2 \left[ \int_{-\infty}^{\infty} [1-F(y)]^{K-2} f^2(y) dy \right]^{2}}
$$

$$
= \left[ \frac{\int_{-\infty}^{\infty} [1-F(y|x)]^{K-2} f^2(y|x) dy}{\int_{-\infty}^{\infty} [1-F(y)]^{K-2} f^2(y) dy} \right]^2
$$

It follows that, if the covariable is independent of the response, then $\text{ARE}(G^*,V) = 1$. So if the AMK sign test is performed when Bhapkar's test is appropriate (i.e., when a covariable need not be controlled for, but yet is incorporated into the analysis) asymptotically, nothing is lost. It should be noted, however, that this is an asymptotic result, and that in finite samples, some loss of efficiency may be expected if an irrelevant covariable is incorporated into the analysis. In general, if a covariable should be incorporated into the analysis, but is not, Bhapkar's test may not detect a true population difference, or may depict a false difference.

2. Example 1:

Refer to the data of Example 1 of Section A.

\text{AMK Sign Test:} \quad n_1 = n_2 = n_3 = 8; \ n = 24

$$
\hat{\varepsilon}_1 = \frac{7}{24} \quad \hat{W}_1 = \left(\frac{7}{24}\right)^{-2} = \frac{(24)^2}{49}
$$
\[ \hat{c}_2 = \frac{8}{24} \quad \hat{w}_2 = (\frac{8}{24})^{-2} = 9 \]
\[ \hat{c}_3 = \frac{9}{24} \quad \hat{w}_3 = (\frac{9}{24})^{-2} = \frac{64}{9} \]

\[ V_1 = \frac{(24)^2}{49} \cdot 10 + 9.18 + \frac{64}{9} \cdot 24 = 450.21768 \]
\[ V_2 = \frac{(24)^2}{49} \cdot 2 + 0 + \frac{64}{9} \cdot 3 = 44.843537 \]
\[ V_3 = 0 \]

\[ (\Pi \binom{n_i}{i}) = 512 \]

\[ U_1 = 0.879314 \]
\[ U_2 = 0.087585 \]
\[ U_3 = 0 \]

\[ \bar{\Sigma}U_1 = 0.966899, \quad \bar{U} = 0.3222996 \]

\[ \tilde{G}^* = \frac{n(2K-1)}{K} \sum_{i=1}^{k} (U_i - \bar{U})^2 = \frac{(24)(5)}{3} \left[ 0.3102649 + 0.0550909 + 0.10387 \right] \]
\[ \tilde{G}^* = 18.7693 \]

Compare with \( \chi^2_{(2)} \).

**Bhapkar's V-test:**

\[ V_1 = 234 \quad U_1 = 0.4570313 \]
\[ V_2 = 191 \quad U_2 = 0.3730469 \]
\[ V_3 = 87 \quad U_3 = 0.1699219 \]
\[ V = \frac{n(2k-1)}{k} \sum_{i=1}^{k} (u_i \cdot \frac{1}{k})^2 \]

\[ = \frac{(24)(5)}{3} [0.0153012 + 0.0015772 + 0.0267033] \]

\[ = 40[0.0435817] \]

\[ V = 1.7433 \]

Compare with \( X^2_{(2)} \).

Example 1 is an extreme case where a translation difference exists, and the covariable is correlated with the response. Bhapkar's V-test failed to detect any difference, while the AMK test shows a dramatic result.

In general, the ARE(G*, V) depends on the relationship between the covariable and the response.

3. Example 2:

Refer to Example 2 of Section A.

\[ V_1 = 70.530612 + 36 + 56.\bar{8} = 163.4195 \]

\[ V_2 = 23.510204 + 63 + 78.\bar{2} = 164.73242 \]

\[ V_3 = 47.020408 + 63 + 56.\bar{8} = 166.90929 \]

\[ U_1 = 0.3191787 \]

\[ U_2 = 0.321743 \]

\[ U_3 = 0.3259947 \]

\[ \Sigma u_i = 0.9669164, \quad \bar{U} = 0.3223054 \]
\[ G^* = 40[0.0000097 + 0.0000003 + 0.0000136] = 0.0009 \]

Compare with \( X^2(2) \).

C. Comments on "Handpicked" Examples and Example with Real Data

Examples 1 and 2 illustrate how the AMP and AMK sign tests are performed, and also that, (for \( n_i = 8 \) at least), if the data are extreme indicating a translation difference (when matching on \( \tilde{z} \)), the statistics tend to be large; if the data are extreme indicating no translation difference, the statistics tend to be close to zero.

An Example with Real Data:

The following unpublished data are a subsample of a large study with three treatment groups. The covariable is univariate continuous; however, the measuring process converts it into a discrete measure. The response is univariate continuous. The subsample is selected with \( n_1 = n_2 = n_3 = 9 \).

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>The Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment 1</td>
<td>Treatment 2</td>
</tr>
<tr>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>4</td>
<td>0.20</td>
</tr>
<tr>
<td>3</td>
<td>0.31</td>
</tr>
<tr>
<td>5</td>
<td>0.79</td>
</tr>
<tr>
<td>5</td>
<td>0.93</td>
</tr>
<tr>
<td>6</td>
<td>0.17</td>
</tr>
<tr>
<td>10</td>
<td>2.11</td>
</tr>
<tr>
<td>10</td>
<td>1.00</td>
</tr>
<tr>
<td>25</td>
<td>1.60</td>
</tr>
<tr>
<td>25</td>
<td>1.43</td>
</tr>
</tbody>
</table>
TABLE 2
Data Displayed by X

<table>
<thead>
<tr>
<th></th>
<th>Rx 1</th>
<th>Rx 2</th>
<th>Rx 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>X = 3,4</td>
<td>0.20</td>
<td>0.33</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>0.31</td>
<td>0.54</td>
<td>2.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.00</td>
</tr>
<tr>
<td>X = 5,6</td>
<td>0.17</td>
<td>1.43</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>0.79</td>
<td>1.76</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>0.93</td>
<td>2.38</td>
<td></td>
</tr>
<tr>
<td>X = 10</td>
<td>1.00</td>
<td>1.88</td>
<td>1.43</td>
</tr>
<tr>
<td></td>
<td>2.11</td>
<td>3.164</td>
<td>1.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.162</td>
</tr>
<tr>
<td>X ≥ 20</td>
<td>1.43</td>
<td>2.14</td>
<td>2.31</td>
</tr>
<tr>
<td></td>
<td>1.60</td>
<td>2.14</td>
<td></td>
</tr>
</tbody>
</table>

AMP Sign Test for K-samples:

\[ \hat{W}_1 = \frac{27}{7} \]
\[ \hat{W}_2 = \frac{27}{8} \]
\[ \hat{W}_3 = \frac{27}{7} \]
\[ \hat{W}_4 = \frac{27}{5} \]

\[ W_1 = \frac{27}{7}(90) + \frac{27}{8}(45) + \frac{27}{7}(36) + \frac{27}{5}(54) = 929.475 \]
\[ W_2 = \frac{27}{7}(18) + \frac{27}{8}(-135) + \frac{27}{7}(-54) + \frac{27}{5}(-18) = -691.6821 \]
\[ W_3 = \frac{27}{7}(-108) + \frac{28}{8}(90) + \frac{27}{7}(18) + \frac{27}{5}(-36) = -226.5429 \]
\[ T_1 = 1.2750 \]
\[ T_2 = -0.9488 \]
\[ T_3 = -0.3108 \]
\[ T = 0.0051 \]

\[ \tilde{D}^* = \frac{3(27)}{(27)} [1.6125 + 0.9100 + 0.0998] \]

\[ \tilde{D}^* = 7.8669 \]

AMK Sign Test: \[ n_1 = n_2 = n_3 = 9, \ n = 27 \]

\[ \hat{\xi}_1 = \frac{7}{27} \quad \hat{W}_1 = \left(\frac{7}{27}\right)^{-2} \]

\[ \hat{\xi}_2 = \frac{8}{27} \quad \hat{W}_2 = \left(\frac{8}{27}\right)^{-2} \]

\[ \hat{\xi}_3 = \frac{7}{27} \quad \hat{W}_3 = \left(\frac{7}{27}\right)^{-2} \]

\[ \hat{\xi}_4 = \frac{5}{27} \quad \hat{W}_4 = \left(\frac{5}{27}\right)^{-2} \]

\[ V_1 = \frac{27^2}{49}(12) + \frac{27^2}{64}(3) + \frac{27^2}{49}(7) + \frac{27^2}{25}(4) = 433.48534 \]

\[ V_2 = 0 + 0 + \frac{27^2}{49}(1) + 0 = 14.877551 \]

\[ V_3 = 0 + \frac{27^2}{64}(15) + \frac{27^2}{49}(4) + 0 = 230.3696 \]

\[ \prod_{i=1}^{n_1} = 729 \]

\[ U_1 = 0.5946 \]

\[ U_2 = 0.0204 \]
\[ U_3 = 0.3160 \]

\[ \Sigma U_1 = 0.9310, \quad \bar{U} = 0.3103 \]

\[ \tilde{G}^* = \frac{(27)(5)}{3} [0.0808 + 0.0841 + 0.0104] = 7.888 \]

D. Efficiency Comparisons Between AMK Sign Statistic and AMP Sign Statistic for K-samples for the Special Case: \[ n_1 = n_2 = \ldots = n_K \]

Given the two new methods of matched analysis of covariance advanced in this dissertation, it is desirable to compare them to each other with respect to efficiency. It is stressed that the efficiency results presented in this chapter are asymptotic results, and in finite samples of reasonable size, the relative efficiencies may differ. However, Pitman ARE is used in this section to address the question: Which test is better?

The noncentrality parameters are:

\[ \lambda_{G^*} = K (2K-1) \sum_{i=1}^{k} (\theta_i - \bar{\theta})^2 \left[ \int_{-\infty}^{\infty} [1-F(y|x)]^{K-2} F'(y|x) dF(y|x) \right]^2 \]

\[ \lambda_{D^*} = 12 \cdot \frac{1}{K} \sum_{i=1}^{k} (\theta_i - \bar{\theta})^2 \left[ \int_{-\infty}^{\infty} F'(y|x) dF(y|x) \right]^2 \]

Hence,

\[ \text{ARE}(D^*, G^*) = \lambda_{D^*}/\lambda_{G^*} = \frac{12 \cdot \frac{1}{K} \sum_{i=1}^{k} (\theta_i - \bar{\theta})^2 \left[ \int_{-\infty}^{\infty} F'(y|x) dF(y|x) \right]^2}{K (2K-1) \sum_{i=1}^{k} (\theta_i - \bar{\theta})^2 \left[ \int_{-\infty}^{\infty} [1-F(y|x)]^{K-2} F'(y|x) dF(y|x) \right]^2} \]
\[
\text{ARE}(D^*, G^*) = \frac{12}{K^2(2K-1)} \cdot \left\{ \frac{\int_{-\infty}^{\infty} F'(y|x) \, dF(y|x)}{\left[ 1-F(y|x) \right]^{K-2} \int_{-\infty}^{\infty} F'(y|x) \, dF(y|x)} \right\}^2
\]

Note that \( [1-F(y|x)]^{K-2} \leq 1 \), hence, \( \text{ARE}(D^*, G^*) \geq 12/[K^2(2K-1)] \).

Consider special cases:

a. \( K = 2 \):

\[
\text{ARE}(D^*, G^*) = \frac{12}{(4)(3)} \left\{ \frac{\int_{-\infty}^{\infty} F'(y|x) \, dF(y|x)}{\int_{-\infty}^{\infty} F'(y|x) \, dF(y|x)} \right\}^2
\]

\[= 1, \text{ regardless of } F(y|x).\]

This has to be the case, since both the AMP test for K-samples and the AMK test reduce to Schoenfelder's AMP sign test when \( K=2 \).

b. \( F(y|x) \sim \text{Uniform} \):

\[
\text{ARE}(D^*, G^*) = \frac{12}{K^2(2K-1)} \left\{ \frac{1}{\int_{0}^{\infty} (1-x)^{K-2} \, dx} \right\}^2
\]

\[= \frac{12}{K^2(2K-1)} \left\{ \frac{1}{\frac{1}{K-1}} \right\}^2
\]

\[= \frac{12(K-1)^2}{K^2(2K-1)}.\]
Clearly, this is the ARE(D*,G*) if F(y|x) is uniform (for any scale parameter).

c. \( F(y|x) \sim \text{Exponential} \):

\[
\text{ARE}(D^*, G^*) = \frac{12}{K^2(2K-1)} \left\{ \frac{\int_{0}^{\infty} e^{-2(t-\theta)} \, dt}{\int_{0}^{\infty} [1-(1-e^{-(t-\theta)})]^{K-2} e^{-2(t-\theta)} \, dt} \right\}^2
\]

\[
= \frac{12}{K^2(2K-1)} \left\{ \frac{\int_{0}^{\infty} e^{-2(t-\theta)} \, dt}{\int_{0}^{\infty} e^{-K(t-\theta)} \, dt} \right\}^2
\]

\[
= \frac{12}{K^2(2K-1)} \left( \frac{K}{2} \right)^2
\]

\[
= \frac{3}{2K-1}
\]

d. \( F(y|x) \sim \text{Pareto}(1,1) \):

\[
\text{ARE}(D^*, G^*) = \frac{12}{K^2(K+1)} \left\{ \frac{\int_{1}^{\infty} (t^2)^2 \, dt}{\int_{1}^{\infty} [1-(1-\frac{1}{t})]^{K-2} (t^2)^2 \, dt} \right\}^2
\]

\[
= \frac{12}{K^2(2K-1)} \left\{ \frac{1/3}{1/(K+1)} \right\}^2
\]

\[
= \frac{4(K+1)^2}{3K^2(2K-1)}
\]
e. \( F(y|x) \sim \text{Beta}(r, 1) \):

\[
\text{ARE}(D^*, G^*) = \frac{12}{K^2(2K-1)} \left[ \frac{\int_0^1 r^2 t^{(2r-2)} dt}{\int_0^1 (1-t^r)^K 2 \, r^2 t^{(2r-2)} dt} \right]^2
\]

\[
= \frac{12}{K^2(2K-1)} \left\{ \frac{r^2 \cdot \frac{1}{2r-1}}{B(K, 1, 2 - \frac{1}{r})} \right\}
\]

where \( B(\cdot, \cdot) \) indicates the beta function

\[
= \frac{12 \, r^2}{K^2(2K-1)(2r-1)^2[B(K-1, 2 - \frac{1}{r})]^2}
\]

For large \( r \), \( \text{ARE}(D^*, G^*) \to \frac{12 \, r^2}{K^2(2K-1)(2r-1)^2[B(K-1, 2)]^2} = \frac{3(K-1)^2}{2K-1} \).

\[
\cdot \frac{4 \, r^2}{(2r-1)^2} = \sim \frac{3(K-1)}{(2K-1)}.
\]

(Note that case (b) is a special case of the above.)

f. \( F(y|x) \sim \text{Double Exponential} \):

\[
\text{ARE}(D^*, G^*) = \frac{12}{K^2(2K-1)} \left\{ \begin{array}{c}
\int_\theta^\infty \frac{1}{4} e^{2(t-\theta)} dt + \int_0^\infty \frac{1}{4} e^{-2(t-\theta)} dt \\
\int_{-\infty}^\theta \left[ \frac{1}{2} e^{(t-\theta)} \right] K^2 - \frac{1}{4} e^{2(t-\theta)} dt + \int_\theta^\infty \left[ 1 - \left\{ \frac{1}{2} e^{(t-\theta)} \right\} \right] K^2 - \frac{1}{4} e^{-2(t-\theta)} dt
\end{array} \right\}
\]
\[
\frac{12}{K^2(2K-1)} \left\{ \frac{1}{4} \cdot \left[ \frac{1}{2} e^{2(t-\theta)} \right]_0^\infty - \frac{1}{2} e^{-2(t-\theta)} \right\} \left( \frac{1}{4} \int_0^\infty e^{2(t-\theta)} \left[ 1 - \frac{1}{2} e^{(t-\theta)} \right]^{K-2} dt \right.
\]
\[
+ \left. \int_0^\infty \frac{1}{2K} e^{-(t-\theta)K} dt \right\} ^{118}
\]

\[
\frac{1}{4} \int_0^\infty e^{2(t-\theta)} \left[ 1 - \frac{1}{2} e^{(t-\theta)} \right]^{K-2} dt = \frac{1}{K2^K} - \frac{1}{(K-1)2^{K-1}} + \frac{1}{K-1} - \frac{1}{K}
\]

\[
\int_0^\infty \frac{1}{2K} e^{-(t-\theta)K} dt = \frac{1}{K2^K}
\]

\[
= \frac{12}{K^2(2K-1)} \left\{ \frac{1}{4} \left[ \frac{1}{K(K-1)} \left[ 1 - \frac{1}{2^{K-1}} \right] \right] \right\}
\]

\[
= \frac{3(K-1)^2}{2K-4} \frac{2^{K-4}}{(2K-1)(2^{K-1}-1)^2}
\]

For the normal distribution, the ARE's can be computed from Table 1 given by Hojo (1931). The results are as follows:

<table>
<thead>
<tr>
<th>K</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>10</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARE(D*,G*)</td>
<td>1.06</td>
<td>1.16</td>
<td>1.25</td>
<td>1.35</td>
<td>1.45</td>
<td>1.54</td>
<td>1.72</td>
<td>1.89</td>
</tr>
</tbody>
</table>

Then, if \( F(y|x) \) is normal, then the preferable test is \( D^* \) for \( K \geq 3 \), since \( ARE(D^*,G^*) \geq 1 \) for \( K \geq 3 \). A summary table follows.
| F(y|x)       | 2  | 3  | 4  | 5  | 6  | 10 |
|-------------|----|----|----|----|----|----|
| Uniform     | 1.00 | 1.07 | 0.96 | 0.85 | 0.76 | 0.51 |
| Exponential | 1.00 | 0.60 | 0.43 | 0.33 | 0.27 | 0.16 |
| Double Exp. | 1.00 | 1.07 | 1.26 | 1.52 | 1.82 | 3.20 |
| Beta (2,1)  | 1.00 | 1.67 | 2.05 | 2.30 | -   | -   |
| Normal      | 1.00 | 1.06 | 1.16 | 1.25 | 1.35 | 1.72 |
| Pareto (1,1)| 1.00 | 0.47 | 0.30 | 0.21 | 0.16 | 0.08 |
| Lower Bound | 1.00 | 0.27 | 0.11 | 0.05 | 0.03 | 0.01 |

Based on the efficiency results reported for these particular conditional distributions, it appears that for distributions that are bounded below, the AMK test has better asymptotic efficiency. For those not bounded below, the AMP test tends to perform better. Which test is preferable relative to efficiency really depends on the conditional distribution. Bhapkar (1961) notes that for the normal distribution, the ARE of the V-test relative to the Kruskal-Wallis test tends to zero as the number of populations tends to infinity. It follows that for the normal distribution, the ARE of the AMK test relative to the AMP test for k-samples also tends to zero as the number of populations tends to infinity, since the \( \text{ARE}(D^*, G^*) \) is the same as the ARE of the Kruskal-Wallis test relative to Bhapkar's V-test.
CHAPTER V

COMMENTS AND SUGGESTIONS FOR FUTURE RESEARCH

This dissertation advances two new analysis techniques for the K-sample homogeneity problem in the presence of a discrete covariate. The method presented in Chapter II combines the features of Schoenfelder's AMP sign test and the Kruskal-Wallis test. Chapter III presents a method extending Schoenfelder's AMP test and Bhapkar's V-test. In each of these chapters, the test statistic is presented for the general K-sample framework as a quadratic form. The asymptotic distribution of the quadratic form is, in each case, derived under the null and translation-type alternative hypotheses, and an expression for the noncentrality parameter is provided. For the special case where the sample sizes are equal, the quadratic form is expanded and simplified yielding a well-defined explicit test statistic. Also for the special case, the noncentrality parameter is simplified. ARE result are provided in Chapter IV.

A. Comments on the AMP Sign Test for K-Samples

Quade (1982) suggests a method called rank analysis of covariance by matching. Assuming the framework of this dissertation, the test may be defined as follows:
Let

$$S_{ij} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \text{sgn}(Y_{ij} - Y_{i',j'}) I\{X_{ij} = X_{i',j'}\} / M_{ij}$$

where $M_{ij}$ is the number of observations matched with $S_{ij}$. (If $X_{ij} = V_c$, then $M_{ij} = M_c$. ) An analysis of variance can be performed on the $S_{ij}$'s. Quade shows that this test is a generalization of the Kruskal-Wallis test. Clearly, the weights are Schoenfelder-type weights and, in fact, it can be shown that analysis of variance on these scores is equivalent to the AMP sign test for $K$-samples. The results in this dissertation support the intuition presented in Quade's paper (for discrete $X$ with zero tolerance), and provide efficiency results for a special case.

Quade's rank ANOCOVA by matching is defined for a more general framework. The only restriction on $X$ is that it be concomitant. A matching function is imposed on $X$ which defines two observations to be matched if their covariables assume values that are within a given tolerance. This generalization can be pursued for the AMP sign test. With the intuition and framework provided by Quade, it might be possible to use the methods of Chapter II to verify the asymptotic distribution and derive efficiency results of such a test.

Puri (1964) proposed a class of test statistics which generalizes the Kruskal-Wallis test, among others. The test statistics are of the form

$$L = \sum_{i=1}^{k} n_i \frac{[(T_{n,i} - \nu_{n,i}) / A_n]^2}{A_n}$$

where $\nu_{n,i}$ and $A_n$ are normalizing constants and

$$T_{n,i} = n_i^{-1} \sum_{j=1}^{n} E_{n,j} z_{n,j}^{(i)}$$
where
\[ z_{n,j}^{(i)} = \begin{cases} 1 & \text{if the } j\text{th smallest observation is from sample } i \\ 0 & \text{else.} \end{cases} \]

It appears that a generalization of the AMP sign test for K-samples based on Puri's general class of statistics would be possible. As with the AMP sign test, the weights (that are functions of the \( n_i \)'s only) are fixed across categories. Hence, consider

\[
L_{c} = \sum_{i=1}^{k} n_i [(T_{n,i} - \mu_{n,i})/\sigma_{n,i}]^2 \mathbb{I}\{X_{ij} = V_{c}\}
\]

and \( L^* = \sum_{c} W_c L_c \).

It should be possible to show that \( L^* \) is asymptotically chi-squared with \( K-1 \) degrees of freedom under the hypothesis of homogeneity; however, this has not been investigated.

B. Comments on the AMK Sign Test

Deshpande (1970) presented a unified treatment of Bhapkar's V-test, the L-test, the W-test, and other variations of ANOVA problems based on the analysis of K-tuples. As in Chapter I, let \( V_{ij} \) be the number of K-tuples in which the observation from sample \( i \) has rank \( j \), and let \( U_{ij} = V_{ij}/\Pi n_i \). Then let

\[
L_i = \sum_{j=1}^{k} f_i U_{ij},
\]

where the \( f_i \)'s are real constants such that they are not all equal, and
\[
F = \sum_{j=1}^{k} \sum_{\ell=1}^{k} f_j f_\ell \left[ \frac{(K-1)!}{(j-1)! (\ell-1)!} \frac{1}{(2K-1)(2K-2)} \right].
\]

Then define

\[
L = \frac{n(K-1)^2}{FK^2} \left[ \sum_{i=1}^{k} a_i L_i^2 - \left( \sum_{i=1}^{k} a_i L_i \right)^2 \right].
\]

It is easy to see that if \( f_1 = 1, f_j = 0 \) for \( j \neq 1 \), the result is Bhapkar's \( V \) statistic. If \( f_1 = -1, f_K = 1, f_j = 0 \) for \( j \neq 1, K \), the result is the \( L \)-test. If \( f_j = j-1 \), the result is the \( W \)-test. The AMK sign test can obviously be extended to a more general form in this way. Let \( V_{ij}^{(c)} \) be the number of \( K \)-tuples for which the observation from sample \( i \) has rank \( j \), calculated only for those observations whose covariable assumes the value \( V_c \). Then let

\[
V_{ij} = \sum_c \varepsilon_c^{K+1} V_{ij}^{(c)}, \quad (\varepsilon_c = \mathbb{P}(X=V_c) > 0)
\]

\[
U_{ij} = V_{ij}/n_i,
\]

\[
L_i^* = \sum_{j=1}^{k} f_j U_{ij},
\]

with \( f_i \) and \( F \) as defined above. Then

\[
L^* = \frac{n(K-1)^2}{FK^2} \left[ \sum_{i=1}^{k} a_i (L_i^*)^2 - \left( \sum_{i=1}^{k} a_i L_i^* \right)^2 \right]
\]

defines a class of test statistics for ANOCOVA by matching. It should be possible (by using the methods of Chapter III and those of Deshpande (1970)) to show that members of this class have null
distributions which are chi-squared (K-1), and noncentral chi-squared distributions under translation-type alternative hypotheses.

C. General Comments and Suggestions for Future Research

There are many obvious directions future research can take. A few of these are now listed.

1. Continuous covariable:

Both of the methods presented assume a discrete covariable. The extension to continuous X would certainly prove to be a useful analysis tool. The next step from what is developed in this research would be to a discrete X defining a matching function, where two observations are matched if their covariables assume values within a specified tolerance. This would isolate the issue of coping with "overlapping" categories. Next in the process of generalizing to continuous X would be to let X be continuous and define the matching function with fixed, positive tolerance.

2. Discrete response:

The framework assumed in this research specifies that the response variable is continuous. The assumption was made so that ties on Y might be ignored. An adjustment to the significance level for approximating a discrete distribution with a continuous one is made frequently. The impact of this can be assessed by assuming tied values at certain levels of the covariable, breaking the tie(s) in all possible ways, and comparing the results. Another possibility might be to derive an
analogue of the "midrank" method commonly used for the Kruskal-Wallis test and other rank procedures.

Future results in many other respects are desirable. The mathematical inversion of $\Sigma_0$ could be pursued, methods accommodating a multivariate response, ordered alternatives, multiple comparisons procedures, and methods of point and interval estimation of effects (assuming shift alternatives) are other suggestions for future research.
REFERENCES


Hojo, T (1931), "Distribution of the Median, Quartiles, and Inter-quartile Distance in Samples from a Normal Population," Biometrika 23: 315-360.


Kupper, KL; Karon, JM; Kleinbaum, DG; Lewis, DK; Morgenstern, M (1979), "Matching in Epidemiologic Studies: Validity and Efficiency Considerations," Institute of Statistics Mimeo Series No. 1239, UNC.

Lehmann, EL (1975), Non-Parametrics: Statistical Methods Based on Ranks. Holden-Day.


Mann, HB and Whitney, DR (1957), "On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other," Annals of Mathematical Statistics 18: 50-60.


(1969), "Individual Matching with Multiple Controls in the Case of All-or-None Responses," Biometrics 25(2): 339-354.


Slutsky, E (1925), 'Uber Stochastiche Asymptoten und Grenzwerte, m, 5, No. 3, p. 3.


Walter, SD (1979), 'Matched Case-Control Studies with a Variable Number of Controls Per Case,' Presented at ENAR Biometrics Meetings, New Orleans, April 1979.

