This research was supported by the National Institutes of Health Grant No. T 01GM00038 from the National Institute of General Medical Sciences.

A STUDY OF THE SMALL SAMPLE PROPERTIES OF TESTS OF LINEAR HYPOTHESES FOR CATEGORIZED ORDINAL RESPONSE DATA

By

Noel Mohberg

Department of Biostatistics
University of North Carolina
Chapel Hill, N. C.

Institute of Statistics Mimeo Series No. 815

APRIL 1972
A STUDY OF THE SMALL SAMPLE PROPERTIES
OF TESTS OF LINEAR HYPOTHESES FOR
CATEGORIZED ORDINAL RESPONSE DATA

by

NOEL MOHBerg
A STUDY OF THE SMALL SAMPLE PROPERTIES
OF TESTS OF LINEAR HYPOTHESES FOR
CATEGORIZED ORDINAL RESPONSE DATA

by

Noel Mohberg

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

Chapel Hill
1972

Approved by:

Adviser
Reader
Reader
NOEL MOHBERG. A Study of the Small Sample Properties of Tests of Linear Hypotheses for Categorized Ordinal Response Data. (Under the direction of JAMES E. GRIZZLE.)

The linear models analysis of categorized data with ordinal response variables has been investigated. A relationship between the \( \chi^2_L \) statistic, which arises in a natural way from weighted linear regression, and the Kruskal-Wallis \( H \) statistic was derived. It was found that if rank scores based on the true cell probabilities are used, \( \chi^2_L = HN/(N-1) \) when the null hypothesis is true. However, under alternative hypotheses, \( \chi^2_L \) generally exceeds \( HN/(N-1) \). The variance of the \( H \) statistic, conditional on the observed ties, was derived and a computer program to evaluate \( \text{Var}_c(H) \) was prepared.

Preliminary simulations employing \( \chi^2_L \) as the test statistic showed that the distribution of the observed values of \( \chi^2_L \) under the null hypothesis was poorly approximated by \( \chi^2 \), as suggested by \( \chi^2_L > HN/(N-1) \), so the formulation of the variance used in \( \chi^2_L \) was changed to use the covariance matrix based on the response marginal distribution added over treatments, rather than within treatments. The resulting statistic, called \( H' \), was then studied in a simulation experiment which was designed to investigate the effects of the following variables: sample size, equality of the treatment subsample sizes, response distribution form, number of categories of the response variables, and the number of levels of the factor variables. The results of the simulations indicated that for small samples in two-way cross classification problems, the null hypothesis distribution of \( H' \) can be quite generally approximated by \( \chi^2 \). Further simulations under alternative hypotheses showed that the \( H' \) test has power comparable to the \( F \) test.
ACKNOWLEDGEMENTS

The author wishes to express his appreciation for the assistance of his advisor, Dr. J. E. Grizzle, who suggested the problem and who provided counsel and thoughtful guidance throughout the research and its writing. In addition, he would like to thank the other members of his advisory committee, Drs. R. W. Helms, P. A. Lachenbruch, M. J. Symons, and S. J. Zyzanski for their helpful criticisms and suggestions.

He also wishes to thank his wife, Betty Jane, his children, and his friends for their support and encouragement throughout his studies.

Finally, the author expresses his appreciation to Mrs. Pamela Smith for her excellent typing of the paper.
# TABLE OF CONTENTS

| LIST OF TABLES | .......... | v |
| LIST OF FIGURES | .......... | vi |

## CHAPTER

### I. INTRODUCTION AND REVIEW OF THE LITERATURE .......... 1

1.1 Introduction ........................................ 1  
1.2 Definitions and Categorical Data Framework .......... 2  
1.3 Methods of Analysis for Ordinal Response  
Categorical Data ........................................ 3  
1.3.1 Kruskal-Wallis Rank Analysis of Variance .......... 4  
1.3.2 General Linear Models Approach .................. 6  
1.3.3 Discussion of Applicable Methods ................. 8  

### II. THE KRUSKAL-WALLIS TEST, THE LINEAR MODELS APPROACH,  
AND THE MOMENTS OF THEIR TEST STATISTICS .......... 11

2.1 Introduction ........................................ 11  
2.2 Derivation of the Relationship Between the Kruskal- 
Wallis $H$ Statistic and the Linear Models $\chi^2_L$  
Statistic ............................................. 11  
2.3 The Moments of the Kruskal-Wallis $H$ Statistic .... 21  
2.4 The Moments of the $\chi^2_L$ Statistic ............. 24  

### III. SIMULATION STUDIES ................................. 26

3.1 Introduction ........................................ 26  
3.2 The Computer Simulation Program .................... 27  
3.2.1 The Program and Its Workings .................... 27  
3.2.2 Test Cases .................................... 31  

3.3 The Variables of Interest and Analytical Frame- 
work of the Simulation Studies ......................... 32  
3.4 Preliminary Simulations; Examination and Modifi- 
cation of $\chi^2_L$ ................................... 37

### IV. ANALYSIS OF NULL HYPOTHESIS SIMULATION RESULTS .......... 40

4.1 Introduction ........................................ 40  
4.2 Analysis of the Simulation Study of $H^1$ .......... 41
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.1</td>
<td>The Means of the Test Statistics</td>
<td>41</td>
</tr>
<tr>
<td>4.2.2</td>
<td>The Variances of the Test Statistics</td>
<td>45</td>
</tr>
<tr>
<td>4.2.3</td>
<td>The Cumulative Distributions of the Test Statistics</td>
<td>50</td>
</tr>
<tr>
<td>4.3</td>
<td>The F Approximation of the Distribution of the H' Statistic</td>
<td>52</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Background for the Use of a Beta Approximation</td>
<td>52</td>
</tr>
<tr>
<td>4.3.2</td>
<td>An Empirical Comparison of the $F$ and $\chi^2_L$ Approximations</td>
<td>72</td>
</tr>
<tr>
<td>4.4</td>
<td>Summary of Analysis of Simulation Results</td>
<td>77</td>
</tr>
<tr>
<td>V.</td>
<td>TEST PERFORMANCE UNDER ALTERNATIVE HYPOTHESES</td>
<td>79</td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>79</td>
</tr>
<tr>
<td>5.2</td>
<td>The Simulation Experiment</td>
<td>82</td>
</tr>
<tr>
<td>5.2.1</td>
<td>The Analytical Framework of the Simulation Experiment</td>
<td>82</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Specification of Noncentralities in Simulations</td>
<td>83</td>
</tr>
<tr>
<td>5.3</td>
<td>Results of Simulation Experiment</td>
<td>84</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Expected Significance Level Results</td>
<td>84</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Other Results</td>
<td>88</td>
</tr>
<tr>
<td>5.4</td>
<td>The Blocking Factor Problem</td>
<td>90</td>
</tr>
<tr>
<td>5.5</td>
<td>Discussion and Summary</td>
<td>95</td>
</tr>
<tr>
<td>VI.</td>
<td>DISCUSSION, SUMMARY, AND RECOMMENDATIONS FOR FURTHER RESEARCH</td>
<td>96</td>
</tr>
<tr>
<td>6.1</td>
<td>Summary and Discussion</td>
<td>96</td>
</tr>
<tr>
<td>6.2</td>
<td>Recommendations for Further Research</td>
<td>99</td>
</tr>
</tbody>
</table>

APPENDICES

<p>| I.     | DERIVATION OF RELATIONSHIP BETWEEN $\chi^2_L$ AND H                    | 101  |
| II.    | DERIVATION OF $\text{VAR}_C(H)$                                       | 108  |
| III.   | LISTING OF COMPUTER PROGRAM FOR COMPUTING $\text{VAR}_C(H)$            | 115  |
| LIST OF REFERENCES                           | 119  |</p>
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>A General r×s Contingency Table</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>Expected Cell Probabilities</td>
<td>6</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Assumed Distributions of Response Variables</td>
<td>35</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Summary of Simulations Employing $\chi^2$</td>
<td>38</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Means of Test Statistics</td>
<td>42</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Observed Variances of Test Statistics</td>
<td>46</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Observed Variances of Test Statistics</td>
<td>49</td>
</tr>
<tr>
<td>4.2.4</td>
<td>Averages of Conditional Variances</td>
<td>51</td>
</tr>
<tr>
<td>4.2.5</td>
<td>Designs</td>
<td>53</td>
</tr>
<tr>
<td>4.3.1</td>
<td>F Approximations</td>
<td>74</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Variances by Numbers of Categories of Response and Response Distribution</td>
<td>85</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Sample Averages Producing ESL's of 0.05</td>
<td>85</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Observed ESL's and Their Standard Deviations</td>
<td>87</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Observed Powers of Principal Tests</td>
<td>89</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Observed Powers of Principal Tests</td>
<td>91</td>
</tr>
<tr>
<td>5.3.4</td>
<td>Sample Moments of Test Statistics under Alternative Hypotheses</td>
<td>92</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Sample Moments of Test Statistics for Null Factor Effects in the Presence of Non-null Effects</td>
<td>93</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.1</td>
<td>Two Sample Univariate Tests</td>
<td>55</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Four Sample Univariate Tests</td>
<td>55</td>
</tr>
<tr>
<td>4.2.3</td>
<td>B1 Contrast and $\chi^2$ Cumulatives</td>
<td>56</td>
</tr>
<tr>
<td>4.2.4</td>
<td>B2 Contrast and $\chi^2$ Cumulatives</td>
<td>56</td>
</tr>
<tr>
<td>4.2.5</td>
<td>B3 Contrast and $\chi^2$ Cumulatives</td>
<td>57</td>
</tr>
<tr>
<td>4.2.6</td>
<td>Two Sample Bivariate Tests, Bell Response, Equal Samples</td>
<td>58</td>
</tr>
<tr>
<td>4.2.7</td>
<td>Two Sample Bivariate Tests, Bell Response, Unequal Samples</td>
<td>58</td>
</tr>
<tr>
<td>4.2.8</td>
<td>Two Sample Bivariate Tests, J Response, Equal Samples</td>
<td>59</td>
</tr>
<tr>
<td>4.2.9</td>
<td>Two Sample Bivariate Tests, J Response, Unequal Samples</td>
<td>59</td>
</tr>
<tr>
<td>4.2.10</td>
<td>Two Sample Bivariate Tests, U Response, Equal Samples</td>
<td>60</td>
</tr>
<tr>
<td>4.2.11</td>
<td>Two Sample Bivariate Tests, U Response, Unequal Samples</td>
<td>60</td>
</tr>
<tr>
<td>4.2.12</td>
<td>Four Sample Bivariate Tests, Bell Response, Equal Samples</td>
<td>61</td>
</tr>
<tr>
<td>4.2.13</td>
<td>Four Sample Bivariate Tests, Bell Response, Unequal Samples</td>
<td>61</td>
</tr>
<tr>
<td>4.2.14</td>
<td>Four Sample Bivariate Tests, J Response, Equal Samples</td>
<td>62</td>
</tr>
<tr>
<td>4.2.15</td>
<td>Four Sample Bivariate Tests, J Response, Unequal Samples</td>
<td>62</td>
</tr>
</tbody>
</table>
4.2.16 Four Sample Bivariate Tests, U Response, Equal
Samples ......................................................... 63

4.2.17 Four Sample Bivariate Tests, U Response, Unequal
Samples ......................................................... 63

4.2.18 Two Sample Univariate Tests, Bell Response, Equal
Samples ......................................................... 64

4.2.19 Two Sample Univariate Tests, Bell Response, Unequal
Samples ......................................................... 64

4.2.20 Two Sample Univariate Tests, J Response, Equal
Samples ......................................................... 65

4.2.21 Two Sample Univariate Tests, J Response, Unequal
Samples ......................................................... 65

4.2.22 Two Sample Univariate Tests, U Response, Equal
Samples ......................................................... 66

4.2.23 Two Sample Univariate Tests, U Response, Equal
Samples ......................................................... 66

4.2.24 Four Sample Univariate Tests, Bell Response, Equal
Samples ......................................................... 67

4.2.25 Four Sample Univariate Tests, Bell Response, Unequal
Samples ......................................................... 67

4.2.26 Four Sample Univariate Tests, J Response, Equal
Samples ......................................................... 68

4.2.27 Four Sample Univariate Tests, J Response, Unequal
Samples ......................................................... 68

4.2.28 Four Sample Univariate Tests, U Response, Equal
Samples ......................................................... 69
4.2.29 Four Sample Univariate Tests, U Response, Unequal

Samples .................................................. 69

4.3.1 Two Sample Univariate H' and $\chi^2$ Cumulatives .......... 75

4.3.2 Four Sample Univariate H' and $\chi^2$ Cumulatives .......... 75

4.3.3 Transformed H' and F Cumulatives, Two Samples ...... 76

4.3.4 Transformed H' and F Cumulatives, Four Samples ...... 76
1.1 Introduction

Contingency tables, in which the response variables are ordinal and which have moderate-to-small sample sizes, are frequently encountered in research. Goodman (1968), Feinberg (1970), Mosteller (1968), Bishop (1969), and Birch (1963), have developed a variety of maximum likelihood based techniques for the analysis of many types of contingency tables, but in general, the maximum likelihood techniques developed thus far do not take into account any structure that may be present in the response variable. Hence, these analyses do not utilize all the information contained in the response variable. On the other hand, the general linear models approach of Grizzle, Starmer, and Koch, (1969), can be easily adapted to take advantage of ordinality of the response variable, but the small sample properties of the linear models approach are largely unknown.

The goals of this study are thus:

(1) To evaluate the properties of the linear models techniques for the analysis of ordinal response categorical data, and,

(2) To suggest modifications to these techniques that may improve their
small sample properties.

1.2 Definitions and Categorical Data Framework, Ordinal Response Case

In order to fix notation and to facilitate further discussion, an analytical description of the contingency table situation of interest now follows:

Contingency table data are usually representative of \( s \) independent multinomial distributions, each with \( r \) possible responses or outcomes. The appropriate likelihood is

\[
\prod_{i=1}^{s} \frac{n_{i}!}{\prod_{j=1}^{r} n_{ij}!} \prod_{j=1}^{r} \pi_{ij}^{n_{ij}} \tag{1.2.1}
\]

where: \( n_{i} \) is the number of individuals in the \( i^{th} \) sample, 
(a fixed number),

\( n_{ij} \) is the number of individuals in the \( i^{th} \) sample who exhibit the \( j^{th} \) response, (a random variable),

and \( \pi_{ij} \) is the probability of response \( j \) in sample \( i \).

Restrictions are \( \sum_{j=1}^{r} \pi_{ij} = 1 \) and \( \sum_{j=1}^{r} n_{ij} = n_{i} \) for all \( i \).

\[
\sum_{i=1}^{s} n_{i} = N , \text{ where } N \text{ is the total sample size.}
\]

Responses 1, 2, \ldots r can be ranked according to size in some sense.

We shall denote observed probabilities by \( p_{ij}'s \) and multinomial parameters by \( \pi_{ij}'s \). Hypotheses concerning the data can be formulated as

\[
H_{0}: F_{k}(n) = 0, \text{ } k = 1, 2 \ldots t, \text{ } t \leq s-1 , \tag{1.2.2}
\]

where the \( F_{k}'s \) are \( t \) independent functions of the \( \pi_{ij}'s \), and
\[ \pi' = (\pi_{11}', \pi_{12}', \ldots, \pi_{1r}'; \pi_{21}', \pi_{22}', \ldots, \pi_{2r}'; \ldots, \pi_{s1}', \pi_{s2}', \ldots, \pi_{sr}') \]

The restriction \( \sum_{j=1}^{r} \pi_{ij} = 1 \) would allow deletion of each of the \( \pi_{ir} \) parameters, but for this ordinal response case it is computationally easier to retain these redundant parameters. Table 1.1 is an example of a contingency table as described above.

**Table 1.1**

A GENERAL \( r \times s \) CONTINGENCY TABLE

<table>
<thead>
<tr>
<th>Categories of Response</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Samples</strong></td>
</tr>
<tr>
<td>( n_{11} )</td>
</tr>
<tr>
<td>( n_{21} )</td>
</tr>
<tr>
<td>( \vdots )</td>
</tr>
<tr>
<td>( s )</td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
</tbody>
</table>

The \( s \) different samples will be treated as forming \( s \) levels of a factor variable, and the categories of response form levels of the response variable. The categories of response can be ordered so that \( 1 < 2 < 3 < \ldots < r \).

**1.3 Methods of Analysis for Ordinal Response Categorical Data**

To utilize the ordinality structure of the response variables in our analyses, we will replace the observations by scores and analyze the scores. The form of the analysis of the scores will be discussed in the following sections, but the scores substituted for the
observations will be rank scores. The motivation for the selection of rank scores arises from work by Ghosh (1969), who showed that for the underlying logistic distribution in the continuous data case, rank scores provided an optimal test. In a comparative Monte Carlo study dealing with the continuous case, Bhattacharyya, Johnson, and Neave (1971) found the bivariate rank sum test to have power comparing favorably with the Hotelling $T^2$ test, and in the tied data case, Krauth (1971) showed a rank sum test is locally most powerful against the class of locally linear alternatives. Other possible scores are the less intuitively appealing arbitrary scores 1, 2, and 3, suggested by Grizzle, Stramer, and Koch (1969) in their multiple sample example, and normal scores, which would provide optimal tests if the underlying distributions were normal.

1.3.1 **Kruskal-Wallis Rank Analysis of Variance**

For the experimental situation where (1) Numerous different ordinal responses are present, i.e., the "$r$" of Section 1.2 is not small, and (2) The data can be cast into a simple $r \times s$ table such as Table 1.1, the well-known Kruskal-Wallis test may be a useful analytic method. (We note that for a multiple factor variable case we could treat combinations of factors as forming levels of a single composite factor and thus retain a simple contingency table as shown in Table 1.1.) To perform the Kruskal-Wallis test, one substitutes rank scores for the observations and essentially performs an analysis of variance on these ranks. A full description of the procedure is given in Kruskal and Wallis, (1952).
Since in contingency tables all individuals who exhibit response
j would be considered to be tied, for our purposes, the treatment of
ties and their effects are very important. Kruskal (1952) showed
that after correction for ties, the Kruskal-Wallis test statistic is
still asymptotically \( \chi^2 \), but the distribution of the test statistic
is essentially untabulated in the ties-present case.

The test statistic for the Kruskal-Wallis test is

\[
H = \left\{ \frac{12}{N(N+1)} \right\} \sum_{i=1}^{S} \frac{R_i^2/n_i}{3(N+1)}/\left\{ 1 - \sum_{j=1}^{r} \frac{T_j}{N^3-N} \right\}
\]

(1.3.1)

where \( s \) is the number of samples or levels of the factor variable,

\( N \) is the total sample size,

\( R_i \) is the sum of ranks in the \( i^{\text{th}} \) sample,

\( n_i \) is the size of the \( i^{\text{th}} \) sample,

and \( T_j = n_j^3 - n_j \) where \( n_j \) is the number of individuals who

exhibit the \( j^{\text{th}} \) response.

\( H \) is approximately distributed as a chi-square with \((s-1)\) degrees of

freedom.

Kruskal and Wallis (1952) pointed out that this is not, rigorously,
a test comparing average ranks among the \( s \) samples; it tests a null
hypothesis of homogeneity among the \( s \) underlying distributions. This
null hypothesis relates to the hypothesis (1.2.2) through

\[
H_0 : \pi_{11} = \ldots = \pi_{1s_1} ; \ \pi_{12} = \ldots = \pi_{2s_2} ; \ldots ; \ \pi_{1(r-1)} = \ldots = \pi_{s(r-1)}
\]

(1.3.2)

Hence, functions of differences within these \((r-1)\) sets of equalities
will be zero if they are properly chosen and the null hypothesis is
true. Kruskal and Wallis also pointed out that the two sample Kruskal-
Wallis test is equivalent to the Wilcoxon test.
1.3.2 General Linear Models Approach

The general linear models system of analysis, as presented by Grizzle, Starmer, and Koch, (1969), allows a good deal of flexibility in formulating the analysis of ordinal response categorical data, because multiple factors and their interactions can be handled with established methods and the system is set up with the aim of analyzing categorized, and hence tied, data. A brief description of the system is as follows:

Consider the following table of expected cell probabilities:

Table 1.2
EXPECTED CELL PROBABILITIES

<table>
<thead>
<tr>
<th>Categories of Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Samples</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>s</td>
</tr>
</tbody>
</table>

Define $\pi_i = [\pi_{i1}, \pi_{i2}, \ldots, \pi_{ir}]$, and $\pi_s = [\pi_{s1}, \pi_{s2}, \ldots, \pi_{sr}]$. 

$p_{ij} = n_{ij}/n_i$. 

$p_{i}^\prime = [p_{i1}, p_{i2}, \ldots, p_{ir}]$, and $p_{s}^\prime = [p_{s1}, p_{s2}, \ldots, p_{sr}]$. 


\begin{equation}
\begin{pmatrix}
\pi_{11}(1-\pi_{11}) & -\pi_{11}\pi_{12} & \cdots & -\pi_{11}\pi_{1r} \\
-\pi_{11}\pi_{12} & \pi_{12}(1-\pi_{12}) & \cdots & -\pi_{12}\pi_{1r} \\
\vdots & \vdots & \ddots & \vdots \\
-\pi_{11}\pi_{ir} & -\pi_{12}\pi_{ir} & \cdots & \pi_{ir}(1-\pi_{ir})
\end{pmatrix}
\end{equation}
(1.3.3)

\(V(p)\) is the sample estimate of \(V(\pi)\) and \(V(p)\) is a block diagonal matrix with \(V(p)\) matrices forming the blocks on the main diagonal.

Suppose that \(f_m(\pi)\) is a function of \(\pi\) which has continuous second order partial derivatives with respect to \(\pi_m\), \(m = 1, 2, \ldots, u(s-1)\).

Then

\[
F'(\pi) = [f_1(\pi), f_2(\pi), \ldots, f_u(\pi)], \text{ and }
F'(p) = [f_1(p), f_2(p), \ldots, f_u(p)].
\]

Further define

\[
H = \begin{bmatrix}
\frac{\partial f_m(\pi)}{\partial \pi_{ij}} \\
\frac{\partial \pi_{ij}}{\partial \pi_{ij}}
\end{bmatrix}
\quad \text{and} \quad
S = H V(p) H'
\]

The hypothesis (1.2.2) can be formulated as

\[
F(\pi) = X \beta, \quad \text{where} \quad X \quad \text{is a known design matrix of rank} \quad v \leq u, \quad \text{and} \quad \beta \quad \text{is a vector of unknown population parameters.}
\]

From (1.3.4), an estimate of \(\beta\) can be obtained by minimizing the expression

\[
(F - X \beta)' S^{-1}(F - X \beta)
\]

with respect to \(\beta\). The minimum value of this expression is used to test the fit of the model \(F(\pi) = X \beta\), and if the fit is adequate, a test of the hypothesis \(H_0 : C \beta = 0\) is performed employing weighted multiple regression. The estimate of the \(\beta\) vector is

\[
\beta = (X' S^{-1} X)^{-1} X' S^{-1} F. \quad (1.3.5)
\]
Taking $C$ to be a $(d \times \nu)$ matrix of arbitrary constants and of rank $d \leq \nu$, and given that the model fits the data, the test of the hypothesis $C \beta = 0$ is produced by

$$SS \begin{bmatrix} C & \beta = 0 \end{bmatrix} = b' C' [C(X' S^{-1} X)^{-1} C']^{-1} C b,$$  \hspace{1cm} (1.3.6)

which is asymptotically chi-square distributed with $d$ degrees of freedom if $H_o$ is true.

1.3.3 Discussion of Applicable Methods

The exact distribution of the Kruskal-Wallis $H$ statistic can be found in a somewhat laborious but straightforward manner, provided there are no ties in the data. Kruskal and Wallis (1952) included a table of the exact distribution of $H$ in the three sample, ties-absent case, where each sample is of size five or less. They tabulated the distribution of $H$ in the neighborhoods of the 10, 5, and 1 percent points, and Alexander and Quade (1968) extended the Kruskal and Wallis tables to the case where $N = 26$ and $n_i \geq 2$, $i = 1, 2, 3$. Kruskal and Wallis, Wallace (1952) and Alexander and Quade discussed a number of approximations to the exact distribution of $H$; these approximations will be examined in Section 4.3.

When ties are possible, enumeration of the possible configurations becomes complex because one must condition on the number of ties present in the data. With the exception of the two sample case, where Klotz (1966), and Lehman (1961), have published small tables of the Wilcoxon statistic, the ties-present case remains untabulated. Kruskal (1952) proved that when ties are treated according to the midrank method and the $H$ statistic is adjusted for ties as in (1.3.1), $H$ is still
asymptotically chi-square distributed with \((s-1)\) degrees of freedom. Very little work has been done in developing approximations to the distribution of \(H\) in this situation; Klotz (1966), studied the Wilcoxon statistic with ties present but was unable to improve on the usual normal approximation of \(W\) by employing the first four moments of \(W\) in an approximation. The moments of the \(H\) statistic for the ties-present case will be discussed further in Section 2.2.

When one examines the properties of the Grizzle, Starmer, and Koch general linear model analysis of categorical data, the foremost consideration is the performance of the modified chi-squared statistic \(\chi^2_L\).

The asymptotic properties of \(\chi^2_L\) were established by Neyman (1949). He showed that the minimum modified chi-square (\(\min \chi^2_L\)) estimation procedure is asymptotically equivalent to minimum chi-square (\(\min \chi^2\)) and maximum likelihood (ML) estimation. If the condition that \(F(\pi)\) has continuous partial derivatives up to the second order is satisfied, then the three estimation methods share the following properties:

1. They are functions of \(p_{ij}\) and do not depend directly on \(N\).
2. They are consistent.
3. As \(N \to \infty\), their distributions tend to normality.
4. Their variances are less than or equal to the variances of other estimators satisfying (2) and (3).

These properties imply that \(\min \chi^2_L\), \(\min \chi^2\), and ML estimates are Best Asymptotically Normal (BAN) estimates, and hence they are asymptotically equivalent.
Neyman further showed that the Pearson $\chi^2$, $\chi^2_\ell$, and the likelihood ratio $\chi^2$ are asymptotically equivalent test statistics; as $N \to \infty$, each of the test statistics has a limiting $\chi^2$ distribution with $t$ degrees of freedom, $(t \leq r(s-1))$, if (1) the ratio $n_{ij}/N$ remains constant as $N \to \infty$, (2) there exists at least one solution to the equation $F(\pi) = 0$ such that $\pi_{ij} > 0$ for all $i$ and $j$ and (3) the null hypothesis is true.

Beyond the sampling studies of Berkson (1968) and Odoroff (1968), little is known about the small sample properties of $\chi^2_\ell$ as an estimator or a test statistic. C. R. Rao (1963) found that under his large sample concept of Second Order Efficiency maximum likelihood estimators have the smallest variances of the three methods, but Odoroff found in his small sample empirical work that $\min \chi^2_\ell$ estimates compared favorably with ML estimates.

All of the above discussion of the properties of $\chi^2_\ell$ deals with the nominal response case, and it is possible that these properties may be somewhat different in the ordinal response categorical data case. For further work, analogies drawn with the Kruskal-Wallis statistic or $F$ ratio statistic will probably be more useful.
CHAPTER II

THE KRUSKAL-WALLIS TEST, THE LINEAR MODELS APPROACH,
AND THE MOMENTS OF THEIR TEST STATISTICS

2.1 Introduction

In this chapter, we shall develop some theoretical results for later use in a simulation study of the small sample properties of the linear models analysis of ordinal response categorical data. We shall first show a relationship between the Kruskal-Wallis $H$ statistic and the modified chi square statistic, $\chi^2_{\ell}$. After exploring this relationship, we will review some material concerning the moments of the $H$ statistic in the ties-absent case, and then derive an expression for the conditional variance of $H$ in the ties-present, or categorical data, situation. Then, following a description of a computer program designed to evaluate the conditional variance of $H$ in the ties-present case, we will consider the problem of finding the moments of $\chi^2_{\ell}$.

2.2 Derivation of the Relationship Between the Kruskal-Wallis $H$
Statistic and the Linear Models $\chi^2_{\ell}$ Statistic

When we substitute rank scores for the observations in a categorical data problem, and analyze the rank data using the linear models
approach, this analysis is related to the Kruskal-Wallis analysis. We can intuitively recognize this relationship by observing that the linear models analysis is a weighted least squares procedure, and that the Kruskal-Wallis test, since it is basically an analysis of variance, is a special case of weighted least squares procedures. Assuming that the same rank scores are substituted for the observations, and that the assumption of equal variance within the samples is satisfied, the two analyses should yield similar results.

An outline of the derivation of the relationship between $\chi^2_\ell$ and $H$ is presented here; the detailed derivation is contained in Appendix 1.

Suppose we have a contingency table similar to Table (1.1) and we wish to analyze it using the Kruskal-Wallis test, adjusted for ties. We would replace the observations exhibiting the $j^{th}$ response by the midrank score $a_j = n_{j1} + \ldots + (n_{rj} + 1)/2$, $j = 1, \ldots, r$, compute the totals of the scores over the sample and thence compute the test statistic (1.3.1),

$$H = \left\{ \frac{1}{12/N(N+1)} \sum_{i=1}^{S} R_i^2/n_{i1} - 3(N+1)/\left\{1 - \sum_{1}^{T} j/(N^3-N) \right\} \right\}.$$

(2.2.1)

This formulation of the $H$ statistic can be rewritten as

$$H = \frac{(N-1)/N}{\left[\frac{1}{4} \sum_{i \neq j} \frac{n_{i1} n_{j1} + 2}{n_{i1} n_{j1}} \sum_{i < k} \frac{n_{i1} n_{j1} n_{k1}}{n_{i1} n_{j1}} \right] \left( \sum_{1}^{S} \sum_{j=1}^{S} n_{i1} n_{j1} (f_i - f_j)^2 \right)}.$$

(2.2.2)

To perform the linear models analysis, we would also replace the observations by the rank scores $a_j$, and compute the test statistic $\chi^2_\ell$ as shown at (1.3.5) and (1.3.6), using suitable $C$ and $X$ matrices.
This yields the test statistic

\[ \chi^2 = \sum_{i<j} \sum_{k \neq i,j} (f_{i,j} - f_{j,j})^2 / \sum_{k \neq i} \sum_{1}^{s} n_i \cdot n_j \cdot z_k, \]  

(2.2.3)

where \( z_k \) is the variance within the \( k \)th sample. Under the null hypothesis \( p_{ij} = p_{i'j} = p_j \), \( i, i' = 1, 2...s, j = 1,...r \), \( \chi^2 \) can be rewritten as

\[ \chi^2 = \sum_{i<j} \sum_{1}^{s} n_i \cdot n_j \cdot (f_{i,j} - f_{j,j})^2 / zN, \]  

(2.2.4)

where \( z_1 = z_2 = ... = z_s = z \).

Comparing (2.2.2) with (2.2.4), as shown in Appendix I, we obtained the relationship

\[ \chi^2 \frac{(N-1)}{N} = H \]  

(2.2.5)

It should be recalled that we have assumed that the \( p_{ij} = p_{i'j} \), \( j = 1,...r \), in deriving (2.2.5). This implies that (2.2.5) will hold exactly only when the null hypothesis is true, which in turn implies that (2.2.5) will hold exactly only when \( \chi^2 \) and \( H \) are both identically zero. However, the relationship may be of use as an estimate, so we will now investigate the relationship between \( \chi^2 \) and \( H \) when their variance terms are not assumed equal. From (2.2.3), \( \chi^2 \) in the two sample case is

\[ \chi^2 = (f_1 - f_2)^2 / (1/n_1 \cdot \text{Var} X_1 + 1/n_2 \cdot \text{Var} X_2), \]  

(2.2.6)

where \( \text{Var} X_i = n_i \cdot \text{Var} X_i \) is the variance of a single observation in sample \( i \). In the Kruskal-Wallis test, we assume \( \text{Var} X_1 = \text{Var} X_2 \) and base the variance on the response marginal. Hence, we have

\[ H = ((N-1)/N)(f_1 - f_2)^2 / [(1/n_1 + 1/n_2) \cdot \text{(Var} X)], \]  

(2.2.7)
where Var X = \(1/N[\sum_{l=1}^{n_1} X_{1l}^2 + \sum_{l=1}^{n_2} X_{2l}^2 - N(N+1)^2/4]\). Adding and subtracting \((n_1.f_1)^2/n_1\) and \((n_2.f_2)^2/n_2\) from Var X, we have

\[
\text{Var X} = \frac{1}{N}\left[\sum_{l=1}^{n_1} X_{1l}^2 - (n_1.f_1)^2/n_1 + \sum_{l=1}^{n_2} X_{2l}^2 - (n_2.f_2)^2/n_2\right] + \frac{1}{n_1}(n_1.f_1)^2/n_1 + \frac{1}{n_2}(n_2.f_2)^2/n_2 - \frac{N(N+1)^2}{4}
\]

Substituting \((n_1.f_1+n_2.f_2)^2/N\) for \(N(N+1)^2/4\), and noting that

\[
\sum_j X_{1j}^2 - (n_1.f_1)^2/n_1 = n_1 \cdot \text{Var } X_1,
\]

\[
\text{Var X} = \frac{1}{N}[n_1 \cdot \text{Var } X_1 + n_2 \cdot \text{Var } X_2 + n_1 n_2 \cdot (\text{Var } X_1 + \text{Var } X_2 + (f_1-f_2)^2)]
\]

Because \((1/n_1 + 1/n_2) = (n_1 + n_2)/(n_1 n_2)\), (2.2.7) becomes

\[
H = \frac{(N-1)/N}{(f_1-f_2)^2/[1/N^2(n_1 + n_2)/n_1 \cdot \text{Var } X_1 + n_2/n_1 \cdot \text{Var } X_2 + \text{Var } X_1 + \text{Var } X_2 + (f_1-f_2)^2)}
\]

(2.2.8)

Now forming the difference between the denominators of (2.2.6) and (2.2.8), and calling it Y, we have

\[
Y = \frac{1}{N}[n_1 \cdot \text{Var } X_1 + n_2/n_1 \cdot \text{Var } X_2 + \text{Var } X_1 + \text{Var } X_2 + (f_1-f_2)^2]
\]

\[
- \frac{(n_1 + n_2)/(n_1 \cdot \text{Var } X_1 - (n_1 + n_2)/n_2 \cdot \text{Var } X_2)}
\]

\[
= \frac{1}{N}[n_1/n_2 \cdot \text{Var } X_1 + n_2/n_1 \cdot \text{Var } X_2 + (f_1-f_2)^2 - n_2/n_1 \cdot \text{Var } X_1 - n_1/n_2 \cdot \text{Var } X_2]
\]

(2.2.9)

In (2.2.9), it is clear that if \(n_1 = n_2\), then \(Y > 0\) if \(f_1 \neq f_2\). If \(n_1 \neq n_2\), then \(Y\) can be negative if \((f_1-f_2)^2\) is small enough and \(n_i > n_j\) while \(\text{Var } X_i < \text{Var } X_j\), \(i \neq j = 1, 2\).
From (2.2.9), for the two equal sample case, we can write

\[ H = \frac{(N-1)/N}{((f_1-f_2)^2/N + a'V_1a + a'V_2a)} , \]  

(2.2.10)

and hence we have the relationship

\[ H = \chi^2((N-1)/N)(a'V_1a + a'V_2a)/[(f_1-f_2)^2/N + a'V_1a + a'V_2a] . \]  

(2.2.11)

An example of the effect of this relationship follows:

**Example 2.1**

Consider the following data:

<table>
<thead>
<tr>
<th>Response</th>
<th>Sample</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>5 ( R_1 = 16.5, f_1 = 3.3 )</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>5 ( R_2 = 38.5, f_2 = 7.7 )</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>10</td>
</tr>
</tbody>
</table>

\[
H = \frac{12}{10 \times 11} \left( \frac{16.5^2}{5} + \frac{38.5^2}{5} \right) - 3(11) / (1 - 72/990) 
\]

\[ = 5.694 \]

\[
\chi^2 = \frac{(4.4)^2}{.432+.692} = 17.2242 
\]

From (2.2.10), \( H = (9/10) (19.36)/[19.36/10 + .432 + .692] \)

\[ = 5.694 \]

or, working from (2.2.7),

\[
V = \begin{bmatrix}
.09 & -.02 & -.01 & -.04 & -.02 \\
.16 & -.02 & -.08 & -.04 \\
.09 & -.04 & -.02 \\
.24 & -.08 & .16 \\
\end{bmatrix}, \ a = (1, 2.5, 4.0, 6.5, 9.5) ,
\]
and \( n_1 a' V a = a_1^2 v_{11} + a_2^2 v_{22} + \ldots + 2a_4 a_5 v_{45} \)
\[
= 27.11 - 2(9.73)
\]
\[
= 7.65.
\]

Hence, \( H = (9/10)(19.36)/(7.65/5+7.65/5) \)
\[
= 5.694
\]

This example points out rather vividly how \( \chi^2_\ell \) and \( H \) are not equal
when the null hypothesis relationship \( p_{1j} = p_{1'j} = p_j \) is not true.

The relationship between \( \chi^2_\ell \) and \( H \) in the multiple sample case
is not so easily handled as the two sample case; we can write down \( H \),
given \( \chi^2_\ell \), by replacing \( a' V a \) terms by \( a' V a \), but it does not
appear that a closed form relationship between the two statistics can
be found without making the assumptions \( p_{1j} = p_{1'j} = p_j \).

For the three sample case, from (2.2.3),
\[
\chi^2_\ell = \frac{n_2 n_3 z_1 (f_2 - f_3)^2 + n_1 n_3 z_2 (f_1 - f_3)^2 + n_1 n_2 z_3 (f_1 - f_2)^2}{n_3 z_1 z_2 + n_2 z_1 z_3 + n_1 z_2 z_3}
\]

If we assume that \( (z_1 + z_2 + z_3)/3 = z \), the overall variance, and we
write \( z_i = z + \delta_i \), \( i = 1, 2, 3 \), where \( \delta_1 + \delta_2 + \delta_3 = 0 \) then
\[
\chi^2_\ell = \frac{n_2 n_3 (z + \delta_1)(f_2 - f_3)^2 + n_1 n_3 (z + \delta_2)(f_1 - f_3)^2 + n_1 n_2 (z + \delta_3)(f_1 - f_2)^2}{n_3 (z + \delta_1)(z + \delta_2) + n_2 (z + \delta_1)(z + \delta_3) + n_1 (z + \delta_2)(z + \delta_3)}
\]
(2.2.12)

Expanding, assuming \( n_1 = n_2 = n_3 = n \), and noting that
\( \delta_1 = 0 \) implies \( \delta_1^2 + \delta_2^2 + \delta_3^2 = -2(\delta_1 \delta_2 + \delta_1 \delta_3 + \delta_2 \delta_3) \), i.e., that
\( \delta_1 \delta_2 + \delta_1 \delta_3 + \delta_2 \delta_3 < 0 \) if at least one \( \delta_i \neq 0 \), (2.2.12) reduces to
\[
\chi^2_\ell = \frac{n(z(f_2 - f_3)^2 + z(f_1 - f_3)^2 + z(f_1 - f_2)^2)}{3z^2 + \delta_1 \delta_2 + \delta_1 \delta_3 + \delta_2 \delta_3}
\]
\[ n \left( \delta_1 (f_{22} - f_{3})^2 + \delta_2 (f_{1} - f_{3})^2 + \delta_3 (f_{1} - f_2)^2 \right) \]
\[ + \frac{3\delta_1^2 + \delta_2^2 + \delta_3^2}{3\delta_1^2 + \delta_2^2 + \delta_3^2} \]

(2.2.13)

The expectations of all the \((f_i - f_j)^2\) are the same, so, since \(\sum \delta_i = 0\), the expected value of the numerator of the right-most term of (2.2.13) will be zero. Further, since \(\delta_1 \delta_2^2 + \delta_1 \delta_3^2 + \delta_2 \delta_3^2 < 0\), the term immediately to the right of the equal sign in (2.2.13) is greater than \(H\) since its numerator is identical to that of \(H\) while its denominator is strictly smaller. Hence, for the three equal sample case, we should expect \(\chi^2_L\) to exceed \(H \frac{N}{N-1}\). Again, this relationship is not necessarily true for unequal sized samples.

The four sample case is more difficult to work with than the three sample case; an argument similar to the three sample case is troubled by \(\sum \delta_i^2 \delta_j\) and \(z(\sum \delta_i \delta_j) + \sum \delta_i \delta_j \delta_k\), which are of uncertain signs, depending on the \(\delta_i\) and \(z\). The five sample case is more difficult still, so we will turn to less specific analysis.

Examination of the \(\chi^2_L\) test statistics at (2.2.3) indicates that if the \(z_i\) are approximately equal, \(\chi^2_L\) is approximately expressed by (2.2.4), where \(z = \text{ave}(z_i)\). If \(z\) is less than the Kruskal-Wallis variance \(N(N-1) - \sum T_j^2) / 12\), and it will be if there is any grouping of the scores within the samples, then \(\chi^2_L\) at (2.2.4) will be greater than \(H \frac{N}{N-1}\). Another way of looking at the question is by treating the \(z\) as an "error mean square" and the variance of the sample averages as a "treatments mean square". Clearly, only if the treatments all have the same average will the error mean square reflect the population variance.
During the program checkout work, in trials with four samples of size 100 each, \( \chi^2 \) and \( H \) were computed for each of 32 contingency tables. The range of \( (\chi^2 - H) \) was from +0.5060 to -0.2042; for nine tables \( H \) exceeded \( \chi^2 \). The average of \( (\chi^2 - H) \) was 0.0702, adjusted for the \( N/(N-1) \) factor.

We conclude that \( \chi^2 \) will often exceed \( NH/(N-1) \). The difference will be marked for two samples, and decrease as the numbers of samples increases and/or the sample size increases.

The preceding work has been for the one way classification problem; of more interest in our investigation of complex contingency tables is the relationship between \( \chi^2 \) and \( H \) when \( \chi^2 \) is computed in cross classification problem. To keep arguments concrete, we shall look at a specific case and attempt to generalize from it.

Suppose in the model (1.3.4) we take

\[
\begin{bmatrix}
 1 & 1 & 1 & 0 & 0 \\
 1 & 1 & 0 & 1 & 0 \\
 1 & 1 & 0 & 0 & 1 \\
 1 & 1 & -1 & -1 & -1 \\
 1 & -1 & 1 & 0 & 0 \\
 1 & -1 & 0 & 1 & 0 \\
 1 & -1 & 0 & 0 & 1 \\
 1 & -1 & -1 & -1 & -1
\end{bmatrix}
\]

and \( C = (0, 1, 0, 0, 0) \)

in (1.3.6). This yields a test for the effect of the two level factor in a \( 2 \times 4 \) cross classification. Assuming that the sample sizes in each of the eight cells are equal to \( n \), then the variance for the test for an effect caused by the two level factor is

\[
[C'[X'X^{-1}X]^{-1}C]
\]

where
\[
S^{-1} = \begin{bmatrix}
\frac{1}{a_1^2} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \frac{1}{a_8^2}
\end{bmatrix} = \begin{bmatrix}
n/z_1 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & n/z_8
\end{bmatrix}
\]

Hence,

\[
(X'S^{-1}X)^{-1} = \begin{bmatrix}
4 & \sum_{i=1}^{8} \frac{n}{z_i} & \sum_{i=1}^{8} \frac{n}{z_i^2} & \sum_{i=1}^{8} \frac{n}{z_i^3} & \sum_{i=1}^{8} \frac{n}{z_i^4} & \sum_{i=1}^{8} \frac{n}{z_i^5} & \sum_{i=1}^{8} \frac{n}{z_i^6} & \sum_{i=1}^{8} \frac{n}{z_i^7} & \sum_{i=1}^{8} \frac{n}{z_i^8} \\
\sum_{i=1}^{8} \frac{n}{z_i} & n & \sum_{i=1}^{8} \frac{n}{z_i^2} & \sum_{i=1}^{8} \frac{n}{z_i^3} & \sum_{i=1}^{8} \frac{n}{z_i^4} & \sum_{i=1}^{8} \frac{n}{z_i^5} & \sum_{i=1}^{8} \frac{n}{z_i^6} & \sum_{i=1}^{8} \frac{n}{z_i^7} & \sum_{i=1}^{8} \frac{n}{z_i^8} \\
\sum_{i=1}^{8} \frac{n}{z_i^2} & \sum_{i=1}^{8} \frac{n}{z_i^3} & n & \sum_{i=1}^{8} \frac{n}{z_i^4} & \sum_{i=8}^{8} \frac{n}{z_i^5} & \sum_{i=1}^{8} \frac{n}{z_i^6} & \sum_{i=1}^{8} \frac{n}{z_i^7} & \sum_{i=1}^{8} \frac{n}{z_i^8} \\
\sum_{i=1}^{8} \frac{n}{z_i^3} & \sum_{i=1}^{8} \frac{n}{z_i^4} & \sum_{i=1}^{8} \frac{n}{z_i^5} & n & \sum_{i=8}^{8} \frac{n}{z_i^6} & \sum_{i=1}^{8} \frac{n}{z_i^7} & \sum_{i=1}^{8} \frac{n}{z_i^8} \\
\sum_{i=1}^{8} \frac{n}{z_i^4} & \sum_{i=1}^{8} \frac{n}{z_i^5} & \sum_{i=1}^{8} \frac{n}{z_i^6} & \sum_{i=1}^{8} \frac{n}{z_i^7} & n & \sum_{i=8}^{8} \frac{n}{z_i^8} \\
\sum_{i=1}^{8} \frac{n}{z_i^5} & \sum_{i=1}^{8} \frac{n}{z_i^6} & \sum_{i=1}^{8} \frac{n}{z_i^7} & \sum_{i=1}^{8} \frac{n}{z_i^8} & \sum_{i=8}^{8} \frac{n}{z_i^8} & n \\
\sum_{i=1}^{8} \frac{n}{z_i^6} & \sum_{i=1}^{8} \frac{n}{z_i^7} & \sum_{i=1}^{8} \frac{n}{z_i^8} & \sum_{i=1}^{8} \frac{n}{z_i^8} & \sum_{i=8}^{8} \frac{n}{z_i^8} & \sum_{i=1}^{8} \frac{n}{z_i^8} & n \\
\sum_{i=1}^{8} \frac{n}{z_i^7} & \sum_{i=1}^{8} \frac{n}{z_i^8} & \sum_{i=1}^{8} \frac{n}{z_i^8} & \sum_{i=1}^{8} \frac{n}{z_i^8} & \sum_{i=8}^{8} \frac{n}{z_i^8} & \sum_{i=1}^{8} \frac{n}{z_i^8} & \sum_{i=1}^{8} \frac{n}{z_i^8} & n \\
\sum_{i=1}^{8} \frac{n}{z_i^8} & \sum_{i=1}^{8} \frac{n}{z_i^8} & \sum_{i=1}^{8} \frac{n}{z_i^8} & \sum_{i=1}^{8} \frac{n}{z_i^8} & \sum_{i=8}^{8} \frac{n}{z_i^8} & \sum_{i=1}^{8} \frac{n}{z_i^8} & \sum_{i=1}^{8} \frac{n}{z_i^8} & \sum_{i=1}^{8} \frac{n}{z_i^8} & n
\end{bmatrix}
\]

The \( C \) matrix in use here will pick off the \((2,2)\) element of \((X'S^{-1}X)^{-1}\) as the variance for the test. It is apparent that the \((2,2)\) element of \((X'S^{-1}X)^{-1}\) will be algebraically difficult unless the \(z_i\) are all equal; if the \(z_i\) are all equal, all the off diagonal elements of \((X'S^{-1}X)\) in the first and second rows will be zero and thence the \((2,2)\) element of \((X'S^{-1}X)^{-1}\) will be just \[\frac{1}{n/z_1}\]. From (1.3.5), the vector of the eight cell means.
\[ X' S^{-1} X = \begin{bmatrix} \sum_{i=1}^{8} \frac{n_f_i}{z_i} \\
\sum_{i=1}^{4} \frac{n_f_i}{z_i} - \sum_{i=5}^{8} \frac{n_f_i}{z_i} \\
n_f_1/z_1 - n_f_4/z_4 + n_f_5/z_5 - n_f_8/z_8 \\
n_f_2/z_2 - n_f_4/z_4 + n_f_6/z_6 - n_f_8/z_8 \\
n_f_3/z_3 - n_f_4/z_4 + n_f_7/z_7 - n_f_8/z_8 \end{bmatrix} \]

Again, if the \( z_i \) are all equal to \( z \),

\[ C \beta = C (X' S^{-1} X)^{-1} X' S^{-1} F \]

\[ = (8n/z)^{-1} \left( n_f_1 + n_f_3 + n_f_4 - n_f_5 - n_f_6 - n_f_7 - n_f_8 / z \right) \]

\[ = \left( \text{level 1 mean} - \text{level 2 mean} \right) / 2. \]

Hence,

\[ SS(\beta = 0) = \frac{1}{4} \left( \text{level 1 mean} - \text{level 2 mean} \right)^2 \left( \sum_{i=1}^{8} \frac{1}{n_i} \right)^{-1} \]

\[ = \left( \text{level 1 mean} - \text{level 2 mean} \right)^2 \left( 2n/z \right). \]

But \( (2n/z) = (z/2n)^{-1} = (z/4n + z/4n)^{-1} \)

So \( SS(\beta = 0) = \left( \text{level 1 mean} - \text{level 2 mean} \right)^2 / (z/4n + z/4n). \) \hspace{1cm} (2.2.14)

which is exactly the same test statistic as obtained in the one-way classification case, at (2.2.10), under the assumptions of equal sample sizes and variances, since \( n \) here corresponds to \( n/4 \) at (2.2.6).

The four level factor test is orthogonal to the two level test, so it follows through to yield the same statistic as obtained at (2.2.3) in the one-way classification case, and so, under the null hypothesis, the cross classification model yields exactly the same test statistics as does the one-way classification. We should expect that the mean of the sample variances here will be still smaller than the corresponding mean in the one-way classification case, since the population is broken
into more samples.

2.3 The Moments of the Kruskal-Wallis H Statistic

For the ties-absent case, Kruskal (1952) found the expectation of the Kruskal-Wallis H statistic under the null hypothesis to be (s-1), where S is the number of samples under analysis. He then obtained the null hypothesis variance of H for the multiple sample case:

\[ \text{Var}(H) = 2(S-1) - 2/5N(N+1)[3S^2 - 6S + N(2S^2 - 6S + 1)] - 6/5 \sum_{i=1}^{s} 1/n_i. \]  \hspace{1cm} (2.3.1)

He noted that as \( N \rightarrow \infty \), \( \text{Var}(H) \rightarrow 2(s-1) \), the variance of a central chi square with (s-1) degrees of freedom.

In the ties-present case, Kruskal showed that the expectation of H is again (s-1), but did not find the variance of H. The variance of H will be useful in later work, so an expression for it was derived; the derivation was based on Kruskal's derivation of \( \text{Var}(H) \) in the ties-absent case, and employed some results of Klotz (1966) on the moments of the Wilcoxon test in the ties-present case. The derivation is shown in full in Appendix 2. A representation of the variance is

\[ \text{Var}_c(H) = \frac{144}{N^2(N+1)^2} \left\{ \sum_{i=1}^{S} \frac{1}{n_i^2} \sum_{j=1}^{S} \frac{1}{n_j} \cdot \text{ER}_{ij}^4 \right\} \left\{ \sum_{i=1}^{S} \frac{1}{n_i} \cdot \text{ER}_i^2 \right\} - \left\{ \sum_{i=1}^{S} \frac{1}{n_i} \cdot \text{ER}_i \right\}^2, \]  \hspace{1cm} (2.3.2)

where the elements \( \sum_{i=1}^{S} \text{ER}_i^4 \), etc., are evaluated in Appendix 2. The variance is conditional on both the numbers and locations of the ties.

Because of the length and complexity of calculation involved in computing a variance through the use of (2.3.2), a FORTRAN computer program for evaluating \( \text{Var}_c(H) \) was prepared. The program computes the conditional (or unconditional) variance of H in the ties-present
(or absent) case, given the number of samples, $S$, the number of responses, $A$, the total sample size, $TN$, the vector of numbers of ties, $T_j^l$, and the sample sizes, $N(I)$, by sample. The program was written to be run on a Call-A-Computer terminal; a listing of the program is given in Appendix 3.

A number of test cases were run with the program. The program was first checked for agreement with Kruskal's variance for the ties-absent case, and in all cases, exact agreement was found. (This is a trivial result of all the difficult parts of the expression for $\text{Var}_C(H)$ going to zero.) The critical test of the program was in evaluating the conditional variances for Lehman's (1961) five completely enumerated distributions of the Wilcoxon statistic. These distributions arose from five data sets, each of which had two samples of size five each, but which had varying numbers of tied observations. For each of the five completely enumerated distributions of the Wilcoxon statistic, the Kruskal-Wallis $H$ statistics were computed for each of the possible outcomes. The variances of these $H$ statistics were then computed and checked against the output of the conditional variance program. The comparison was as follows:

<table>
<thead>
<tr>
<th>Case</th>
<th>Exact Variance Based on Lehman's Complete Enumerations</th>
<th>Variance found by Program</th>
<th>Variance by Kruskal's Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.6081</td>
<td>1.6080</td>
<td>1.6291</td>
</tr>
<tr>
<td>2</td>
<td>1.5987</td>
<td>1.5988</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>1.6214</td>
<td>1.6210</td>
<td>&quot;</td>
</tr>
<tr>
<td>4</td>
<td>1.6249</td>
<td>1.6251</td>
<td>&quot;</td>
</tr>
<tr>
<td>5</td>
<td>1.6339</td>
<td>1.6339</td>
<td>&quot;</td>
</tr>
</tbody>
</table>
Because there were relatively few ties in the Lehman data, the variance as evaluated by Kruskal's formula for the ties-absent was also computed, and we note rather close agreement everywhere. Because there were no more complex test cases available, no more direct checking could be done. Hence, checking was done by assuring that: (1) The response vector \((n_{.1}, n_{.2}, \ldots, n_{.r})\) produced the same variance as \((n_{.r}, n_{.r-1}, \ldots, n_{.1})\), and (2) making the sample sizes slightly unequal produced only slight changes in the variance.

The large sample moments of the bivariate Kruskal-Wallis statistic in untied data are given by Puri and Sen (1971). They showed that the bivariate test statistic is asymptotically chi-square distributed with \(2(s-1)\) degrees of freedom, where \(s\) is the number of samples. However, to the best of our knowledge, the moments of the bivariate Kruskal-Wallis statistic, when applied to small samples of tied data, are unknown, so a computation of the bivariate \(\chi^2_L\) test as suggested by Williams and Grizzle (1970) was followed algebraically in order to obtain an expression for the bivariate test statistic. Equal variances within each response variables were used, and equal sample sizes assumed, so the resulting univariate components of the bivariate test were exactly \(HN/(N-1)\), aside from covariance terms. (In further work we will refer to this formulation of \(\chi^2_L\) as \(H'N/(N-1)\).

It was found that

\[
\chi^2_L \text{ (bivariate)} = \frac{[\text{Var } 2(\text{Contrast } 1)^2 + \text{Var } 1(\text{Contrast } 2)^2 - 2 \text{ Cov}(\text{Contrast } 1)(\text{Contrast } 2)]}{[\text{Var } 1 \text{ Var } 2 - (\text{Cov})^2]} .
\]

(2.3.3)

where \(\text{Contrast } j\) is a contrast among the sample averages for for response \(j\), \(j = 1, 2\).
Var \( j \) is the observed variance of the \( j^{th} \) contrast, \( j = 1, 2 \), and Cov is the observed covariance between responses 1 and 2.

If the covariance term is negligible, then the bivariate test statistic is approximately the sum of the first two terms on the right side of (2.3.3). After cancelling, these two terms are approximately the two univariate tests for the effects of the factor on the first and second responses.

Therefore, it appears that in view of Puri and Sen's results for the asymptotic case and the above, a reasonable approximation for the expectation of the bivariate Kruskal-Wallis statistic is the sum of the expectations of the corresponding univariate statistics. If the response variables are independent, then the conditional variance of the bivariate statistic should be well approximated by the sum of the conditional variances of the univariate statistics.

2.4 The Moments of the \( \chi^2_{\ell} \) Statistic

Because there is a relationship between the Kruskal-Wallis \( H \) statistic and the modified chi-square statistic \( \chi^2_{\ell} \), we can gain some insight concerning the moments of \( \chi^2_{\ell} \). For the simplest two sample case, recalling equation (2.2.11) we write

\[
H = \chi^2_{\ell} \times (N-1)/N \times \text{Term 1}.
\] (2.4.1)

We know the first two moments of \( H \), conditional on the number of ties in the data, so the corresponding moments of \( \chi^2_{\ell} \) follow, except that Term 1 of (2.4.1) is a function of random variables. If we fix Term 1 to condition upon it, then we also fix \( H \), because \( H \) is uniquely determined by the elements of Term 1. However, because of the
overall variance formulation of $H$, $\chi^2_{\ell}$ is not uniquely determined by the elements of $H$.

Recalling the work of Section 2.2 which indicated that $\chi^2_{\ell} > HN/(N-1)$, and the structures of the test statistics which indicated that the distribution of $\chi^2_{\ell}$ will have a much longer right tail than the distribution of $H$, it appears that the moments of $H$ will provide approximate lower limits for the moments of $\chi^2_{\ell}$. Little else can be said directly, since the non-random variable relationship between $\chi^2_{\ell}$ and $H$ that we have holds exactly only when both are zero. There is a similar situation in the multiple sample case.

Suppose now that we change the formulation of $\chi^2_{\ell}$ by multiplying it by $(N-1)/N$ and by basing the variance term on the total sample response marginal rather than pooling the within-sample variances. Calling this revised statistic $H'$, we note that for the one-way classification case and the balanced cross classification cases, the work in Section 2.2 indicates that the moments of $H'$ would be equal to those of $H$, since the test statistics are identical. In the unbalanced cases, we could reasonably hope for some averaging to keep the moments of $H'$ near the moments of $H$. 
CHAPTER III

SIMULATION STUDIES

3.1 Introduction

In order to delve further into the small sample properties of the linear model approach to the analysis of ordinal response categorical data, it is now necessary to turn to computer simulation of the test statistic. We do so for the following reasons:

1. From Section 2.2 we know that $\chi^2\ell$ is related to the $H$ statistic, but the properties of the $H$ statistic in the presence of ties are essentially unknown for the small sample case.

2. We should find $\chi^2\ell > HN/(N-1)$, caused by the differences in their variances discussed in Section 2.2, but we do not know if this is of practical importance.

3. Williams and Grizzle (1970) suggest the use of multivariate methods in analyzing multi-response categorical data problems. We do not know the small sample properties of these tests.

To answer the above questions, a computer simulation program with the following characteristics appeared necessary:

1. The program should employ a random number generator to create data with known properties, and the program should be able to efficiently replicate a desired situation many times.

2. There should be sufficient flexibility in the program so that
one could vary sample sizes, numbers of samples, forms of response
distributions, and experimental structures through inputs to the
program alone.

(3) The program should have the capability of handling bivariate
problems.

(4) The program should be able to summarize its results and
require little input to produce much output.

Hence, with the general goals in mind, a computer program was
prepared, input factors and their levels selected, and simulations
performed, first using $\chi^2_L$ as the test statistic and later using $H'$
as the test statistic.

3.2 The Computer Simulation Program

3.2.1 The Program and Its Workings

The computer program which was prepared is an extensively modified
version of the LINCAT program of Forthofer, which is described in
Forthofer, Starmer, and Grizzle (1969). The program generates pseudo-
random data, analyzes the data and summarizes the results in the follow-
ing manner:

With the use of the well known IBM Scientific Subroutines Package
Subroutine RANDU, and the IBM 370/165 computer at the Triangle Univer-
sities Computation Center, random uniform numbers in the interval $(0, 1)$
are generated. These numbers are classified into $r_1 \times r_2$ tables within
each of $s$ samples, where $r_1$ is the number of levels or categories of
response 1, and $r_2$ is the number of levels or categories of response
2. The cell boundaries in the $r_1 \times r_2$ tables are chosen to reflect the
probabilities of observations falling into the various cells, given the
exact distributions of the response variables. The numbers of samples and their numbers of observations are controlled by the input, also. Marginal rank scores are formed by collapsing the \( r_1 \times r_2 \times s \) table over \( r_2 \) and \( s \) for the \( r_1 \) marginal and \( r_2 \) and \( s \) for the \( r_2 \) marginal. The program then sets the marginal scores and generated data into the form used as input to the LINCAT program and proceeds with the linear model analysis.

To perform the linear model analysis, the program accepts a reparametrized \( \mathbf{X} \) (i.e., experimental design) matrix, and a series of \( \mathbf{C} \) (i.e., contrast) matrices. The \( \mathbf{X} \) matrix is read so as to be appropriate for testing a bivariate hypothesis as suggested in Williams and Grizzle (1970). For instance, a 2 level factor \( \times \) 4 level factor problem would perhaps have as the \( \mathbf{X} \) matrix:

\[
\begin{bmatrix}
\mu_1 & \mu_2 & \alpha_1 & \alpha_2 & \beta_{11} & \beta_{12} & \beta_{21} & \beta_{22} & \beta_{31} & \beta_{32} \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 \\
0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 \\
1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 \\
0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & -1 \\
\end{bmatrix}
\]

\[X_1 = \quad (3.2.1)\]
where the column headings $\mu_j$, $\alpha_j$ and $\beta_{ij}$ are included to illustrate bivariate classification of an observation. This type of $X$ matrix is discussed fully in Williams and Grizzle. The first group of $C$ matrices would be those for performing bivariate tests of homogeneity among the levels of the factor variables; e.g., a bivariate test for the $A$ effect requires in this case the $C$ matrix

$$C_1 = \begin{bmatrix}
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix} \quad (3.2.2)$$

After the bivariate tests are completed, the program begins univariate testing for effects of the factor variables on the response variables. If there are equal sample sizes and equal variances in either the cross classification or one-way classification models, then we can obtain univariate tests for factor effects by deleting rows of the $C$ matrices such as (3.2.2). However, if the equal variance and/or equal sample size restrictions are not met, then the $(X'S^{-1}X)$ matrix will not be block diagonal, and the elements of its inverse will be functions of both response variables. Hence, in the general case, even if the correlation between the two responses is zero, the information concerning the two responses is inextricably intermingled and makes clean univariate tests impossible without reformulating the problem. Therefore, to allow study of the unequal sample size case, we turn to the following approach:

The program collapses the $X$ matrix given at (3.2.1) into one consisting of the alternate elements of the alternate rows of (3.2.1), i.e.,
\[
\begin{bmatrix}
1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 \\
1 & -1 & 0 & 0 \\
1 & -1 & -1 & -1 \\
\end{bmatrix}
\]

(3.2.3)

Selecting the appropriate elements of \((A'V A)\) and \(F(p)\), the program reads the next group of \(C\) matrices. These contrast matrices are set up to test univariate hypotheses of homogeneity concerning response 1 only. An example of the \(C\) matrix required to test for a univariate A factor effect is

\[
C_2 = [0 \ 1 \ 0 \ 0 \ 0] .
\]

(3.2.4)

When the testing concerning the first response variable is complete, the program repeats the univariate analysis on the second marginal, employing the \(X_2\) matrix given at (3.2.3) and the same \(C\) matrices as illustrated in (3.2.4) but using different elements of \(F(p)\) and \((A'V A)\), of course.

To summarize, the above data processing produces (1) A series of bivariate tests of homogeneity of the factor variables with respect to the two response variables and an associated test for lack of fit of the model depicted by the \(X\) matrix.

(2) Two series of univariate tests, one for each response, and their associated lack of fit tests, which in this two factor situation may be interpreted as a test for interaction.

The data generation-analysis procedure can be repeated as many times as desired for a particular set of cell probabilities, an \(X\) matrix, and set of \(C\) matrices. The program employs the same subroutines as the LINCAT program, with the following two exceptions:
(1) this program requires the random number generator RANDU, and
(2) this program makes use of the fact that $A'V_A$ is a block diagonal matrix (in this bivariate case, it has $2 \times 2$ blocks on the diagonal) and inverts $A'V_A$ within the main program according to the simpler rules for inverting block diagonal matrices, rather than calling a general inversion routine. Two major portions of the LINCAT program have been eliminated in this program; the log model option and the modified chi-square estimation section. Neither the log model nor the modified chi-square estimation options are appropriate to the problem considered here and the modified chi-square estimation elimination results in considerable time savings. (Perhaps ten seconds in compilation and two seconds per replication in the eight factor combination case.)

When the desired number of replications of a particular experimental situation have been simulated and analyzed, the averages and variances of all the test statistics and a number of estimated model parameters and variances are printed out.

The program as used in production, with a 3 category response $\times$ 5 category response $\times$ 8 sample model took approximately 20 seconds to compile and then one minute to perform 40 simulations. Storage requirements were 266 K bytes. We could have performed these simulations using somewhat less storage; the dimensioning is aimed at a maximum problem of two responses with 24 factor combinations.

3.2.2 Test Cases

A number of test cases were run to assure this program would produce the same results as the LINCAT program; in the series of 12
problems listed below the only numerical discrepancy noted was of one digit in the sixth decimal place in one instance. This program also reacts differently, but with the same net results, to a singular $A'VA$ matrix than does LINCAT, since $A'VA$ is inverted differently here. The test cases actually run are as follows: (Each case had two or more problems, and the experimental situation is denoted by $r_1 \times r_2 \times f_1 \times f_2$ where $r_j$ is number of categories of response $j$ and $f_i$ the number of levels of factor $i$.)

(i) $3 \times 3 \times 2 \times 2$ with an average of 20 observations per treatment combination

(ii) $3 \times 3 \times 2 \times 4$ with 3 observations per treatment combination

(iii) $3 \times 5 \times 2 \times 6$ with 6 observations per treatment combination

(iv) $3 \times 5 \times 2 \times 4$ with 6 observations per treatment combination

(v) $3 \times 3 \times 1 \times 4$ with 100 observations per treatment level

3.3 The Variables of Interest and Analytical Framework of the Simulation Studies

The selection of the variables and their levels to be included in the simulation studies is a difficult task for it defines the width of applicability of the results, and is for the most part, arbitrary. Our justifications for selecting variables and their levels follow:

We believed that the variables which were represented in the theoretical work of Chapter 2 held the most promise of producing useful results, so we immediately had the following list:

(1) number of factor variables,

(2) number of levels of factor variables,
(3) sample sizes by combinations of levels of the factor variables, and
(4) some measure of the numbers of ties present in the data.
In addition, we felt we needed to consider the following:
(5) the distributions of the response variables (which can be used to control (4).),
(6) the numbers of categories of the response variables,
(7) the correlation between the response variables, and
(8) the number of replications in the simulation.
The choice of the number of factor variables followed from two considerations: (a) At least two factor variables are necessary in order to study interaction effects, and (b) for a two factor variable case a test for interaction is produced automatically as a test of lack of fit. Hence, rerunning the program with the interaction term included in the model is not necessary in order to get interaction effects adjusted for main effects and main effects not adjusted for interaction effect. Therefore, all simulations will be run with two factor variables.

The numbers of levels of the factor variables were chosen to be two and four; the choice of two levels was appropriate in view of the frequency of encounter of two sample problems and the more easily handled relationship between \( H \) and \( \chi^2 \) in the two sample case. The selection of four levels of the second factor variable was arbitrary; it is perhaps representative of the multi-level situation and yet reasonably economical to simulate.

Choosing the numbers of observations by treatment combinations depends on the numbers of categories of the response variables, so we
next considered the response variables. Assuming that a response with eight ordinal categories would produce results quite similar to those of a continuous response, and since three category ordinal responses are the most frequently encountered in practice, e.g., deterioration, no change, improvement, we settled on three categories and five categories. Because we were interested in small sample properties, we selected three observations per treatment combination and six observations per treatment combination, with twelve observations per treatment combination as an alternative in the event of failure in the three observations per treatment combination case. To answer questions concerning behavior in the case of unequal sample sizes, we decided to run a group with equal sample sizes throughout and another group with unequal sample sizes. The inequalities of sample sizes were aimed at one sample being one-half of another, subject to integer sample size restrictions and avoidance of a single observation in a treatment combination.

The next considerations were the distributions of the response variables. The easiest distribution to deal with was the uniform, but the uniform is also perhaps the least informative, so we elected to simulate Bell shaped, U shaped and J shaped response variable distributions. The percentage points of the response distributions are shown in Table (3.3.1).
Table 3.3.1

ASSUMED DISTRIBUTIONS OF RESPONSE VARIABLES

<table>
<thead>
<tr>
<th>Number of Response Categories</th>
<th>Cell Probabilities</th>
<th>Response Distribution Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \pi_1 )</td>
<td>Bell</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.2525</td>
</tr>
<tr>
<td></td>
<td>( \pi_2 )</td>
<td>.4950</td>
</tr>
<tr>
<td></td>
<td>( \pi_3 )</td>
<td>.2525</td>
</tr>
<tr>
<td></td>
<td>( \pi_1 )</td>
<td>.1151</td>
</tr>
<tr>
<td></td>
<td>( \pi_2 )</td>
<td>.2942</td>
</tr>
<tr>
<td></td>
<td>( \pi_3 )</td>
<td>.1872</td>
</tr>
<tr>
<td></td>
<td>( \pi_4 )</td>
<td>.0950</td>
</tr>
<tr>
<td></td>
<td>( \pi_5 )</td>
<td>.0550</td>
</tr>
</tbody>
</table>

Because the degree of correlation between the response variables is of interest primarily from the bivariate testing viewpoint, and we did not know what would be appropriate levels of correlation to induce, we elected to have the response variables independent.

The last variable considered was the number of replications of the simulation at each point. Because (a) our purpose in simulation was partially verification and partially establishment of results, (b) each point yielded information on two factors relative to two responses, and (c) there should be relatively close agreement between the results obtained at the different points, we settled on 40 replications per point. Additional independent simulations could be added relatively easily if needed.
In summary, the framework of the simulation experiment is as follows:

(i) Two independent response variables, with three and five categories.

(ii) Response variables arising from Bell, J, and U shaped distributions.

(iii) Two factor variables, at two and four levels, yielding eight treatment combination cells.

(iv) The sample sizes, by treatment combination, for the eight treatment combination cells:

\[
\begin{align*}
6 & \quad 6 & \quad 6 & \quad 6 & \quad 6 & \quad 6 & \quad 6 & \quad 6 & \quad \text{equal} & \quad \{6 \text{ observations per treatment combination}\} \\
6 & \quad 6 & \quad 6 & \quad 6 & \quad 5 & \quad 5 & \quad 5 & \quad 3 & \quad \text{unequal} & \quad \{ \text{treatment combination}\} \\
3 & \quad 3 & \quad 3 & \quad 3 & \quad 3 & \quad 3 & \quad 3 & \quad 3 & \quad \text{equal} & \quad \{3 \text{ observations per treatment combination}\} \\
5 & \quad 4 & \quad 4 & \quad 3 & \quad 2 & \quad 2 & \quad 2 & \quad 2 & \quad \text{unequal} & \quad \{ \text{treatment combination}\}
\end{align*}
\]

The variables sample size, sample size equality, and shape of response distribution form a $2 \times 2 \times 3$ factorial experiment with 40 replications per cell. Within each of the twelve cells, we obtain information on the effects of two factor variables on the two response variables through two univariate tests and one bivariate test.

The inputs to the program can now be specified completely. The $X$ matrix cited at (3.2.1) was used in each simulation, and the $C_1$ matrix (3.2.2) was used for the bivariate test for factor A effects. The $C$ matrix used for the bivariate test for B effects was

\[
C_2 = \begin{bmatrix} 0 & 1 \\ 6 \times 4 & 6 \times 6 \end{bmatrix}.
\]
The B effects were broken out into three tests according to the
contrast matrices

\[
\begin{align*}
C_3 &= \begin{bmatrix} 0 & 1 & 0 & -1 & 0 & 0 & 0 \end{bmatrix}^T, \\
C_4 &= \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & -2 & 0 \end{bmatrix}^T, \\
C_5 &= \begin{bmatrix} 0 & 4 & 0 & 4 & 0 & 4 & 0 \end{bmatrix}^T,
\end{align*}
\]

and the C matrices for the univariate tests for the B effects
follow from \( C_2 - C_5 \), as does \( (3.2.4) \) from \( (3.2.2) \), i.e., by deleting
alternate elements and alternate rows. The contrast \( C_5 \) arises from
the contrast \( B_1 + B_2 + B_3 \) vs \( 3B_4 \), when the reparametrization
\( B_1 + B_2 + B_3 + B_4 = 0 \) holds. The contrasts \( C_3, C_4 \) and \( C_5 \) are
not orthogonal, but any one of them has expectation 2.0 and they depict
the problem of finding where differences lie when an overall test of
homogeneity indicates differences are present.

3.4 Preliminary Simulations; Examination and Modification of \( \chi^2_\ell \)

The initial simulations performed were with \( \chi^2_\ell \) as the test sta-
tistic because our goals in these first few simulations were to see if
the magnitude of the difference between \( \chi^2_\ell \) and \( \text{HN/(N-1)} \) was of
practical significance and if so, to see if the difference persisted
for multiple samples and for large sample sizes. A summary of these
\( \chi^2_\ell \) simulations is presented in Table (3.4.1). We note in Table (3.4.1)
that only the eight category response (response 2) in the fourth trial,
with an average of 2.51, had a smaller average than expected under the
null hypothesis. The three and five category response variables had
more or less comparable excessive averages and variances everywhere.
### TABLE 3.4.1
SUMMARY OF SIMULATIONS Employing $X^2$

<table>
<thead>
<tr>
<th>Trial</th>
<th>Response</th>
<th>Responses, $n_i$ per Trt. Comb.</th>
<th>$N$ # Replications</th>
<th>Observed $^*$ Means and Variances of Univariate Test Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Design</td>
<td></td>
<td></td>
<td>Lack of Fit</td>
</tr>
<tr>
<td></td>
<td>Dist'n</td>
<td>Factors</td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>1</td>
<td>U</td>
<td>$3 \times 5 \times 2 \times 4$</td>
<td>6</td>
<td>48</td>
</tr>
<tr>
<td>2</td>
<td>Bell</td>
<td>$3 \times 5 \times 2 \times 4$</td>
<td>6</td>
<td>48</td>
</tr>
<tr>
<td>3</td>
<td>U</td>
<td>$3 \times 5 \times 2 \times 4$</td>
<td>12</td>
<td>96</td>
</tr>
<tr>
<td>4</td>
<td>Uniform</td>
<td>$5 \times 8 \times 1 \times 4$</td>
<td>100</td>
<td>400</td>
</tr>
<tr>
<td>5</td>
<td>Uniform</td>
<td>$3 \times 3 \times 1 \times 4$</td>
<td>100</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Expectation of Average $X^2$**

**Approximate Expected variance of $X^2$**

- * Observed means and variances have been adjusted for $N/(N-1)$ term
- $N = \sum n_i$, $n_i =$ sample size for $i^{th}$ treatment combination

---

*38*
The differences between the observed moments of \( \chi^2_{\ell} \) and the moments of \( H \) decreased as the sample sizes increased, but not quickly. Graphs of the cumulative distributions of the test statistics, when plotted against the appropriate \( \chi^2_{\ell} \), indicated serious anticonservativeness; a typical graph would show an observed \( \alpha \) level to be twice the nominal \( \alpha \) level for the nominal \( \alpha \) level equal 0.10 or less. In only a few instances did the observed cumulative distribution show a nominal or smaller significance level.

Because these preliminary simulations showed that in this situation the null distribution of \( \chi^2_{\ell} \) is not well approximated by \( \chi^2 \), and we had no theoretical basis for suggesting approximations to the null distribution of \( \chi^2_{\ell} \) which would perform significantly better, we chose to change the formulation of the test statistic from \( \chi^2_{\ell} \) to \( H' \). To do this, we changed the variance of \( \chi^2_{\ell} \) from one based on pooled within sample variances to one based on the response marginal for the total sample. This implied no further consideration of the \( \chi^2_{\ell} \) statistic, per se, and we turned to the \( H' \) formulation of the test statistic for further work. The simulation experiment specified in Section 3.3 was run to provide an information source for testing for the effects of the variable of interest. The results of these simulations will be analyzed in Chapter 4.
CHAPTER IV

ANALYSIS OF SIMULATION EXPERIMENT RESULTS,

NULL HYPOTHESIS CASE

4.1 Introduction

In this chapter, we shall examine the results of the $H'$ statistic simulation experiment outlined in Section 3.3. We shall examine the simulated test statistics by comparing (a) averages within the experiment, (b) observed averages with expected averages, (c) observed variances with expected variances, and (d) observed cumulative distributions with the appropriate $\chi^2$ cumulative distributions. From these comparisons we shall formulate answers to the following questions:

(1) Are there differences in the simulations caused by the variables of interest, e.g., sample size?

(2) How well do the means and variances of the simulated test statistics agree with their expectations?

(3) Does a $\chi^2$ distribution provide a reasonable approximation to the distribution of the $H'$ statistic when the null hypothesis is true? Where does it break down?

(4) Another question of interest from the standpoint of applying the $H'$ test is whether the $F$ distribution provides a better
approximation to the null distribution of the $H'$ statistic than does the $\chi^2$. To answer this question, we will review the method of Wallace (1952) for deriving the $F$ approximation to the $H$ statistic, adapt it for use here, and then compare the $\chi^2$ and $F$ approximations for four critical pooled samples of the simulated data.

In Section 4.4, we will summarize the analysis.

4.2 Analysis of the Simulation Study of $H'$

4.2.1 The Means of the Test Statistics

For each of the 12 response distribution $\times$ sample size $\times$ equality of sample size combinations described in Section 3.3, the averages of each of the 18 test statistics were computed and are presented in Table (4.2.1), together with their expectations. To assess the effects the design variables, namely, sample size, response distribution, and equality of sample size, had on the averages of the 18 test statistics, a multivariate analysis of variance was run on the simulated test statistics. The model assumed was

$$E(Y) = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \epsilon_{ijk},$$

where $\alpha_i$ is the effect of differences in the response distributions,
$\beta_j$ is the effect of sample sizes,
$\gamma_k$ is the effect of inequality of sample sizes,
$\alpha\beta_{ij}$ is the effect of the response distribution by sample size interaction,
$\alpha\gamma_{ik}$ is the effect of the response distribution by inequality of sample size interaction,
$\beta\gamma_{jk}$ is the effect of the sample size by inequality of sample
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of Fit</td>
<td>5.24 5.37 5.56 7.08</td>
<td>5.81 6.23 5.22 6.33</td>
<td>5.80</td>
<td>3.91</td>
<td>6.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>2.12 1.86 2.09 1.77</td>
<td>1.50 1.96 2.02 1.90</td>
<td>2.14 2.17 2.28 2.12</td>
<td>1.99</td>
<td>2.03</td>
<td>2.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>6.04 5.73 5.70 6.37</td>
<td>6.10 6.25 5.27 5.30</td>
<td>5.58 6.56 6.16 5.51</td>
<td>6.04</td>
<td>3.72</td>
<td>6.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>1.90 1.95 2.21 2.13</td>
<td>2.06 1.79 1.81 1.81</td>
<td>1.73 2.06 2.16 1.93</td>
<td>1.92</td>
<td>2.01</td>
<td>2.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B2</td>
<td>2.00 1.53 1.65 1.75</td>
<td>2.20 2.33 1.91 1.42</td>
<td>1.98 2.73 2.05 1.90</td>
<td>2.13</td>
<td>1.78</td>
<td>2.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B3</td>
<td>2.13 2.26 1.83 2.48</td>
<td>1.85 2.15 1.55 1.96</td>
<td>1.87 1.76 1.96 1.72</td>
<td>2.00</td>
<td>1.92</td>
<td>2.00</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Lack of Fit | 2.90 2.93 2.57 2.95 | 2.65 3.16 2.86 4.33 | 2.67 3.08 2.49 2.98 | 2.90 | 3.03 | 3.00 |
| A | 1.09 0.89 1.09 0.94 | 0.72 0.83 1.09 1.05 | 1.26 1.20 1.14 1.23 | 1.00 | 1.09 | 1.00 |
| B | 2.71 3.38 3.31 3.35 | 2.99 2.81 2.48 2.38 | 3.02 3.58 2.23 2.35 | 3.08 | 2.87 | 3.00 |
| B1 | 0.69 1.21 1.32 1.11 | 1.03 0.71 0.85 0.86 | 0.99 1.19 0.93 1.01 | 0.97 | 1.01 | 1.00 |
| B2 | 0.98 0.79 1.05 0.94 | 1.10 0.87 0.78 0.70 | 1.01 1.43 0.95 1.12 | 1.03 | 0.92 | 1.00 |
| B3 | 1.04 1.37 0.94 1.31 | 0.86 1.22 0.85 0.78 | 1.02 0.97 0.84 0.87 | 1.08 | 0.93 | 1.00 |

| Lack of Fit | 2.88 3.43 3.00 2.54 | 2.58 2.21 2.70 2.75 | 3.14 3.15 2.73 3.54 | 2.91 | 2.88 | 3.00 |
| A | 1.23 0.98 1.01 0.84 | 0.78 1.13 0.93 0.85 | 0.88 0.98 1.14 0.89 | 1.00 | 0.94 | 1.00 |
| B | 3.33 2.35 2.39 3.02 | 3.11 3.44 2.79 2.93 | 2.56 2.98 3.44 2.53 | 2.96 | 2.85 | 3.00 |
| B1 | 1.21 0.74 0.90 1.02 | 1.03 1.08 0.96 0.95 | 0.74 0.86 1.22 0.91 | 0.94 | 0.99 | 1.00 |
| B2 | 1.03 0.74 0.61 0.81 | 1.09 1.46 1.13 0.73 | 0.97 1.30 1.09 0.78 | 1.10 | 0.86 | 1.00 |
| B3 | 1.09 0.89 0.89 1.18 | 0.99 0.92 0.70 1.18 | 0.85 0.79 1.12 0.85 | 0.92 | 0.99 | 1.00 |

* Expectations are approximate
size interaction, and

c

\[ c_{ijk} \]

is the error term, which is composed of residual variability and within cell variability, since there were 12 cells with 40 replications each.

The findings of this analysis may be summarized as follows:

(i) For each of the 18 responses, the averages for the three response distributions (Bell, J, and U) were not significantly different.

(ii) For each of the 18 responses, the averages for the two sample size levels (6/Trt. Comb. and 3/Trt. Comb.) were not significantly different.

(iii) For the bivariate lack of fit test and the univariate lack of fit test for the three category response, inequality of sample size had a significant effect. Since the bivariate test is approximately the sum of the two univariate tests, the difference can be attributed to the univariate test alone. For no other response was there a significant effect caused by inequality of sample size.

(iv) None of the interaction terms had significant effects on any of the responses.

For an answer to the part of the second question of Section 4.1 which deals with the agreement between the averages of the test statistics and their expectations, we turn again to Table (4.2.1), where the averages and expectations are presented. To obtain the expectations shown in Table (4.2.1), the following argument was used: The expectation of the \( H \) statistic is known. Because \( H \) and \( H' \) have exactly the same formulation for the one-way classification and \( H' \) reduces to
H in the balanced cross classification case, the moments of \( H' \) are equal to the moments of \( H \) for these cases. Because our linear models analysis here is analogous to the linear models analysis in the continuous case, in the unbalanced cross classification case we expect the distribution of \( H' \) to be similar to the distribution of \( H \), and thus the moments of \( H \) should provide good approximations for the moments of \( H' \). Hence, for the four principal univariate test statistics, i.e., tests A and B for the three category and for the five category responses, the expectations of the \( H' \) test statistics are 1.0, 3.0, 1.0, and 3.0 respectively. The work on the formulation of the bivariate \( H' \) statistic in Section 2.3 indicates that the expectations of the bivariate tests for A and B have expectations 2.0 and 6.0 respectively. Further, the breakdown of Test B into contrasts produces tests which are approximately two sample \( H \) statistics, thus each of their expectations are approximately 1.0 and the corresponding bivariate tests then each have expectation equal 2.0. The lack of fit, or interaction, statistic has asymptotically a central \( \chi^2 \) distribution with 3 degrees of freedom, and, assuming independence of response variables, the bivariate interaction test has expectation equal to 6.

Confidence bounds about the Table (4.2.1) averages of the 18 test statistics in each of the 12 factor combinations were computed. The bounds were of the form \( \bar{H'} + 2\sigma/\sqrt{40} \) where \( \sigma \) is obtained from the appropriate Table (4.2.2) or Table (4.2.3) expected variance. The expected variances were used rather than the observed variances in order to remove some of the dependency between the observed means and variances of the statistics. None of the expectations of the 18 test statistics in any of the 12 factor combinations fell outside the 2\( \sigma \)
limits about the observed averages. Further, the averages of the \( H' \) statistics, pooled over equality of sample size status and response distributions, were computed and are presented as Overall \( H' \) Averages in Table (4.2.1). Most of these averages of 240 simulations compare closely with their expectations, and none are further than \( 2\sigma/\sqrt{240} \) from their expectations.

The three contrasts within the levels of the \( B \) factor yielded statistics which approximately added to the overall \( B \) statistic in the equal samples case and almost as closely added to the overall \( B \) statistic in the unequal samples case. Since the three tests behaved similarly, their performance is basically a reflection of the performance of the overall \( B \) test.

4.2.2 The Variance of the Test Statistics

To investigate the agreement between the variances of the test statistics and their expectations, we first needed to compute the expected variances of the test statistics. Hence, within each of the 12 response distribution \( \times \) sample size \( \times \) equal of sample size combinations that were simulated, the conditional variances of \( H' \) were computed, using the computer program mentioned in Section 2.3. The conditional variances were computed for each of the 4 principal tests, i.e., tests A and B, for responses 1 and 2, for each of the 40 replications. The averages of the conditional variances over the 40 replications were then computed, yielding an estimate of the unconditional variances of the four principal test statistics for each of the 12 factor combinations. These average \( \text{Var}_c(H') \) are shown in Table (4.2.2). The variability among the conditional variances, which was
<table>
<thead>
<tr>
<th>Response</th>
<th>n4 per Trt. Comb.</th>
<th>Sample Equality</th>
<th>A (2 Sample)</th>
<th>B (4 Sample)</th>
<th>A (2 Sample)</th>
<th>B (4 Sample)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Observed</td>
<td>Expected</td>
<td>Observed</td>
<td>Expected</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bell</td>
<td>3</td>
<td>Equal</td>
<td>2.50</td>
<td>1.83</td>
<td>4.64</td>
<td>5.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unequal</td>
<td>0.95</td>
<td>1.81</td>
<td>7.18</td>
<td>4.99</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Equal</td>
<td>1.19</td>
<td>1.92</td>
<td>5.18</td>
<td>5.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unequal</td>
<td>1.31</td>
<td>1.91</td>
<td>7.53</td>
<td>5.48</td>
</tr>
<tr>
<td>J</td>
<td>3</td>
<td>Equal</td>
<td>0.34</td>
<td>1.84</td>
<td>5.11</td>
<td>5.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unequal</td>
<td>0.91</td>
<td>1.83</td>
<td>6.01</td>
<td>5.06</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Equal</td>
<td>1.76</td>
<td>1.93</td>
<td>4.71</td>
<td>5.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unequal</td>
<td>2.02</td>
<td>1.91</td>
<td>2.32</td>
<td>5.52</td>
</tr>
<tr>
<td>U</td>
<td>3</td>
<td>Equal</td>
<td>1.76</td>
<td>1.89</td>
<td>4.12</td>
<td>5.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unequal</td>
<td>2.70</td>
<td>1.86</td>
<td>6.47</td>
<td>5.15</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Equal</td>
<td>1.95</td>
<td>1.95</td>
<td>4.77</td>
<td>5.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unequal</td>
<td>3.18</td>
<td>1.93</td>
<td>5.62</td>
<td>5.55</td>
</tr>
<tr>
<td>Pooled</td>
<td>3/Trt. Comb.</td>
<td></td>
<td>1.53</td>
<td></td>
<td>5.56</td>
<td></td>
</tr>
<tr>
<td>Variances</td>
<td>6/Trt. Comb</td>
<td></td>
<td>1.90</td>
<td></td>
<td>4.69</td>
<td></td>
</tr>
</tbody>
</table>
induced by the varying numbers of ties, is quite small; very few fell further from their mean than $\pm 0.05$, most were within $\pm 0.01$ of their mean, and a typical variance of the conditional variances was $10^{-4}$. Hence, the averages of conditional variances are essentially constant for a given population size, set of sample sizes and number of samples. Examining the average observed conditional variances in Table (4.2.2), we note that:

(i) The average conditional variance rises sharply with an increase in the number of levels of the factor variables.

(ii) The average conditional variance is slightly larger for equal sized samples than for unequal sized samples.

(iii) The Bell shaped distribution averages are smaller than those of the J shaped distribution, which in turn are smaller than those of the U shaped distribution.

(iv) The average conditional variances for the five response category variable are larger than the three response category variable averages for the Bell and smaller sample J combinations; for the larger sample J and U distribution combinations the opposite is true.

These findings are in keeping with what one would expect upon examining the formulation of the variance of $H'$ in the ties-absent case, and considering the numbers of tied observations arising from the three response distributions. Further, since the differences (i), (ii) and (iii) are consistent across the Table, interaction of consequence among the factor variables is unlikely. The difference (iv) suggests an interaction between numbers of levels of the response variable and the shape of the response distributions. Having the expected variance in
hand, we compared observed and expected variances. The observed variances and pooled observed variances are also presented in Table (4.2.2). Assuming the \( H' \) statistics are \( \chi^2 \) distributed, approximate \( 2\sigma \) intervals were put about the observed variances. An estimate of \( \sigma \) was obtained through the approach suggested for nonnormal distributions in Section 31.7 of Kendall and Stuart (1967). They derived the asymptotic relationship

\[
V(s^2) = 2\kappa_2^2 \left(1 + \frac{1}{2\gamma_2}\right)/N,
\]

where \( \gamma_2 = 12/\text{(degrees of freedom)} \), and \( \kappa_2 = 2/\text{(degrees of freedom)} \).

All of the observed variances and pooled observed variances in Table (4.2.2) fell within \( \pm 2\sigma \) of their expected variances.

The observed variances and pooled observed variances for the interaction or lack of fit tests and one degree of freedom contrasts are shown in Table (4.2.3). The interaction test statistic is approximately \( \chi^2 \) distributed, so we took its expected variance to be twice the appropriate degrees of freedom. The one degree of freedom contrast statistics were also assumed to be approximately \( \chi^2 \) distributed, and their expected variances taken to be equal to 2.0. Approximate \( 2\sigma \) intervals such as described above were put about the observed variances and those cases where the intervals did not cover the expected variances are denoted by underlined entries in Table (4.2.3).

We note that 3 of the 96 univariate test observed variances shown in Table (4.2.3) were \( 2\sigma \) or further from their expected variances, while all of the pooled variances and 31 of the 32 bivariate test variances were within \( 2\sigma \) of their expectations. There is a suggestion that the three response category tests were more variable than the five response category tests, and the lack of fit tests were less
## Table 4.2.3

**Observed Variances of Simulated Test Statistics**

Response Variable Distribution

<table>
<thead>
<tr>
<th>Response, Test Statistic</th>
<th>J</th>
<th>U</th>
<th>Pooled Variances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bivariate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of Fit B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>2.77</td>
<td>4.20</td>
<td>3.53</td>
</tr>
<tr>
<td>B2</td>
<td>4.39</td>
<td>3.16</td>
<td>2.73</td>
</tr>
<tr>
<td>B3</td>
<td>3.63</td>
<td>5.27</td>
<td>3.45</td>
</tr>
<tr>
<td>Univariate (3 Categories)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of Fit B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.79</td>
<td>4.72</td>
<td>1.51</td>
<td>5.18</td>
</tr>
<tr>
<td>B1</td>
<td>1.12</td>
<td>2.98</td>
<td>2.34</td>
</tr>
<tr>
<td>B2</td>
<td>1.98</td>
<td>1.61</td>
<td>2.60</td>
</tr>
<tr>
<td>B3</td>
<td>2.04</td>
<td>2.83</td>
<td>1.29</td>
</tr>
<tr>
<td>Univariate (5 Categories)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of Fit B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.07</td>
<td>7.86</td>
<td>5.66</td>
<td>3.55</td>
</tr>
<tr>
<td>B1</td>
<td>1.84</td>
<td>1.03</td>
<td>1.39</td>
</tr>
<tr>
<td>B2</td>
<td>1.51</td>
<td>0.95</td>
<td>0.56</td>
</tr>
<tr>
<td>B3</td>
<td>1.89</td>
<td>1.49</td>
<td>1.65</td>
</tr>
</tbody>
</table>

* Underlined entry denotes that observed variance did not fall within ± 20% of expected variance.
variable than expected.

To assess the effects of ties on the variance of the $H'$ statistic, the unconditional or ties-absent variances were computed for each of the four principal test statistics within each of the 12 simulation experiment factor combinations. These variances, denoted by $\text{Var}(H')$, are presented in Table (4.2.4), together with the corresponding averages of the conditional variances, which are denoted by $\overline{\text{Var}}_c(H')$. There is close agreement between $\text{Var}(H')$ and $\overline{\text{Var}}_c(H')$ everywhere in the table; the unconditional variance is an underestimate for the Bell shaped response case and it is an overestimate for the $J$ and $U$ shaped response cases, but the differences are relatively slight.

Hence, we conclude that for the sample sizes, disparity of samples and parent distributions studied here, the unconditional variance gives a good estimate of the average conditional variance, since the average of the conditional variances is quite stable and near the unconditional variance.

4.2.3 The Observed Cumulative Distributions of the Test Statistics

To check the distributional form aspect of the adequacy of the $\chi^2$ approximation to the $H'$ statistic, the cumulative distributions of the test statistics were computed. First, they were computed pooled across response distributions and equality of sample size status, and then were computed separately for each of the 12 cells of the simulation experimental design. The observed cumulative distributions were then plotted against the appropriate $\chi^2$ cumulative. The pooled cumulative distributions for the two and four sample univariate tests and for the single degree of freedom contrast tests are shown in Figures (4.2.1)
Table 4.2.4
Averages of Conditional Variances, $\var(X')$, and Corresponding Unconditional Variances, $\var(H)$

<table>
<thead>
<tr>
<th>Response</th>
<th>$n_s$ per Trt. Comb.</th>
<th>Sample</th>
<th>A (2 Sample)</th>
<th>$\var(H')$</th>
<th>$\var(H)$</th>
<th>$\var(H')$</th>
<th>$\var(H)$</th>
<th>B (4 Sample)</th>
<th>$\var(H')$</th>
<th>$\var(H)$</th>
<th>$\var(H')$</th>
<th>$\var(H)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td></td>
<td>Equal</td>
<td>1.63</td>
<td>1.85</td>
<td>5.00</td>
<td>5.04</td>
<td>1.65</td>
<td>1.85</td>
<td>5.04</td>
<td>5.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unequal</td>
<td>1.61</td>
<td>1.82</td>
<td>4.99</td>
<td>5.03</td>
<td>1.82</td>
<td>1.82</td>
<td>5.03</td>
<td>5.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Equal</td>
<td>1.92</td>
<td>1.92</td>
<td>5.51</td>
<td>5.52</td>
<td>1.92</td>
<td>1.92</td>
<td>5.52</td>
<td>5.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unequal</td>
<td>1.91</td>
<td>1.91</td>
<td>5.48</td>
<td>5.50</td>
<td>1.91</td>
<td>1.91</td>
<td>5.50</td>
<td>5.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Equal</td>
<td>1.84</td>
<td>1.85</td>
<td>5.02</td>
<td>5.04</td>
<td>1.86</td>
<td>1.85</td>
<td>5.07</td>
<td>5.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unequal</td>
<td>1.83</td>
<td>1.82</td>
<td>5.06</td>
<td>5.03</td>
<td>1.83</td>
<td>1.82</td>
<td>5.06</td>
<td>5.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Equal</td>
<td>1.93</td>
<td>1.92</td>
<td>5.55</td>
<td>5.52</td>
<td>1.93</td>
<td>1.92</td>
<td>5.54</td>
<td>5.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unequal</td>
<td>1.92</td>
<td>1.91</td>
<td>5.52</td>
<td>5.50</td>
<td>1.92</td>
<td>1.91</td>
<td>5.51</td>
<td>5.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Equal</td>
<td>1.89</td>
<td>1.85</td>
<td>5.16</td>
<td>5.04</td>
<td>1.87</td>
<td>1.85</td>
<td>5.11</td>
<td>5.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unequal</td>
<td>1.86</td>
<td>1.82</td>
<td>5.15</td>
<td>5.03</td>
<td>1.84</td>
<td>1.82</td>
<td>5.10</td>
<td>5.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Equal</td>
<td>1.95</td>
<td>1.92</td>
<td>5.59</td>
<td>5.52</td>
<td>1.94</td>
<td>1.92</td>
<td>5.58</td>
<td>5.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unequal</td>
<td>1.93</td>
<td>1.91</td>
<td>5.55</td>
<td>5.50</td>
<td>1.92</td>
<td>1.91</td>
<td>5.52</td>
<td>5.50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
through (4.2.5). The bivariate test and both of the univariate tests for the effects of the two level factor and four level factor, by the 12 experimental design combinations, are shown in Figures (4.2.6) through (4.2.29). The pooled bivariate tests are not shown because the bivariate test is essentially the sum of the two corresponding univariate tests. One sample Kolmogorov-Smirnov tests were performed on each of the cumulative distributions, and we found no significant differences between the observed cumulatives and $\chi^2$ at the $\alpha=0.05$ level. Considering the figures separately, we note the following:

(1) Figure (4.2.1) indicates that over the whole simulation experiment the $H'$ comparison of the levels of the two level factor yields a statistic which is well approximated by $\chi^2(1)$.

(2) Figure (4.2.2) indicates that over the whole simulation experiment, the $H'$ statistic test for homogeneity among the levels of the four level factor is well approximated by $\chi^2(3)$.

(3) Figures (4.2.3), (4.2.4), and (4.2.6) indicate that the one degree of freedom contrast test statistics are well approximated by $\chi^2$ when pooled over sample size inequality and response distributions.

(4) Figures (4.2.5) through (4.2.29) show a considerable variability in the observed cumulative distributions of the bivariate and univariate test statistics across the 12 cells of the simulation experiment, but the cumulatives do not show patterns indicative of effects caused by the design variables.

4.3 The F Approximation of the Distribution of the $H'$ Statistic

4.3.1 Background for the Use of the F Approximation

The use of a Beta distribution to approximate the distribution of the Kruskal-Wallis $H$ statistic was first considered by Kruskal (cont'd. p.70)
FOREWORD FOR GRAPHS OF CUMULATIVE DISTRIBUTIONS

Four basic designs were simulated. These designs, with entries denoting the numbers of observations within the respective samples, are as follows:

TABLE 4.2.5

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>A2</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>B1</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>B2</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>B3</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>B4</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

In the following figures, "Two Sample Tests" will denote comparisons of the columns within Designs 1, 2, 3, or 4. "Four Sample Tests" will mean comparisons of the rows of each Design. "Equal Samples" will refer to Designs 1 or 3, while "Unequal Samples" will refer to Designs 2 or 4. The $B_1$, $B_2$, and $B_3$ contrasts will represent partial comparisons within the rows of the designs. For instance, Figure (4.2.3), $B_1$ contrast, N = 24, 3 Response Categories, Designs 1 and 2, will represent the cumulative distribution of 240 three category response case $H'$ statistics formed by the contrast vector (0 0 1 -1 0), and
thus are comparisons between rows 1 and 2 of Designs 1 and 2.
FIGURE 4.2.5
B3 CONTRAST A & D $\chi^2$ CUMULATIVES

Response
N Categories Designs
24 3 1 & 2
48 3 3 & 4 & 4
24 5 1 & 2
48 5 3 & 4 & 4

$\chi^2(1)$
(Each Cumulative Curve Represents 240 Simulations)

(* See Table (4.2.5), Page 53)
FIGURE 4.2.8
TWO SAMPLE 3 AND 5 RESPONSE CATEGORY BIVARIATE TESTS,
J RESPONSE DISTRIBUTION, EQUAL SAMPLES

Sample Sizes
\[ n_1 \ n_2 \ \text{Design}^a \]
- 12 12 1
- 24 24 3
- \( x^2(2) \)
(Each Cumulative Curve Represents 40 Simulations)

FIGURE 4.2.9
TWO SAMPLE 3 AND 5 RESPONSE CATEGORY BIVARIATE TESTS,
J RESPONSE DISTRIBUTION, UNEQUAL SAMPLES

Sample Sizes
\[ n_1 \ n_2 \ \text{Design}^a \]
- 8 16 2
- 16 32 4
- \( x^2(2) \)
(Each Cumulative Curve Represents 40 Simulations)

(* See Table (4.2.5), Page 53)
FIGURE 4.2.12
POUR SAMPLE 3 AND 5 RESPONSE CATEGORY BIVARIATE TESTS,
BELL RESPONSE DISTRIBUTION, EQUAL SAMPLES

FIGURE 4.2.13
POUR SAMPLE 3 AND 5 RESPONSE CATEGORY BIVARIATE TESTS
BELL RESPONSE DISTRIBUTION, UNEQUAL SAMPLES

Sample Sizes
\[ n_1 \quad n_2 \quad n_3 \quad n_4 \]
Design
\[ \text{Design} \]
\[ 6 \quad 6 \quad 6 \quad 6 \quad 1 \]
\[ 12 \quad 12 \quad 12 \quad 12 \quad 2 \]
\[ \chi^2(6) \] (Each Cumulative Curve Represents 40 Simulations)

Sample Sizes
\[ n_1 \quad n_2 \quad n_3 \quad n_4 \]
Design
\[ \text{Design} \]
\[ 5 \quad 6 \quad 6 \quad 7 \quad 2 \]
\[ 9 \quad 9 \quad 15 \quad 15 \quad 4 \]
\[ \chi^2(6) \] (Each Cumulative Curve Represents 40 Simulations)
FIGURE 4.2.14
FOUR SAMPLE 3 AND 5 RESPONSE CATEGORY BIVARIATE TESTS,
J RESPONSE DISTRIBUTION, EQUAL SAMPLES

Sample Sizes
\( n_1 \ n_2 \ n_3 \ n_4 \) Design
- \( n_1 = 6 \ n_2 = 6 \ n_3 = 6 \ n_4 = 1 \)
- \( n_1 = 12 \ n_2 = 12 \ n_3 = 12 \ n_4 = 3 \)
- \( x^2(6) \)
(Each Cumulative Curve Represents 40 Simulations)

FIGURE 4.2.15
FOUR SAMPLE 3 AND 5 RESPONSE CATEGORY BIVARIATE TESTS,
J RESPONSE DISTRIBUTION, UNEQUAL SAMPLES

Sample Sizes
\( n_1 \ n_2 \ n_3 \ n_4 \) Design
- \( n_1 = 5 \ n_2 = 6 \ n_3 = 7 \ n_4 = 2 \)
- \( n_1 = 9 \ n_2 = 9 \ n_3 = 15 \ n_4 = 4 \)
- \( x^2(6) \)
(Each Cumulative Curve Represents 40 Simulations)
FIGURE 4.2.20
TWO SAMPLE UNIVARIATE TESTS, J RESPONSE DISTRIBUTION
EQUAL SAMPLES

Sample Sizes Response
12 12 3 1
24 24 3 3
12 12 5 1
24 24 5 3

\[ X \sim (1) \]
(Each Cumulative Curve Represents 40 Simulations)

FIGURE 4.2.21
TWO SAMPLE UNIVARIATE TESTS, J RESPONSE DISTRIBUTION
UNEQUAL SAMPLES

Sample Sizes Response
8 16 3 2
16 32 3 4
8 16 5 2
16 32 5 4

\[ X \sim (1) \]
(Each Cumulative Curve Represents 40 Simulations)

(* See Table (4.2.5), Page 53)
FIGURE 4.2.24
FOUR SAMPLE UNIVARIATE TESTS, BELL RESPONSE DISTRIBUTION
EQUAL SAMPLES

Sample Sizes Response
\( n_1 \quad n_2 \quad n_3 \quad n_4 \) Categories Design
- \( 6 \quad 6 \quad 6 \quad 6 \) 3 1
- \( 12 \quad 12 \quad 12 \quad 12 \) 3 2
- \( 12 \quad 12 \quad 12 \quad 12 \) 3 3
- \( 12 \quad 12 \quad 12 \quad 12 \) 5 1
- \( 12 \quad 12 \quad 12 \quad 12 \) 5 2
- \( 12 \quad 12 \quad 12 \quad 12 \) 5 3

\( \chi^2(3) \)
(Each Cumulative Curve Represents 40 Simulations)

FIGURE 4.2.25
FOUR SAMPLE UNIVARIATE TESTS, BELL RESPONSE DISTRIBUTION
UNEQUAL SAMPLES

Sample Sizes Response
\( n_1 \quad n_2 \quad n_3 \quad n_4 \) Categories Design
- \( 3 \quad 6 \quad 6 \quad 7 \) 3 2
- \( 9 \quad 9 \quad 15 \quad 15 \) 3 4
- \( 9 \quad 9 \quad 15 \quad 15 \) 5 2
- \( 9 \quad 15 \quad 15 \quad 15 \) 5 4

\( \chi^2(3) \)
(Each Cumulative Curve Represents 40 Simulations)
Figure 4.2.2A: Equal Samples, U Response Distribution

Figure 4.2.2B: Equal Samples, U Response Distribution

Sample Sizes: 6, 8, 10, 12, 14
Categories: 1, 3, 5, 7, 9, 11

(Each curve represents one of the simulations.)
and Wallis (1952). They noted that there were fewer observations (H's) falling at the upper end of the distribution, and a compensating slight excess of observations with values a bit more than the average, than would be expected from a chi-square. Further, they knew that the variance of H was less than 2(s-1) for finite N, so use of an approximation which employed the known moments of H was theoretically more justifiable. Hence, they suggested transforming the H variate to the approximate Beta variate H/M, where M is the maximum value H can assume, and judging the significance of a particular H from the percentiles of a Beta distribution with mean equal E(H/M), variance equal Var(H/M) and maximum value equal H/M.

Wallace (1959), commented that the Kruskal-Wallis Beta approximation tended to underevaluate the probabilities of large observations, because the event \{H/M = 1\} has probability zero under the Beta distribution while in fact the probability that H=M may be not small for small samples. Wallace then proposed an alternative approach to arrive at a Beta approximation of the distribution of H:

Starting with the approximate Beta variate

\[
\text{Beta} = \frac{\text{SSB}}{\text{SSB} + \text{SSW}},
\]

where \[
\text{SSB} = \sum_{i=1}^{S} \frac{R_i^2/n_i}{1} - \frac{N(N+1)^2}{4},
\]

the between sample sum of squares, and

\[
\text{SSB + SSW} = [\frac{N(N^2-1)}{12} - \sum_{i=1}^{S} \frac{T_j}{12}],
\]

the total sum of squares, substituting for SSB and SSB + SSW in (4.3.1), Wallace obtained

\[
\text{Beta} = \left[\frac{12}{N(N^2-1)}\right] \left[\sum_{i=1}^{S} \frac{R_i^2/n_i}{1} - \frac{N(N+1)^2}{4}\right] / \left[\sum_{i=1}^{S} \frac{T_j}{N(N^2-1)}\right]
\]

\[
= \frac{1}{(N-1)} H.
\]

\(F_2\), the F statistic, is \(\text{[SSB/between df]} / \text{[SSW/within df]}\), so
\[ F_2 = \frac{SSB/(S-1)}{SSW/(N-S)} \text{ implies that} \]
\[ F_2 = \left[ \frac{(N-S)/(S-1)}{SSB/(S-1)} \right] \frac{SSB(N-1)/(SSB+SSW)}{SSW(N-1)/(SSB+SSW)} \]  

(4.3.3)

Adding and subtracting \( SSB \) in the denominator of (4.3.3),
\[ F_2 = \left[ \frac{(N-S)/(S-1)}{SSB/(S-1)} \right] \frac{SSB(N-1)/(SSB+SSW)}{(SSB+SSW)-SSB} \cdot \frac{1}{N-1} \]
\[ = \left[ \frac{(N-S)/(S-1)}{SSB/(S-1)} \right] \frac{SSB(N-1)/(SSB+SSW)}{-(1-H/((N-1)(N-1))} \]
\[ = \left[ \frac{(N-S)/(S-1)}{SSB/(S-1)} \right] H/(N-1-H) \]  

(4.3.4)

Hence, the transformation (4.3.4) should produce the approximately \( F \) distributed variate \( F_2 \) when an observed \( H \) statistic is substituted into it.

To obtain the parameters of the approximating \( F \) distribution, Wallace noted that the mean and variance of a Beta variate are
\[ \alpha/\alpha+\beta \text{ and } \alpha\beta/[(\alpha+\beta+1)(\alpha+\beta)^2] \], respectively. The mean and variance of \( H/(N-1) \) are \( E(H)/(N-1) \) and \( Var(H)/(N-1)^2 \), respectively, so equating moments and solving for \( \alpha \) and \( \beta \) in terms of \( E(H) \) and \( V(H) \), Wallace found that
\[ \beta = [N-1-E(H)]\alpha/E(H) \text{ and} \]
\[ \alpha = [E(H)(N-1-E(H)) - Var(H)] E(H)/[(N-1)(Var(H))]. \]

Since the numerator degrees of freedom for the \( F \) distribution is
\[ f_1 = 2\alpha \text{ and the denominator degrees of freedom } f_2 = 2\beta, \]
\[ f_1 = E(H)[E(H)(N-1-E(H)) - Var(H)]/[\frac{1}{2}(N-1) Var(H)], \quad (4.3.5) \]
\[ f_2 = [N-1-E(H)] f_1/E(H) \]

These approximations apply to the ties-absent case, but the results should carry over to the ties-present case because they involve only \( N \) and the first two moments of \( H \); Kruskal (1952) showed that the
expectation of $H$ is unaffected by ties and we have shown in Section 4.2.2 that the variances of $H'$ is well approximated by the variance of $H$ from the ties-absent case.

Wallace based his derivation on a one-way classification analysis of variance, so the within-sample degrees of freedom is $(N-S)$. We will be applying the transformation to statistics arising from a cross classification model, but since we found in Section 2.2 that cross classification did not change the basic structure of the $H'$ statistic, it appears we should resist the temptation to reduce $(N-S)$ to the correct degrees of freedom for the cross classification case. The degrees of freedom (4.3.5) will probably not be integers, so one could compute $f_1$ and $f_2$ and round to the nearest integer as recommended by Alexander and Quade (1968), or compute the desired $F$ distribution with fractional degrees of freedom. Both approaches will be tried in the following section.

4.3.2 An Empirical Comparison of the $F$ and $\chi^2$ Approximations to the Distribution of $H'$

Our empirical comparisons of the $F$ and $\chi^2$ approximations of the distribution of $H'$ are based on the analysis that follows from our results in Sections 4.1 and 4.2; we take the mean of $H'$ to be $(S-1)$ and assume that the variance of $H$ is well approximated by equation (2.3.1), i.e., we use the variance of $H$ from the ties-absent case. Under these restrictions, we pool our $H'$ statistics over response distributions, and, in the case of univariate responses, over numbers of categories of responses, since the response variables were made independent in the simulations. Because the difference
between the $F$ and the $\chi^2$ approximations is greater for smaller
total sample sizes, we will work with the smaller total sample size,
 i.e., $N = 24$. Further, since the deviations from $\chi^2$ noted by the
earlier workers are rather small, it appears that only the largest
number of simulations have much chance of clearly showing a difference
between $F$ and $\chi^2$. Therefore, we will consider only the pooled
three and five response category univariate tests for (1) two equal
samples, (2) two unequal samples, (3) four equal samples, and
(4) four unequal samples. Each of these four pooled samples is then
comprised of 240 $H'$ statistics.

The cumulative distributions for the four pooled samples of $H'$
statistics are plotted against $\chi^2$ in Figures (4.3.1) and (4.3.2);
the graphs suggest that there may be too many $H'$ statistics somewhat
larger than $(S-1)$ than is compatible with $\chi^2$, but a paucity of
large $H'$ is not obvious. (Kolmogorov-Smirnov one sample tests do not
indicate statistically significant differences, at the $\alpha = 0.05$
level, between the cumulatives of any of the four pooled samples of $H'$
and the appropriate $\chi^2$.) Each $H'$ was transformed to an $F$ as
shown in equation (4.3.4) and the transformed cumulative distributions
for the four pooled samples were computed. These four cumulatives are
plotted in Figures (4.3.3) and (4.3.4). The degrees of freedom for
the approximating $F$ distributions plotted in Figures (4.3.3) and
(4.3.4) were obtained from equations (4.3.5), where the $\text{Var}(H)$
terms were taken from Table (4.2.2). Table (4.3.1) summarizes these
eight derived $F$ approximations.
<table>
<thead>
<tr>
<th>Number of Samples</th>
<th>Sample</th>
<th>Degrees of Freedom</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Var (H)</td>
</tr>
<tr>
<td>2</td>
<td>Equal</td>
<td>1.8480</td>
</tr>
<tr>
<td></td>
<td>Unequal</td>
<td>1.8230</td>
</tr>
<tr>
<td>4</td>
<td>Equal</td>
<td>5.0400</td>
</tr>
<tr>
<td></td>
<td>Unequal</td>
<td>5.0286</td>
</tr>
</tbody>
</table>

After rounding the degrees of freedom to the nearest integer, the equal and unequal sample size cases within the two and four sample classes had the same numbers of degrees of freedom. Further, in Figures (4.3.3) and (4.3.4), the differences between the cumulative distributions for the exact and rounded degrees of freedom were too small to show in the plot, so only the integer degrees of freedom curves were plotted.

Upon comparing Figures (4.3.1) with (4.3.3) and (4.3.2) with (4.3.4), it is not obvious that the $F$ distribution is superior to the $\chi^2$ in approximating the upper tails of the cumulatives, but the $F$ does appear to fit better near the mean. Perhaps we need more simulations of $H'$ to settle the question of whether there are too few very large $H'$ relative to $\chi^2$. At any rate, the $F$ approximation offers at best only marginal improvement over $\chi^2$ at any point on the cumulative distribution. The use of the $F$ is not very difficult; one transforms the $H'$ to an $F$ as indicated by (4.3.4), computes the adjusted degrees of freedom to integers as suggested by Alexander and
Quade, and uses the resultant $F$ statistic as the index of significance of an observed $H'$. An advantage in the use of the $F$ is that the $F$ provides slightly higher power. This will be discussed in Section 5.4.

### 4.4 Summary of Analysis of Simulation Study

In order to bring together the rather lengthy preceding discussion of our analysis of the simulation study, we will recount our findings in this section:

1. The multivariate analysis of variance comparisons of the means of the 18 test statistics showed no differences across the experiment, except that inequality of sample size increased the interaction test statistic average. Further, no averages of the 18 test statistics within the 12 experimental combinations were found to be significantly different from their expectations at the $\alpha = 0.05$ level, and pooled averages of the $H'$ statistics agreed closely with their expectations.

2. The non-pooled and pooled variance comparisons in Table (4.2.2) and Table (4.2.3) indicated that the variances of the $H'$ statistics were much as expected under the $\chi^2$ assumption, and the comparisons did not show patterns of large or small variances. The variances of the bivariate tests for the effects of $A$ and $B$ are essentially the sums of the variances of the univariate tests, and so were not reported. The unconditional variance of the $H$ statistic was found to provide a good approximation for the conditional variance of $H'$ in this ties-present case.
(3) The cumulative distributions shown in Figures (4.2.6) through (4.2.29) do not show any patterns of disagreement between the distributions of $H'$ and $\chi^2$, while Figures (4.2.1) and (4.2.2) indicate that when pooled over sample size inequality and response distribution, the principal univariate $H'$ statistics are well approximated by $\chi^2$. The cumulative distributions of the principal bivariate tests are not presented because the univariate statistics are well approximated by $\chi^2$ and their sum approximates the bivariate statistic. Figures (4.2.3) through (4.2.5) indicate that the pooled simulations of the $H'$ statistics for the single degree of freedom tests are well approximated by $\chi^2$.

We draw the following conclusions from the results of the null hypothesis simulations in this chapter:

(1) The $H'$ test is robust to differences in the distribution of the response variable.

(2) The $H'$ test performs approximately as well, with respect to Type I error, with three observations per treatment combination in this eight treatment combinations design as it will for larger samples.

(3) The correction for ties adjusts the test statistics so that the three response category case and the five response category case yield very similar results. It is likely that the five response category test is more sensitive to differences but it does not perform better than the three response category test under the null hypothesis.

(4) The use of the $F$ approximation to the distribution of the $H'$ statistics is well justified theoretically, but we were unable to show any clear advantage of the $F$ approximation over $\chi^2$. 

CHAPTER V

A SIMULATION STUDY OF TEST
PERFORMANCE UNDER ALTERNATIVE HYPOTHESES

5.1 Introduction

The focus of our attention up to this point has been on the performance of the Kruskal-Wallis test statistic under the null hypothesis. In this chapter, we shall explore the properties of the univariate and bivariate Kruskal-Wallis statistics when applied to small samples of ordinal response contingency table data with known factor effects present in the data. We have little theoretical guidance concerning the properties of these test statistics in the nonnull, ties-present, small sample case; Puri and Sen (1971) showed that the statistics are asymptotically noncentral chi-square under the alternative in the ties-present case. In view of the relatively close relationship between the ties-present and ties-absent cases as discussed in Chapter 2, it appears a reasonable first approximation to the distributions of the univariate and bivariate $H$ statistics here would be the noncentral chi-square. However, we would not expect the variance of $H'$ in the small sample case to be as large as that of the noncentral chi-square, since the variance of $H$ is smaller in null case, and the categorization of the data would probably reduce the variance still further.
We shall study the alternative hypothesis true case through the use of the Expected Significance Level (ESL) concept of Dempster and Schatzoff (1965). A brief discussion of ESL follows:

Consider a test statistic $X$ with cumulative distribution function $F(x)$, chosen so that a small value of $X$ casts doubt on the null hypothesis $H$. (A lower one-tailed test). For an observed test statistic $X$, we compute

$$\alpha = F(X),$$

(5.1.1)

the observed significance level corresponding to $X$, or, the so-called $p$ value. The sensitivity of the test statistic $X$ against the simple alternative hypothesis $G$ is determined by $\alpha = F(X)$; the sensitivity to $G(X)$ is $H(\alpha)$, and $H(\alpha)$ is the power of a test of size $\alpha$. Dempster and Schatzoff propose the single criterion

$$ESL = \int_0^1 \alpha dH(\alpha)$$

(5.1.2)

as a measure of the power of the test statistic $X$ over all levels of $\alpha$. Since $\alpha = F(x)$ and $X$ has the distribution $G(x)$ under the alternative hypothesis,

$$ESL = \int_{-\infty}^{\infty} F(x) \, dG(x).$$

(5.1.3)

This integral is the probability that $X \leq Y$, where $X$ is distributed according to $F$ and $Y$ is distributed according to $G$, and $X$ and $Y$ are independent. The expressions in (5.1.2) and (5.1.3) can be integrated by parts to yield

$$ESL = 1 - \int_0^1 H(\alpha) d\alpha \quad \text{and}$$

(5.1.4)

$$ESL = 1 - \int_{-\infty}^{\infty} G(x) \, dF(x).$$

(5.1.5)

Formulae (5.1.3) and (5.1.5) show that the two hypotheses $F$ and $G$
are tested symmetrically, i.e., a test based on the upper tail of $X$
for the null hypothesis $G$ has the same ESL against $F$ as that of
the lower tail test of the null $F$ against the alternative $G$.

Dempster and Schatzoff gave two methods of estimating ESL from simu-
lated data; the more efficient second method requires eliminating, in
turn, each observation while performing a computation on the remaining
observations. This requires a more complex data processing scheme
than could be easily thrust into the already complex LINCAT program, so
the first method was followed. A description of the first method
follows:

Suppose $(X_1, X_2, \ldots, X_m)$ is a simulated random sample from $F$,
the null distribution, and $(Y_1, Y_2, \ldots, Y_n)$ is an independently simulated
sample from $G$, the alternative. $F_m(x)$ is the empirical distribution
function of $X$;

$$F_m(x) = \frac{1}{m} \text{[number of } X_1, X_2, \ldots, X_m \leq x ] , \quad -\infty < x < \infty.$$  

The proposed estimator of ESL is suggested by (5.1.3):

$$W = \frac{1}{n} \sum_{i=1}^{n} F_m(Y_i). \quad (5.1.6)$$

The variance of $W$ is

$$\text{Var}(W) = \frac{1}{mn} [p(1-p) + (n-1)(p_1^2 - p^2) + (m-1)(p_2^2 - p^2)] , \quad (5.1.7)$$

where $p = \int_{-\infty}^{\infty} F(y) \, dG(y)$,

$$p_1 = \int_{-\infty}^{\infty} [1-G(x)]^2 \, dF(x),$$

and $p_2 = \int_{-\infty}^{\infty} [F(y)]^2 \, dG(y)$.

Schatzoff (1966) employed ESL in the comparison of several test
statistics. The system he developed for approximating the noncentrality parameter for producing a desired alternative distribution assumes that the test statistic is $\chi^2_p$ and $\chi^2, \Delta^2$ under the null and alternative hypotheses, respectively. He showed that ESL can be written as a cumulative distribution function of a sum of $f$ independent and identically distributed random variables, with mean $\Delta^2$ and variance $4(f + \Delta^2)$.

In the limit, ESL can then be approximated by $1 - \phi(S)$ where $S = \Delta^2/2(f + \Delta^2)^{\frac{1}{2}}$ and $\phi$ is the cumulative of the normal (0,1) random variable. He then obtained the noncentrality parameter $\Delta^2_\alpha$ corresponding to ESL = $\alpha$ from the formula

$$\Delta^2_\alpha = 2Z_\alpha [Z_\alpha + (Z_\alpha^2 + f)^{\frac{1}{2}}],$$ (5.1.8)

where $Z_\alpha$ is the standard normal deviate corresponding to the $\alpha^{th}$ percentage point, and $f$ is the number of degrees of freedom for the test.

We note in passing that since the estimator of ESL is a Wilcoxon type statistic, assuming ESL to be approximately normally distributed should not be too tenuous.

5.2 The Simulation Experiment

5.2.1 The Analytical Framework of the Simulation Experiment

The problems posed in formulating the framework of the alternative true simulations are more difficult than those of the null hypothesis case, for again we need to consider sample sizes and equality of sample sizes, the distributions of the response variables, and the numbers of categories of the response variables and factor variables. In addition, we should allow the degree of noncentrality to vary, the form or pattern
of noncentrality should vary, as should the number of factors which have a significant effect and the relative magnitudes of their effects. Finally, there is the problem of how to best handle a blocking factor variable. In brief, we have all the variables faced in Chapter 3 in the null hypothesis simulations, in addition to quite a number more which could have considerable effect on the test's power.

Because of the exploratory nature of this section of the research, we felt justified in holding sample size constant for each treatment combination and all simulations. We continued to simulate with the numbers of categories of the response variables equal to three and five. For the four sample case we chose to assume a linear trend in the averages across the samples for the alternative true case. All simulations were run with an ESL of 0.05 for one factor variable and the null true for the other. Since this results in confounding of numbers of levels of response with numbers of levels of the factor variables, we also ran another group of similar simulations but with the significant and null factors interchanged.

5.2.2 Specification of the Noncentralities Employed in the Simulation

We performed all the simulations in this section at the ESL = 0.05 level or with ESL = 0.50 (null hypothesis true), with total sample size \( N = 48 \), divided into two samples of size 24 each and four samples of size 12 each. Assuming a linear trend in the sample averages and maintaining the response variable marginal probabilities shown in Table (3.3.1), we turn to equation (5.1.8):

With \( Z_\alpha = 1.645 \) and \( f = 1 \) or 3
\[ f = 1, \quad \Delta^2_{.05} = 11.7456 \]
\[ f = 3, \quad \Delta^2_{.05} = 13.2709 \]

In this analysis-of-variance-like situation the noncentrality

\[ \Delta^2 = \sum_{i=1}^{g} n_i \cdot (\bar{x}_i - \bar{x})^2 / \sigma^2, \quad (5.2.1) \]

We can solve for \( \bar{x}_i \) given \( \Delta^2 \), \( \bar{x} \), \( n_i \), and \( \sigma^2 \). The appropriate variances are shown in Table (5.2.1). The variances are found from the formula (3.6) in Kruskal and Wallis (1952), with \( n = 1 \), i.e.,

\[ \sigma^2 = \frac{N(N^2-1) - \sum T_j}{12N}, \]

where \( T_j = \sum n_{i,j} - \sum n_{i,j} \).

Substituting 11.7456 for \( \Delta^2 \) in the two sample case, 13.2709 for \( \Delta^2 \) in the four sample case, taking \( \bar{x} = 49/2 \) and substituting 24 or 12 for \( n_i \), we can solve (5.2.1) for \( \bar{x}_i \). The results of solving (5.2.1) are shown in Table (5.2.2) as sample averages which will produce ESL's of 0.05. With these averages, the expected rank scores, and the marginal cell probabilities for the three response distributions, a set of cell probabilities was obtained which would produce the sample averages shown in Table (5.2.2). These cell probabilities were then used to form the cell limits in the simulation program, and thus populations were generated with known differences between the samples.

5.3 Results and Analysis of Simulation Study

5.3.1 Expected Significance Level Studies

For this section of the research, a total of six simulations were performed, each consisting of 40 replications, and each with two response variables and two factor variables. Two experimental
### TABLE 5.2.1
VARIANCES BY NUMBERS OF LEVELS OF RESPONSE AND RESPONSE DISTRIBUTION

<table>
<thead>
<tr>
<th>Response Categories</th>
<th>Response Distribution</th>
<th>Bell</th>
<th>J</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>162.5310</td>
<td>149.0617</td>
<td>164.0183</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>180.0085</td>
<td>176.0393</td>
<td>180.8825</td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 5.2.2
SAMPLE AVERAGES PRODUCING ESL'S OF 0.05

<table>
<thead>
<tr>
<th>Response Distribution</th>
<th>Response Categories</th>
<th>Number of Samples</th>
<th>Average</th>
<th>Average</th>
<th>Average</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sample 1</td>
<td>Sample 2</td>
<td>Sample 3</td>
<td>Sample 4</td>
</tr>
<tr>
<td>Bell</td>
<td>3</td>
<td>2</td>
<td>30.8065</td>
<td>31.1553</td>
<td>33.4936</td>
<td>33.9911</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2</td>
<td>18.1936</td>
<td>17.8447</td>
<td>21.5021</td>
<td>15.5064</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>33.9336</td>
<td>27.4979</td>
<td>21.3363</td>
<td>15.0089</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4</td>
<td>33.9336</td>
<td>27.6637</td>
<td>21.3363</td>
<td>15.0089</td>
</tr>
<tr>
<td>J</td>
<td>3</td>
<td>2</td>
<td>30.5395</td>
<td>31.0633</td>
<td>33.1130</td>
<td>33.8630</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2</td>
<td>18.4605</td>
<td>17.9367</td>
<td>21.6290</td>
<td>15.8870</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>33.1336</td>
<td>27.3710</td>
<td>21.3800</td>
<td>15.1400</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4</td>
<td>33.8630</td>
<td>27.6200</td>
<td>21.3800</td>
<td>15.1400</td>
</tr>
<tr>
<td>U</td>
<td>3</td>
<td>2</td>
<td>30.8352</td>
<td>31.1530</td>
<td>33.5348</td>
<td>33.9878</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2</td>
<td>18.1648</td>
<td>17.8470</td>
<td>21.4884</td>
<td>15.4652</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>33.8352</td>
<td>27.5116</td>
<td>21.3374</td>
<td>15.0122</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4</td>
<td>33.8352</td>
<td>27.6626</td>
<td>21.3374</td>
<td>15.0122</td>
</tr>
</tbody>
</table>
arrangements, Designs 1 and 2, were run once each for each of the three response distributions; in Design 1 Factor A, the two level factor, was given a significant effect on the five category response, and Factor B, the four level factor, was given a significant effect on the three category response. Factor A had no effect on the three category response, and Factor B had no effect on the five category response. In Design 2, Factor A had a significant effect on the three category response while Factor B had a significant effect on the five category response and the complementing factor variables had no effect.

Expected Significance Levels (ESL's) were computed for each of the principal test statistics in each of the six simulation runs, using as the null hypothesis simulations the corresponding simulations from Chapter 3. These ESL's are presented in Table (5.3.1); their most remarkable feature is that only one of the 24 exceeded the 0.05 target. Comparison of the ESL's within Table (5.3.1) is possible, but the differences among the comparable ESL's are probably too small to be of any real importance. The standard deviations of the ESL's are included; they are based on equation (5.1.7) and show some dependence between the ESL and S(ESL). There do not appear to be striking differences among the ESL's by distribution of response variables, so we can conclude that given a standardized difference between sample averages approximately as large as needed to produce an ESL of 0.05 under a noncentral $\chi^2$ alternative, we will detect the difference with an ESL of approximately 0.01, provided we use percentage points of the true null distribution of $H'$ as critical values.
### Table 5.3.1

**Observed ESL's and Their Standard Deviations**

**Distribution of Response Variables**

<table>
<thead>
<tr>
<th>Response, Test</th>
<th>Bell</th>
<th>J</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Design 1</td>
<td>Design 2</td>
<td>Design 1</td>
</tr>
<tr>
<td>Bivariate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A ESL</td>
<td>0.0200</td>
<td>0.0069</td>
<td>0.0344</td>
</tr>
<tr>
<td>S(ESL)</td>
<td>0.0103</td>
<td>0.0177</td>
<td>0.0067</td>
</tr>
<tr>
<td>B ESL</td>
<td>0.0081</td>
<td>0.0238</td>
<td>0.0081</td>
</tr>
<tr>
<td>S(ESL)</td>
<td>0.0050</td>
<td>0.0111</td>
<td>0.0076</td>
</tr>
<tr>
<td>Univariate, 3 Categories</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A ESL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S(ESL)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B ESL</td>
<td>0.0019</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S(ESL)</td>
<td>0.0018</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Univariate, 5 Categories</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A ESL</td>
<td>0.0106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S(ESL)</td>
<td>0.0068</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B ESL</td>
<td>0.0019</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S(ESL)</td>
<td>0.0022</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.3.2 Power and Other Considerations

As another way of looking at the results of the alternative true simulations, we computed the probability of rejection of the null hypothesis for varying $\alpha$ levels, using critical values from the central $\chi^2$. These probabilities are presented in Table (5.3.2). They are not as sensitive to the true null hypothesis as the ESL computation, since they measure proportions of observations exceeding values of the central $\chi^2$ rather than measuring the overlap between two distributions. However, the powers do give an indication of the length of the lower tail of the distribution. Upon examination of Table (5.3.2), we noted the following:

1. The power of the univariate tests at $\alpha = 0.05$ is approximately 0.95 for each of the response variable distributions.

2. There does not appear to be a pattern in power among the univariate tests, either with respect to numbers of categories of the response variables or with respect to numbers of samples.

3. The bivariate tests are clearly less powerful than the univariate tests, as they should be.

4. There is an apparent discrepancy between the ESL results and the power results; from equation (5.1.4), ESL = (1 - power, averaged over uniform $\alpha$), we would expect the powers to be greater than those observed, given the observed ESL's. This discrepancy may be caused by one or more of the following reasons:

1. The $\chi^2$ approximation of $H'$ may be somewhat conservative.

2. Random variability could bear heavily since we are considering the tails of the distributions.

3. The approximation to ESL which was used to obtain the noncentrality may not be very good.
<table>
<thead>
<tr>
<th>Test Statistic</th>
<th>Bell</th>
<th>Distribution of Response Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a = 10</td>
<td>a = 0.05</td>
</tr>
<tr>
<td></td>
<td>0.925</td>
<td>0.675</td>
</tr>
<tr>
<td>Bivariate</td>
<td>0.900</td>
<td>0.800</td>
</tr>
<tr>
<td>A (Design 1)</td>
<td>0.975</td>
<td>0.950</td>
</tr>
<tr>
<td>A (Design 2)</td>
<td>0.975</td>
<td>0.950</td>
</tr>
<tr>
<td>B (Design 1)</td>
<td>0.975</td>
<td>0.950</td>
</tr>
<tr>
<td>B (Design 2)</td>
<td>0.975</td>
<td>0.950</td>
</tr>
<tr>
<td>A (Design 1)</td>
<td>0.975</td>
<td>0.950</td>
</tr>
<tr>
<td>A (Design 2)</td>
<td>0.975</td>
<td>0.950</td>
</tr>
<tr>
<td>B (Design 1)</td>
<td>0.975</td>
<td>0.950</td>
</tr>
<tr>
<td>B (Design 2)</td>
<td>0.975</td>
<td>0.950</td>
</tr>
<tr>
<td>A (Design 1)</td>
<td>0.975</td>
<td>0.950</td>
</tr>
<tr>
<td>A (Design 2)</td>
<td>0.975</td>
<td>0.950</td>
</tr>
<tr>
<td>B (Design 1)</td>
<td>0.975</td>
<td>0.950</td>
</tr>
<tr>
<td>B (Design 2)</td>
<td>0.975</td>
<td>0.950</td>
</tr>
<tr>
<td>A (Design 1)</td>
<td>0.975</td>
<td>0.950</td>
</tr>
<tr>
<td>A (Design 2)</td>
<td>0.975</td>
<td>0.950</td>
</tr>
<tr>
<td>B (Design 1)</td>
<td>0.975</td>
<td>0.950</td>
</tr>
<tr>
<td>B (Design 2)</td>
<td>0.975</td>
<td>0.950</td>
</tr>
</tbody>
</table>
In keeping with the recommendations of Wallace concerning approximation of the $H$ statistic, we computed $F$ distribution critical values for these alternative true test statistics and transformed them to $H'$ statistics, following the general approach described in Section 4.3. The critical values thus obtained were slightly less, in most cases, than the central $\chi^2$ critical values, and these were used to form a basis for power computations. The results of these power computations are shown in Table (5.3.3); they indicate the $F$ test will be slightly more powerful, particularly at $\alpha = 0.01$.

In Table (5.3.4), sample moments of the principal test statistics in the alternative true situation are presented. For the most part, the means of the test statistics are in agreement with the mean of the noncentral chi-square, i.e., $(f + \Delta^2)$, but the variances of the $H'$ statistics are considerably smaller than $2(f + 2\Delta^2)$, the variance of the noncentral chi-square. This is not a very surprising finding, considering the known smaller variance of $H'$ under the null hypothesis. The agreement of means could be taken to show that our analysis is rather similar to a one-way analysis of variance of the rank scores.

5.4 The Blocking Factor Problem

The question of blocking factors such as employed in the Freidman Test, Freidman (1937), is of interest for the multiple factor experiment. In each of the simulations run in Section 5.3, only one factor variable had a significant effect, and we observed that the test statistic for the factor with the null effect had a smaller mean and variance than expected under the null hypothesis. This is shown in Table (5.4.1); it is more of a problem for the two sample case than for the
<table>
<thead>
<tr>
<th>Test Statistic</th>
<th>Bell</th>
<th></th>
<th></th>
<th>J</th>
<th></th>
<th></th>
<th>U</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α = .10</td>
<td>α = .05</td>
<td>α = .01</td>
<td>α = .10</td>
<td>α = .05</td>
<td>α = .01</td>
<td>α = .10</td>
<td>α = .05</td>
<td>α = .01</td>
</tr>
<tr>
<td>Bivariate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A(Design 1)</td>
<td>0.925</td>
<td>0.925</td>
<td>0.675</td>
<td>0.950</td>
<td>0.900</td>
<td>0.725</td>
<td>0.925</td>
<td>0.900</td>
<td>0.800</td>
</tr>
<tr>
<td>A(Design 2)</td>
<td>0.975</td>
<td>0.975</td>
<td>0.825</td>
<td>0.900</td>
<td>0.875</td>
<td>0.725</td>
<td>0.850</td>
<td>0.625</td>
<td>0.450</td>
</tr>
<tr>
<td>B(Design 1)</td>
<td>0.950</td>
<td>0.875</td>
<td>0.550</td>
<td>0.950</td>
<td>0.900</td>
<td>0.775</td>
<td>0.950</td>
<td>0.900</td>
<td>0.700</td>
</tr>
<tr>
<td>B(Design 2)</td>
<td>0.875</td>
<td>0.825</td>
<td>0.600</td>
<td>0.875</td>
<td>0.750</td>
<td>0.550</td>
<td>1.000</td>
<td>0.975</td>
<td>0.975</td>
</tr>
<tr>
<td>Univariate, 3 Categories</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A(Design 2)</td>
<td>0.975</td>
<td>0.950</td>
<td>0.900</td>
<td>0.975</td>
<td>0.950</td>
<td>0.800</td>
<td>0.950</td>
<td>0.900</td>
<td>0.575</td>
</tr>
<tr>
<td>B(Design 1)</td>
<td>1.000</td>
<td>0.975</td>
<td>0.875</td>
<td>1.000</td>
<td>1.000</td>
<td>0.925</td>
<td>0.975</td>
<td>0.975</td>
<td>0.875</td>
</tr>
<tr>
<td>Univariate, 5 Categories</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A(Design 1)</td>
<td>0.975</td>
<td>0.925</td>
<td>0.850</td>
<td>0.950</td>
<td>0.925</td>
<td>0.825</td>
<td>0.950</td>
<td>0.900</td>
<td>0.900</td>
</tr>
<tr>
<td>B(Design 2)</td>
<td>0.975</td>
<td>0.925</td>
<td>0.850</td>
<td>0.975</td>
<td>0.925</td>
<td>0.675</td>
<td>1.000</td>
<td>1.000</td>
<td>0.975</td>
</tr>
</tbody>
</table>
### Table 5.3.4
SAMPLE MOMENTS OF TEST STATISTICS UNDER ALTERNATIVE HYPOTHESES

<table>
<thead>
<tr>
<th>Response, Test Statistic</th>
<th>Distribution of Response Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bell</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td><strong>Bivariate</strong></td>
<td></td>
</tr>
<tr>
<td>A (Design 1)</td>
<td>12.92</td>
</tr>
<tr>
<td>A (Design 2)</td>
<td>13.61</td>
</tr>
<tr>
<td>B (Design 1)</td>
<td>17.12</td>
</tr>
<tr>
<td>B (Design 2)</td>
<td>17.02</td>
</tr>
<tr>
<td><strong>Univariate, 3 Categories</strong></td>
<td></td>
</tr>
<tr>
<td>A (Design 2)</td>
<td>12.88</td>
</tr>
<tr>
<td>B (Design 1)</td>
<td>14.78</td>
</tr>
<tr>
<td><strong>Univariate, 5 Categories</strong></td>
<td></td>
</tr>
<tr>
<td>A (Design 1)</td>
<td>12.20</td>
</tr>
<tr>
<td>B (Design 2)</td>
<td>15.23</td>
</tr>
</tbody>
</table>

* Based on a noncentral chi-square distribution
### TABLE 5.4.1
SAMPLE MOMENTS OF TEST STATISTICS FOR NULL
FACTOR EFFECTS IN THE PRESENCE OF NON-NULL EFFECTS

<table>
<thead>
<tr>
<th>Response, Test Statistic</th>
<th>Bell</th>
<th>J</th>
<th>U</th>
<th>Expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Univariate, 3 Categories</td>
<td>Mean</td>
<td>Variance</td>
<td>Mean</td>
<td>Variance</td>
</tr>
<tr>
<td>A (Design 1)</td>
<td>0.71</td>
<td>1.38</td>
<td>0.83</td>
<td>0.94</td>
</tr>
<tr>
<td>B (Design 2)</td>
<td>1.79</td>
<td>2.23</td>
<td>2.38</td>
<td>1.65</td>
</tr>
<tr>
<td>Univariate, 5 Categories</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A (Design 2)</td>
<td>0.73</td>
<td>0.83</td>
<td>0.69</td>
<td>0.62</td>
</tr>
<tr>
<td>B (Design 1)</td>
<td>2.35</td>
<td>4.29</td>
<td>2.51</td>
<td>3.39</td>
</tr>
</tbody>
</table>
four sample case in this experiment. The phenomena arises from ranking the observations over all the samples. Then, if one factor variable is dominant over the data, the ranks tend to group within the levels of the dominant variable and so the test for less important variable measures mostly variability within the levels of the dominant factor instead of variability over the whole population. This leads to a smaller effect for the null factor than would expected to be caused be random chance. If the unimportant factor has no effect at all we have a one-way classification test for the dominant factor.

The significance of the effect of the less important factor in a two factor contingency table could be evaluated by ranking the observations separately within levels of the dominant factor, i.e., using a known important factor as a blocking variable. The main objection to doing this is erosion of sample size, but the results of Chapter 4 indicate that sample sizes can be rather small and reasonable results still can be obtained. The next problem is forming the test statistic; one could compute an $S$ sample Kruskal-Wallis statistic within each of the $C$ levels of the blocking factor, and pool to obtain a $C(S-1)$ degree of freedom test. This should not be very powerful, however, so a better method would be to fit a linear model to all the $C$ blocks simultaneously and obtain an $(S-1)$ degree of freedom test for the factor of interest. Puri and Sen (1971) suggested aligning the observations across blocks before ranking as another approach. The alignment would be achieved by perhaps subtracting the corresponding block average from each observation before ranking; how we would achieve this in the categorical data case is unclear for we do not have averages until the ranking has been done. In a recent paper, Forthofer and
Koch, (1972), proposed fitting a complex exponential logarithmic model rather than a simple linear model as a way to avoid the problems of the interdependence of the rank scores.

5.5 Discussion and Summary

The H' statistic, in this one alternative hypothesis true case, provides a relatively powerful test for the significant factor effect. We found no real evidence of difference between the performance of the test statistics in the three response category and the five response category cases; the two sample and four sample tests likewise had nearly equal ESL's and powers. Of course, the noncentralities injected into the simulations were selected to yield the same ESL in each of the sample/response combinations, so not finding differences is not surprising. The bivariate tests were less powerful than the univariate tests, and much as expected, given the univariate test performances.

The sensitivity of the modified Kruskal-Wallis test in this situation is similar to what might be expected in doing an analysis of variance on the ranked data.

The discrepancy between ESL and power shows the importance of using the appropriate null distribution critical values; the central chi-square critical points may result in a somewhat conservative test, and the transformation to an F test will slightly improve this conservativeness, particularly at smaller α levels.
CHAPTER VI

SUMMARY, DISCUSSION, AND RECOMMENDATIONS FOR FURTHER RESEARCH

6.1 Summary and Discussion

We have found and explored a relationship between the Kruskal-Wallis H statistic and the modified chi-square statistic, $\chi^2_\ell$. The relationship, $\chi^2_\ell = HN/(N-1)$, was derived under the assumptions that the null hypothesis is true, that rank scores are used in the $\chi^2_\ell$ analysis, and that the H statistic is corrected for ties. When the null hypothesis is not true, the relationship does not hold, and $\chi^2_\ell$ was found to be usually larger than $HN/(N-1)$. The underlying difference between $H$ and $\chi^2_\ell$ was found to be in the variances employed in the formulations of the test statistics; $\chi^2_\ell$ uses a pooled within-sample variance while $H$ uses a response marginal based variance. A relationship between $\chi^2_\ell$ and $H$ which incorporated the difference between the variances was found for the two equal sample case. The variance of $H$, conditional on the observed numbers and locations of ties, was derived and examination of the moments of $\chi^2_\ell$ followed. We concluded that the relationship $\chi^2_\ell = HN/(N-1)$ would provide an approach for estimation of lower bounds on the mean and variance of $\chi^2_\ell$.

In Chapter 3, we set up a simulation experiment involving the following variables; (1) sample sizes, (2) sample size equality,
(3) numbers of levels of the factor variables, (4) numbers of categories of the response variables, and (5) form of the response variable distribution. Some preliminary simulations showed $\chi^2_L$ to be quite anti-conservative relative to $\chi^2$, and that the moments of $\chi^2_L$ were considerably larger than the corresponding moments of $\text{HN}/(N-1)$. On the basis of these simulations and the analytical work in Chapter 2, we decided to modify the formulation of $\chi^2_L$ by using a response marginal based variance rather than a pooled within-sample variance. This resulted in the test statistic we refer to as $H'$, which is identical to $H$ wherever $H$ is defined, but is also defined for unbalanced cross classification problems.

In Chapter 4 we performed the null hypothesis true simulation experiment outlined in Chapter 3, and analyzed the simulation results. We compared the means, variances, and cumulative distributions of the $H'$ statistics and found that under the null hypothesis the distribution of $H'$ was approximately $\chi^2$. For the samples of 40 $H'$ statistics, we found no situations where the $\chi^2$ approximation was clearly rejected by the various tests of significance that were employed in the analysis, and for the samples of 240 observations, the $\chi^2$ approximation was good.

In Chapter 5 we studied the properties of the $H'$ statistic under alternative hypotheses. We found the power of the $H'$ test was comparable to the power of an $F$ test. However, the Expected Significance Level of the $H'$ test was smaller than would be expected from $\chi^2$.

From the foregoing, we conclude that the $H'$ statistic offers a useful test statistic for the analysis of one-way or two-way cross classified categorized data. The $H'$ statistic should be approximately
distributed under the null hypothesis for two factor variable problems with eight or more treatment combinations and three or more observations per treatment combination, i.e., for tables such as shown in Table (4.2.5) of the Foreword to Cumulative Distributions in Section 4.2.3. The $H'$ statistic is not very sensitive to the form of the response distribution or to the number of categories of the response variable. Bivariate response problems can be analyzed through the use of $H'$, and the power of the $H'$ test is comparable to the $F$ test.

Investigation showed that the $\chi^2$ statistic did not behave like $\chi^2$ under the null hypothesis; we, therefore, modified $\chi^2$ by replacing its pooled within-sample variance by a variance based on the marginal distribution of the response variables. Although the performance of the modified $\chi^2$, or $H'$, statistic was found to be acceptable in Chapter 4 and 5, it is interesting to consider the possible advantages of using a pooled within-sample variance. The pooled within-sample variance estimates random variability, regardless of whether the null hypothesis is true. Further, for rank scores data, the within-sample variance will decrease as the between-sample variance increases, for their sums of squares add to the fixed total sum of squares. Hence, $\text{Between SS/Within SS}$ increases quickly as $\text{Between SS}$ grows, and so we might expect to obtain a more sensitive test statistic from the pooled within-sample variance approach. On the other hand, use of the response marginal variance has a minor advantage in that it avoids the difficulty of zero variances within the samples.

Another consideration is that of the applicability of these results. We have dealt with only two-way cross classification designs,
and the situation may be different for more or fewer factor variables. We are cautious about extending our conclusions to three or more factor problems because of the negative relationship between the test statistics for the crossed factors that was observed in Chapter 5. To see whether there was an important negative relationship between the test statistics for the two factors in the null hypothesis case, the rank correlations between the test statistics were computed. We pooled over sample sizes and response distributions, but used no unequal sample size results because the factors were clearly not orthogonal for unequal samples. The following rank correlations were found:

<table>
<thead>
<tr>
<th></th>
<th>3 Response Levels</th>
<th>5 Response Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.087</td>
<td>-0.124</td>
</tr>
</tbody>
</table>

We note that the three response level results indicate a small positive correlation while the five response level results suggest a small negative correlation. We conclude that correlation between the test statistics is not a very serious problem under the null hypothesis.

6.2 **Recommendations for Further Research**

A number of extensions of the preceding work would probably yield interesting results. They are the following:

1. Extension to three way cross classifications.
2. Extension to the blocking factor problem mentioned in Chapter 5.
3. A further study of the test properties under the alternative hypothesis.

In addition to these extensions, it would be interesting to
modify $\chi^2_p$ in other ways; one possibility is to use $(n_i - 1)$ rather than $n_i$ as the division of $V(p_i)$.
APPENDIX I

DERIVATION OF RELATIONSHIP BETWEEN $\chi^2$ AND $H$

Suppose we have a contingency table such as Table 1.1. To analyze it using the Kruskal-Wallis $H$ test, we assign the midrank scores

\[
a_1 = (n_{1.} + 1)/2, \\
\vdots \\
a_r = n_{r.} + 1)/2, \quad (A1.1)
\]

to the individual exhibiting responses $1, 2, \ldots, r$, respectively.

From (1.3.1), the $H$ statistic can be written as

\[
H = \left\{ \frac{(N-1)/N[1/(1-\sum_j T_j/(N^3-N))]^{12}/(N^2-1)}{\sum_i n_i} \right\} \frac{\sum_i (\bar{R}_i - \frac{1}{2}(N+1))^2}{} \quad (A1.2)
\]

The total of the rank scores for sample $i$ is $R_i$;

\[
R_i = n_{i1}a_1 + n_{i2}a_2 + \ldots + n_{ir}a_r. \quad (A1.3)
\]

The tie correction terms $T_j$ are

\[
T_j = n_{.j}(n_{.j}^2-1), \quad j = 1, 2, \ldots, r, \quad \text{so}
\]

\[
\frac{\sum_i T_j}{n} = \sum_{.j} n_{.j}^3 - N. \quad (A1.4)
\]

Substituting (A1.4) into (A1.2) and letting $f_i = \bar{R}_i = R_i/n_i$.,
we have
\[ H = \left[ (N-1)/N \right] \left[ \frac{12N}{(N^3 - \sum_{j=1}^{N} n_j^3)} \right] \left[ n_1, (f_1-(N+1)/2)^2 + \ldots \right. \]
\[ \left. + n_s, (f_s-(N+1)/2)^2 \right]. \quad (A1.5) \]

Noting that \( (f_1-(N+1)/2)^2 = f_1^2 - N f_1 - f_1 + (N+1)^2/4 \),
\[ \sum_{i=1}^{s} n_i, (f_i-(N+1)/2)^2 = n_1, f_1^2 - n_1, N f_1 + n_1, (N+1)^2/4 \]
\[ + \ldots \]
\[ + n_s, f_s^2 - n_s, N f_s - n_s, f + n_s, (N+1)^2/4 \],
\[ = n_1, f_1^2 + \ldots + n_s, f_s^2 - (N+1) \left( n_1, f_1 + \ldots + n_s, f_s \right) + \]
\[ ((N+1)^2/4) (n_1 + \ldots + n_s). \quad (A1.6) \]

Since \( (N+1)/2 = (n_1, f_1 + \ldots + n_s, f_s)/N \),
\[ (N+1)^2/4 = (1/N^2) \left( \sum_{i=1}^{s} n_i, f_i^2 + 2 \sum_{i<j}^{s} n_i, n_j, f_i f_j \right), \]
and hence
\[ ((N+1)^2/4) (n_1 + \ldots + n_s) = 1/N \left[ \sum_{i=1}^{s} n_i, f_i^2 + 2 \sum_{i<j}^{s} n_i, n_j, f_i f_j \right]. \]

Multiplying (A1.6) by \( N/N \), and using \( N(N+1) = 2(\sum_{i=1}^{s} n_i, f_i) \) and
\[ N = \sum_{i=1}^{s} n_i, \]
\[ \sum_{i=1}^{s} n_i, (f_i-(N+1)/2)^2 = 1/N \left\{ \sum_{i=1}^{s} n_i, f_i^2 + \sum_{i<j}^{s} n_i, n_j, f_i f_j - 2 \sum_{i=1}^{s} n_i, f_i^2 \right. \]
\[ - 4 \sum_{i<j}^{s} n_i, n_j, f_i f_j \]
\[ + 2 \sum_{i<j}^{s} n_i, n_j, f_i f_j \left\} \right. \quad (A1.7) \]
Cancelling, (Al.7) becomes

$$\sum_{i=1}^{S} n_i \left( f_i - (N+1)/2 \right)^2 = 1/N \left\{ \sum_{i \neq j}^{s} n_i n_j f_{ij}^2 - 2 \sum_{i < j} n_i n_j f_{ij} f_{ji} \right\}. $$

We have terms of the type

$$n_i n_j (f_{ij}^2 + f_{ji}^2 - 2f_{ij} f_{ji}) = n_i n_j (f_{ij} - f_{ji})^2, \text{ so}$$

$$\sum_{i=1}^{S} n_i (f_i - (N+1)/2)^2 = 1/N \sum_{i < j}^{s} n_i n_j (f_{ij} - f_{ji})^2. \tag{Al.8}$$

We can express \((N^3 - \sum_{i,j} n_i n_j)^3\) in other terms:

$$N^3 = \sum_{i,j}^{r} n_i n_j^2 + 3 \sum_{i \neq j}^{r} n_i n_j n_k, \quad \text{so}$$

$$N^3 - \sum_{i,j}^{r} n_i n_j^2 = 3 \left( \sum_{i \neq j}^{r} n_i n_j+2 \sum_{i < j}^{r} n_i n_j n_k \right) \text{ and, using (Al.8),}$$

\((Al.5)\) can be expressed as

$$H = \frac{(N-1)/N}{[4/(\sum_{i \neq j}^{r} n_i n_j + 2 \sum_{i < j}^{r} n_i n_j n_k)]} \left[ \sum_{i < j}^{s} n_i n_j (f_{ij} - f_{ji})^2 \right]. \tag{Al.9}$$

The linear model that is appropriate for this comparison work is

$$E(Y_i) = \mu + \beta_i, \quad \text{or} \quad E(Y_i) = \sum_{i=1}^{r} a_j \pi_{ij}, \quad i = 1, 2, \ldots s. \quad \text{The s}$$

functions of \(\pi\) cited in (1.3.4), with \(u = s\), are then

$$f_{ij} (\pi) = \mu + \beta_i = \pi_{11} a_1 + \ldots + \pi_{1r} a_r$$

$$f_{ij} (\pi) = \mu + \beta_s = \pi_{sl} a_1 + \ldots + \pi_{sr} a_r$$

$$H = \left[ \frac{\partial f_i (\pi)}{\partial \pi_{ij}} \right]_{\pi_{ij} = p_{ij}}$$

becomes, since its elements are of the form
\[
\begin{align*}
\frac{\partial f_1(\pi)}{\partial \pi_{11}} &= a_1, \quad \frac{\partial f_2(\pi)}{\partial \pi_{11}} = 0, \; \text{etc.,}
\end{align*}
\]

\[
H = \begin{bmatrix}
(a_1 & a_2 & \ldots & a_r) & 0 & 0 \\
\vdots & & & & \vdots & \vdots \\
0 & (a_1 & a_2 & \ldots & a_r) & \ldots & 0 \\
\vdots & & & & \vdots & \vdots \\
0 & 0 & & & \ldots & (a_1 & a_2 & \ldots & a_r)
\end{bmatrix}
\]

Thus \( H \) imposes the rank scores onto the analysis through \( S = HV(p)H' \).

As in (1.3.4), \( F(\pi) = X \beta^* \), where we choose the reparametrization

\[
\beta_i^* = \beta_i - \bar{\beta}, \; i = 1 \ldots (s-1), \; \text{so} \quad \beta_s^* = -\beta_1^* - \beta_2^* - \ldots - \beta_{(s-1)}^*.
\]

To achieve this reparametrization, we take

\[
\begin{bmatrix}
1 & 1 & 0 & \ldots & 0 \\
1 & 0 & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & -1 & -1 & \ldots & -1
\end{bmatrix}
\begin{bmatrix}
\mu \\
\beta_1^* \\
\vdots \\
\beta_{s-1}^*
\end{bmatrix}
\]

Now to test the hypothesis of homogeneity among samples we use the contrast matrix

\[
C = \begin{bmatrix}
0 & 1 & 0 & \ldots & 0 \\
0 & 0 & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 1
\end{bmatrix}
\]

and the test statistic for the hypothesis \([ C \beta^* = 0 ] \) is given at (1.3.5).
In particular, for the two sample, \( r \) response problem,

\[
X = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad C = [0, 1], \quad F = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}
\]

and

\[
S = \begin{bmatrix} a_1 a_2 \ldots a_r & 0 \\ 0 & a_1 a_2 \ldots a_r \end{bmatrix} \begin{bmatrix} V(p_1) & 0 \\ 0 & V(p_2) \end{bmatrix} \begin{bmatrix} a_1 a_2 \ldots a_r & 0 \\ 0 & a_1 a_2 \ldots a_r \end{bmatrix},
\]

where

\[
V(p_i) = 1/n_i, \begin{bmatrix} p_{i1}(1-p_{i1}) & \ldots & -p_{i1}p_{ir} \\ -p_{i1}p_{ir} & \ldots & p_{ir}(1-p_{ir}) \end{bmatrix} i = 1, 2.
\]

Performing the matrix algebra indicated by (1.3.5) and (1.3.6),

\[
\chi^2 = (f_1-f_2)^2/(a' V(p_1) a + a' V(p_2) a), \tag{A1.10}
\]

where \( a' = (a_1 a_2 \ldots a_r) \).

For the three sample, \( r \) response case we can do similarly and obtain the test statistic

\[
\chi^2 = \frac{n_2. n_3. z_1(f_2-f_3)^2+n_1. n_2. z_2(f_1-f_3)^2+n_1. n_2. z_3(f_1-f_2)^2}{n_3. z_1 z_2 + n_2. z_1 z_3 + n_1. z_2 z_3}, \tag{A1.11}
\]

where \( z_i = n_i. a' V(p_i) a \).

The test statistic in the \( s \) sample case is

\[
\chi^2 = \frac{\sum_{i<j} (n_i. n_j. (f_i-f_j)^2}{\sum_{k \neq i,j} (n_i. (z_k))^2}/[\sum_{i=1}^{s} (n_i. (z_i))] \tag{A1.12}
\]

Now expanding a \( z_i \) expression for two levels of response,

\[
z_i = n_i. a' V(p_i) a = p_{ii} p_{i2}((n_1+n_2)/2)^2,
\]

and for \( r \) levels of response,
\[ z_1 = p_{i1}p_{12}((n_{1} + n_{2})/2)^2 + p_{i1}p_{i3}(n_{1}/2 + n_{2} + n_{3}/2)^2 + \cdots + \\
+ p_{i1}p_{ir}(n_{1}/2 + n_{2} + \cdots + n_{r}/2)^2 + p_{i2}p_{3}(n_{2} + n_{3}/2)^2 \\
+ \cdots + p_{i(r-1)}p_{ir}((n_{(r-1)} + n_{r})/2)^2 . \] (A1.13)

Under the null hypothesis (1.3.2),

\[ p_{ij} = p_{ij} = p_j, \quad j = 1, 2, \ldots, r, \quad i \neq i' = 1, 2, \ldots, s . \] (A1.14)

This implies all the \( z_i \) are equal, as can be observed at (1.3.3), so, taking \( z = z_1 \), from (A1.12),

\[ \chi^2 = \left[ z^{s-2} \sum_{i \neq j} n_i n_j (f_i - f_j)^2 \right] / z^{s-1} \sum_i n_i . \]

\[ = \left[ \sum_{i \neq j} n_i n_j (f_i - f_j)^2 \right] / zN . \] (A1.15)

We can now directly relate the \( \chi^2 \) and \( H \) statistics by noting that the extreme right hand term of (A1.9) is identical to the numerator of (A1.15). If we assume \( \chi^2 \) \( Y = H \) and solve for \( Y \), we obtain a term which reflects the difference between the statistics.

Taking \( \chi^2 \) \( Y = H \) and rearranging (A1.15) and (A1.9), we have

\[ Y/zN = [(N-1)/N] \left[ 4 \sum_{i \neq j} n_i n_j \sum_{i \neq j} n_i n_j + 2 \sum_{i \neq j < k} n_i n_j n_k \right] . \] (A1.16)

Solving for \( Y \), and using (A1.13) and (A1.14) to evaluate \( z \),

\[ Y = [(N-1)/N] \left[ 4N \right] \left[ p_{1}p_{2}((n_{1} + n_{2})/2)^2 + \right. \\
+ p_{1}p_{3}(n_{1}/2 + n_{2} + n_{3}/2)^2 + \cdots \\
+ p_{(r-1)}p_{r}((n_{(r-1)} + n_{r})/2)^2 / \left[ \sum_{i \neq j} n_i n_j \sum_{i \neq j < k} n_i n_j n_k \right] . \]
Substituting \( n_j/N \) for \( p_j \),

\[
Y = [(N-1)/N][4N][1/(4N^2)]\left[n_1^2 + n_2^2 + n_3^2 \right]^{1/2} \\
+ \ldots + n_r^2 \\
+ \left[ \sum_{i=1}^{r-2} n_i^2 \right]^{1/2} \\
+ \ldots + n_1^2 + n_2^2 + n_3^2 + \ldots + 2n_r(n_1^2 + n_2^2 + \ldots + n_r^2) \\
+ \ldots + n_r^2 \right].
\]

The denominator cancels into the numerator, leaving

\[
Y = [(N-1)/N][4N][1/(4N^2)]\left[ \sum_{i=1}^{r-1} n_i \right]^{1/2}, \quad \text{so}
\]

\[
Y = (N-1)/N,
\]

and hence

\[
\chi^2_{k}((N-1)/N) = H.
\]  \hspace{1cm} (A1.17)
APPENDIX II

DERIVATION OF THE CONDITIONAL VARIANCE OF THE KRUSKAL-WALLIS H STATISTIC IN THE TIES-PRESENT CASE

The Kruskal (1952) derivation of the variance of the $H$ statistic arose directly from the definition of $H$. He noted that

$$\text{Var}(H) = \frac{144}{N^2(N+1)^2} \left( \sum_{i=1}^{N} \frac{2}{n_i} \text{ER}_i^2 + \sum_{i=1}^{N} \frac{1}{n_i(n_i-1)} \text{E}(R_{i,j}^2) - \left[ \sum_{i=1}^{N} \frac{1}{n_i} \text{ER}_i \right]^2 \right). \quad (A2.1)$$

He evaluated the expression term by term; his representation of $\text{E}(R_{i,j}^2)$ was of the form

$$\text{E}(R_{i,j}^2) = \sum \sum \sum \text{E}[X_i^{(1)} X_i^{(2)} X_j^{(1)} X_j^{(2)}], \quad i, i' = 1, \ldots n_i, \quad j, j' = 1, \ldots n_j. \quad (A2.2)$$

Breaking the quadruple sum into components and taking $i = 1$ and $j = 2$,

$$\text{E}(R_{1,2}^2) = \frac{n_1 n_2 (n_1 -1)(n_2 -1)}{N(N-1)(N-2)(N-3)} \sum \sum \sum (X_1^{(1)} X_2^{(1)} X_1^{(2)} X_2^{(2)})$$

$$+ \frac{n_1 n_2 (n_1 -1)}{N(N-1)(N-2)} \sum \sum (X_1^{(1)} X_2^{(1)} X_1^{(1)} X_2^{(1)})$$

$$+ \frac{n_1 n_2 (n_2 -1)}{N(N-1)(N-2)} \sum \sum (X_1^{(1)} X_2^{(2)} X_1^{(2)} X_2^{(2)})$$

$$+ \frac{n_1 n_2}{N(N-1)} \sum (X_1^{(1)} X_1^{(1)}) \quad (A2.3)$$
To evaluate \( \text{Var}_c(W) \) we need \( E R_1^4 \), \( E R_1 \), and an evaluation of \( (A2.3) \) for the case where the \( X_{1k} \) are not the integers \( 1 \) through \( N \).

Klotz (1966) derived the first four conditional moments of the Wilcoxon statistic for the ties-present, null hypothesis true case. He pointed out that the Wilcoxon test statistic, \( W \), can be expressed as \( U Q t' \), where \( U \) is an \( r \) element vector of \( U_1 \), \( Q \) is an \( r \times r \) matrix with entries of \( 1/2 \) on the main diagonal, \( 1 \) below the diagonal, and \( 0 \) above the diagonal, and \( t' \) is an \( r \) element vector with entries representing the numbers of observations involved in each set of ties, i.e., the \( t_j \) are equivalent to our \( n_{ij} \). Klotz then observed that since we are sampling from a finite population of size \( N = \sum_l n_{1i} \), and in which there are \( r \) sets of ties, we are sampling from the multivariate hypergeometric distribution:

\[
P[U=u|T=t] = \frac{\left(\begin{array}{c} t_1 \\ u_{1i} \end{array}\right) \left(\begin{array}{c} t_2 \\ u_{12} \end{array}\right) \cdots \left(\begin{array}{c} t_r \\ u_{1r} \end{array}\right)}{\begin{array}{c} N \\ n_{1i} \end{array}}
\]

(Johnson and Kotz (1969) refer to this distribution as the multivariate hypergeometric distribution. Klotz calls it a generalized hypergeometric distribution.) The probability element is conditional on the numbers of ties \( t_j \), and the \( u_{ij} \) correspond to our cell frequencies \( n_{ij} \).

Klotz found the first few moments of the \( U_j \) and employed them in finding the moments of \( W \), the Wilcoxon statistic.

\[
\sigma^2(W) = \text{Var}(U Q t') = n_{1i} n_{2i} (N+1)/12 - \sum_{j=1}^{r} (t_j^3 - t_j)n_{1i} n_{2i} /12N(N-1) .
\]

(A2.4)

Defining a vector of coefficients \( b' = (b_1 b_2 \ldots b_r) \), where \( b_1 = t_1/2N \),
\[ b_2 = \frac{(t_1 + t_2/2)N}{\ldots b_j = \sum_{j'=1}^{j-1} \left( t_{j'} + t_{j'}/2 \right) N}, \quad \text{and} \quad U \cdot Q \cdot t' = \sum_{j=1}^{r} U_j b_j N. \]

Under this notation, Klotz found that

\[ \mu_3(W) = \frac{n_{1, n_2, \nu_{1, n_2, \nu_2, n_2}}}{(N-1)(N-2)} \left[ \sum_{j=1}^{r} b_1 (1-b_j) (b_j - \frac{1}{2}) t_j / N \right]. \quad (A2.5) \]

\[ \mu_4(W) = \frac{3n_{1, n_1, n_2, n_2}}{N(N-1)(N-2)(N-3)} \left[ \sum_{j=1}^{r} b_{1, j}^2 t_j / N - b_{2, j}^2 t_j / N \right] - b_{3, j} (b_j - \frac{1}{2})^4 t_j / N, \quad (A2.6) \]

where \( B = \frac{6n_{1, n_1, n_2, n_2}}{N^2(N-1)(N-2)(N-3)} - \frac{n_{1, n_2, n_2}}{N^2(N-1)}. \)

To find the conditional variance of \( H \), we recall the expression \( (A2.1) \); to evaluate it we need \( E(R_1^4) \), \( E(R_1^2 R_1^2) \), and \( E(R_1^2) \). Klotz has \( \mu_4(W) = \mu_4(R_1) \); we need

\[ E(R_1^4) = \mu_4(R_1) + 4\mu_1^4(R_1) + 6\mu_1^2(R_1) \mu_2(R_1) + 6\mu_1^2(R_1) \mu_2(R_1) + \mu_1^4(R_1), \quad (A2.7) \]

so from \( (A2.4), (A2.5), (A2.6) \), and using \( \mu_1^4(R_1) = n_1(N+1)/2 \),

\[ E(R_1^4) = \frac{3n_{1, n_1, -1}(N-n_{1, n_1, -1})(N-n_1, 1)}{N(N-1)(N-2)(N-3)} N^6 \left[ \sum_{j=1}^{r} b_{1, j}^2 t_j / N - b_{2, j}^2 t_j / N \right] - b_{3, j} (b_j - \frac{1}{2})^4 t_j / N \]

\[ - 4\left( n_{1, (N+1)/2} \right) \left( n_{1, n_1, (N-n_{1, n_1, -1})} \right) \left( N-2n_{1, n_1} \right) \]

\[ + 6\left( n_{1, (N+1)/2} \right)^2 \left( n_{1, (N-n_{1, n_1}, 1)} \right) / 12 - \sum_{j=1}^{r} (t_{j}^3 - t_{j}) n_{1, n_1, -1} / 12N(N-1) \]

\[ + \left( n_{1, (N+1)/2} \right)^4. \quad (A2.8) \]

Cancelling, rearranging, and substituting for \( B \),

\[ E(R_1^4) = \frac{3N^5 \left[ \sum_{j=1}^{r} b_{1, j}^2 t_j / N - b_{2, j}^2 t_j / N \right]^2}{(N-1)(N-2)(N-3)} \left[ n_{1, n_1, -1}(N-n_{1, n_1, -1})(N-n_{1, n_1, -1}) \right] \]

\[ - \frac{N^3 \left[ b_{3, j} (b_j - \frac{1}{2})^4 t_j \right]}{N-1} \left[ n_{1, n_1, -1} \right] - \left[ \frac{6(n_{1, n_1, -1})(N-n_{1, n_1, -1})}{N(N-2)(N-3)} \right] - 1. \]
\[-\frac{2(N+1)^2}{(N-1)(N-2)} \left( \sum_{j=1}^{r} b_j (1-b_j)(b_j - \frac{1}{2}) t_j \right) n_1^2 (N-n_1)(N-2n_1) \]

\[+ \frac{1}{8} (N+1)^2 \left( n_1^3 (N-n_1)(N+1) - \sum_{j=1}^{r} (t_j^3 - t_j) n_1^3 (N-n_1) / N(N-1) \right) + \frac{1}{16} (N+1)^4 n_1^4 . \] (A2.9)

From \( \mu_2 = \mu_2 + (\mu_1')^2 \) and (A2.4),

\[E(R_1^2) = n_1 (N+1) (N-n_1) / 12 - \left( n_1 (N-n_1) / 12N(N-1) \right) \left( \sum_{j=1}^{r} (t_j^3 - t_j) \right) \]

\[+ n_1^2 (N+1)^2 / 4 . \] (A2.10)

For \( E(R_{11}^2) \), \( i \neq i' \), we note the following algebraic results:

\[(X_1 + \ldots + X_n)^2 = \sum_{i=1}^{N} X_i^2 + 2 \sum_{i < i'} X_i X_{i'} , \text{ and so} \] (A2.11)

\[(X_1 + \ldots + X_n)^4 = \sum_{i=1}^{N} X_i^2 \sum_{i=1}^{N} X_i^2 + 4 \sum_{i=1}^{N} X_i \sum_{i < i'} X_i X_{i'} + 4 \sum_{i=1}^{N} X_i \sum_{i < i'} X_i X_{i'} + \sum_{i < i'} \sum_{i < i'} X_i X_{i'} . \] (A2.12)

\[\sum_{i=1}^{N} X_i^2 = \sum_{i=1}^{N} X_i^2 \sum_{i=1}^{N} X_i^2 . \] (A2.13)

\[\sum_{i=1}^{N} X_i = \sum_{i=1}^{N} X_i^3 , \text{ and so} \] (A2.14)

\[\sum_{i<1'} X_i X_{1'} = \sum_{i<1'} X_i X_{1'} X_i X_{1'} . \] (A2.15)

Further,

\[\sum_{i=1}^{N} X_i^3 = \sum_{i=1}^{N} X_i^4 + \sum_{i \neq 1'} X_i X_{1'} . \] (A2.16)

Taking the rank scores to be the \( X_i \),

\[\sum_{i=1}^{N} X_i = N(N+1)/2 \text{ and} \]
\[
\sum_{i=1}^{N} x_i^2 - \frac{(\sum x_i)^2}{N} = \frac{N(N^2-1)}{12} - \frac{x(t^3-t_j)}{j=1}^{12}.
\]

Hence,
\[
\sum_{i=1}^{N} x_i^2 = \frac{1}{12}[N(N^2-1) + 3N(N+1)^2 - \sum_{j=1}^{r}(t_j^3-t_j)]
= A_1.
\]

Since \((\sum x_i)^2 = N^2(N+1)^2/4\), using (A2.11)
\[
\sum_{i<1',j} x_{i} x_{j} = N^2(N+1)^2/8 - 1/24[N(N^2-1) + 3N(N+1)^2 - \sum_{j=1}^{r}(t_j^3-t_j)]
= N^2(N+1)^2/8 - A_1/2.
\]

(A2.17)

From (A2.12), we have
\[
(\sum x_i)^4 = A_1^2 + 4A_1[N^2(N+1)^2/8-A_1/2] + 4[N^2(N+1)^2/8-A_1/2]^2.
\]

(A2.18)

We can write
\[
\sum_{i=1}^{N} x_i^2 = \sum_{j=1}^{r} [t_j(\sum_{k=1}^{j-1} t_k + \frac{t_{j+1}}{2})^2],
\]

(A2.19)

and similarly
\[
\sum_{i=1}^{N} x_i^3 = \sum_{j=1}^{r} [t_j(\sum_{k=1}^{j-1} t_k + \frac{t_{j+1}}{2})^3], \quad \text{and}
\]

\[
\sum_{i=1}^{N} x_i^4 = \sum_{j=1}^{r} [t_j(\sum_{k=1}^{j-1} t_k + \frac{t_{j+1}}{2})^4].
\]

(A2.20)

(A2.21)

Now solving for \(\sum_{i \neq 1'} x_i x_{1'}\), \(\sum_{i \neq j', j} x_i x_{j}\), and \(\sum_{i < 1', j < j'} x_{i} x_{j}\), and \(\sum_{i \neq 1'} \sum_{i \neq j', j} \sum_{j < j'} x_i x_{j} x_{1'} x_{j'}\).

from (A2.13), (A2.14), (A2.15) and (A2.16), in terms of
\[
\sum_{i=1}^{N} x_i^2, \sum_{i < 1'} x_{1'}, \sum_{i=1}^{N} x_i^3, \quad \text{and} \quad \sum_{i=1}^{N} x_i^4, \quad \text{and} \quad N,
\]

we have from (A2.13)
\[
\sum_{i \neq 1'} \sum_{i \neq j', j} \sum_{j < j'} x_i x_{j} x_{1'} x_{j'}, \quad \text{in terms of}
\]
\[
\sum_{i=1}^{N} x_i^2, \sum_{i < 1'} x_{1'}, \sum_{i=1}^{N} x_i^3, \quad \text{and} \quad \sum_{i=1}^{N} x_i^4, \quad \text{and} \quad N,
\]

\[
\sum_{i \neq 1'} \sum_{i \neq j', j} \sum_{j < j'} x_i x_{j} x_{1'} x_{j'} = \{(\sum x_i^4) + [\sum x_i^2] \}
\]

(A2.22)

From (A2.14), (A2.16), and (A2.11)
\[
\sum_{i \neq j} \frac{1}{i} \sum_{j \neq i} x_1^2 x_j x_i = - \sum_{i=1}^{N} x_1^3 i + \sum_{i=1}^{N} x_1^4 i + \sum_{i=1}^{N} x_1^2 \sum_{i \neq j} x_j x_i.
\] (A2.23)

From (A2.15),
\[
\sum_{i < j} x_1 x_1 x_i x_j = \frac{N N}{i < j} \sum_{i=1}^{N} x_1^2 - \frac{N N}{i < j} \sum_{i=1}^{N} x_1^2 - \frac{N N}{i < j} \sum_{i=1}^{N} x_1^2 x_i x_j,
\]
\[
= \frac{1}{2} \left( \frac{1}{8} (N^2 (N+1)^2 - 4 \sum_{i=1}^{N} x_i^2)^2 \right) - \frac{1}{4} \left( \sum_{i=1}^{N} x_i^4 + \sum_{i=1}^{N} x_i^2 \right)
\]
\[
+ \frac{N N}{i=1} x_1^3 i + \sum_{i=1}^{N} x_1^2 \sum_{i=1}^{N} x_i x_j + \frac{N N}{i=1} x_1^2 \sum_{i=1}^{N} x_i^2 x_j.
\] (A2.24)

Following Kruskal's evaluation of \( E(R_{1 R_{1,1}}^2) \), i.e. (A2.3),
\[
E(R_{1 R_{1,2}}^2) = \frac{8 n_1 n_2 (n_1 - 1)(n_2 - 1)}{N(N-1)(N-2)(N-3)} \sum_{i \neq j} x_1 x_i x_j x_j,
\]
\[
+ \frac{2 n_1 n_2 (n_1 + n_2 - 2)}{N(N-1)(N-2)} \sum_{i \neq j} x_1 x_i x_j x_j + \frac{n_1 n_2}{N(N-1)} \sum_{i \neq j} x_1 x_i x_j x_j.
\] (A2.25)

Hence, \( \text{Var}(H) \), conditional on an observed set of ties, is as in (A2.1), where \( E(R_{1}^4) \) is given by (A2.9), \( E(R_{1 R_{1,2}}^2) \) can be evaluated through equations (A2.11) through (A2.25), and \( \text{Var}(H) \) is given by (A2.10).

Dividing and summing (A2.9) as indicated in (A2.1),
\[
\sum_{i=1}^{N} \frac{1}{n_i} \frac{1}{2} ER_{1}^4 = \frac{3 N^5 \left( \sum_{j=1}^{N} \frac{1}{2} t_j \right)^2 - \left( SN^2 - 2N^2 + \frac{1}{2} n_1 \sum_{i=1}^{N} + SN - S \right)}{(N-1)(N-2)(N-3)}.
\]
\[
- \frac{N^3}{(N-1)(N-2)(N-3)} \left( 7 N^2 S - 12 N^2 + NS + 6 \right) \frac{n_1}{N} \left( N_1^2 - N_1^3 \right) \frac{1}{1}.
\]
\[
- \frac{2 (N+1)^2}{(N-1)(N-2)} \left[ \sum_{j=1}^{N} b_j (1-b_j) \left( \frac{1}{2} t_j \right) \right] \left( SN^2 - 2N^2 + \frac{1}{2} n_1 \sum_{i=1}^{N} \right).
\]
\[
+ \frac{1}{8} (N+1) \left[ (N+1) \frac{1}{N(N-1)} \left( \frac{1}{2} t^2 \right) \right] \left( N_1^2 - n_1 \sum_{i=1}^{N} \right) + \frac{(N+1)^4}{16} \frac{n_1}{N}.
\] (A2.26)
Similarly, generalizing (A2.25) to $E(R_i^2 R_j^2)$, dividing by $n_i n_j$, and summing over $i \neq j$, we have

$$
\sum_{i \neq j} \frac{1}{n_i n_j} E(R_i^2 R_j^2) = \frac{1}{N(N-1)} \left\{ \frac{8 \left[ \frac{1}{128} (N^2(N+1)^2 - 4\sum_{i=1}^N X_i^2) + \sum_{i=1}^N X_i^2 \right]}{(N-2)(N-3)} 
+ \frac{2}{(N-2)} \left[ -\sum_{i=1}^N X_i^3 \sum_{i=1}^N X_i + \frac{N}{2} \sum_{i=1}^N X_i^4 \right] + \frac{(N-1)(N-S)}{2(N-S)} \sum_{i=1}^N X_i X_j \right\}
+ \left[ -\frac{N}{2} X_1^4 + \left( \frac{N}{2} X_1^2 \right)^2 \right] \right\}. \tag{A2.27}
$$

Dividing and summing (A2.10),

$$
\sum_{i=1}^N \frac{1}{n_i} E R_i^2 = \frac{N(N+1)(S-1)}{12} - \frac{1}{12(N-1)} \left( \frac{(S-1)(t_j^3 - t_j)}{N} \right) + \frac{N(N+1)^2}{4}, \tag{A2.28}
$$

Recalling that the $X_i$ are the assigned rank scores and that

$$
b_j = \sum_{j'=1}^{j-1} \frac{(t_{j'} + t_j/2)}{N}, \tag{A2.26}
$$

and substituting (A2.26), (A2.27), and (A2.28) into (A2.11), we have an expression for $\text{Var}(H)$, in $X_i, t_j, b_j, S, N$, and $n_i$. \hfill \blacksquare
A COMPUTER PROGRAM FOR EVALUATING THE CONDITIONAL VARIANCE OF THE KRUSKAL-WALLIS H STATISTIC IN THE PRESENCE OF TIES

5 DIMENSION T(10),A(10),B(10),N(8),Q(8)
10 INTEGER S,R
12 SUM=0.0
13 SUMSQ=0.0
20 INPUT, S,R,TN
25 INPUT, (N(I),I=1,S)
27 DO 100 KKK=1,40
28 PRINT++
30 INPUT, (T(J),J=1,R)
31 NN=0
32 DO 53 J=1,R
33 53 NN=NN+T(J)
34 PRINT,"CHECK TOTAL OF ENTRIES = ",NN
35 J=1
40 A(J)=(T(J)+1.)/2.
45 B(J)=T(J)/(2.*TN)
50 DO 1 J=2,R
55 A(J)=A(J-1)+(T(J-1)+T(J))/2.

60 \( B(J) = B(J-1) + (T(J-1) + T(J)) / (2. * TN) \)
65 1 CONTINUE
80 SNI2 = 0.0
85 SINI = 0.0
90 DO 2 I = 1, S
95 SNI2 = N(I) * N(I) + SNI2
100 SINI = 1. / N(I) + SINI
105 2 CONTINUE
115 SX1 = 0.0
120 SX2 = 0.0
125 SX3 = 0.0
130 SX4 = 0.0
135 DO 3 J = 1, R
140 SX1 = T(J) * A(J) + SX1
145 SX2 = T(J) * A(J) * A(J) + SX2
150 SX3 = T(J) * A(J) * A(J) * A(J) + SX3
155 SX4 = T(J) * A(J) * A(J) * A(J) * A(J) + SX4
156 3 CONTINUE
165 Q(1) = 0.0
170 Q(2) = 0.0
175 Q(3) = 0.0
180 Q(8) = 0.0
185 DO 4 J = 1, R
190 Q(2) = (B(J) - 0.5) * (B(J) - 0.5) * (B(J) - 0.5) * (B(J) - 0.5) * T(J) + Q(2)
195 Q(3) = B(J) * (1 - B(J)) * (B(J) - 0.5) * T(J) + Q(3)
200 Q(1) = B(J) * B(J) * T(J) / TN + Q(1)
205 Q(8) = T(J) * T(J) * T(J) - T(J) + Q(8)
4 CONTINUE
Q(1)=Q(1)-.25
Q(4)=TN+1.-1/(TN*(TN-1.))*Q(8)
Q(7)=-SX4+sx2*sx2
Q(6)=-sx3*sx1+sx4+sx2*(sx1*sx1-sx2)/2.
Q(5)=TN*TN*(TN+1.)*(TN+1.)-4.*sx2
Q(5)=1./128.*Q(5)*Q(5)-Q(7)/4.-Q(6)
F=144./(TN*TN*(TN+1.)*(TN+1.))
TERM11=3.*F*TN*TN*TN*TN*Q(1)*Q(1)/((TN-1.)*(TN-2.))
TERM11=TERM11/(TN-3)
TERM11=TERM11*((S-2.)*TN*TN+SNI2-TN*(TN-1.)*SNI1+S*TN-S)
TERM12=-F*TN*TN*Q(2)/((TN-1.)*(TN-2.)*(TN-3.))
TERM12=TERM12*((S-12.)*TN*TN+SNI2*SNI2-TN*(TN+1.)*SNI)
TERM13=2.*F*TN*TN*(TN+1.)*Q(3)*(S-3.)*TN*TN+2.*SNI2
TERM13=TERM13/((TN-1.)*(TN-2.))
TERM13=-TERM13
TERM14=(TN+1.)*(TN+1.)*F*Q(4)*(TN*SNI2)/8.
TERM14=TERM14+F*(TN+1.)*(TN+1.)*(TN+1.)*SNI2/16.
TERM1=TERM11+TERM12+TERM13+TERM14
TERM21=F*8.*Q(5)/((TN*(TN-1.))*(TN-2.))*(TN-3.))
TERM21=TERM21*((TN-S)*(TN-S)+2.*TN-S-SNI2)
TERM22=4.*F*Q(6)/((TN-2.)*(S-1.)*(TN-S))/((TN*(TN-1.))
TERM23=F*Q(7)*S*(S-1.)/(TN*(TN-1.))
TERM2=TERM21+TERM22+TERM23
TERM31=TN*(TN+1.)*(S-1.)/12.-(S-1.)*Q(8)/(12.*(TN-1.))
TERM31=TERM31+TN*TN*(TN+1.)*(TN+1.)/4.
TERM3=F*TERM31
340   VARH=TERM1/TERM31-TERM3
341   VARH=VARH+TERM2/TERM31
342   VARH=VARH*TERM31
345   C=1.-Q(8)/(TN*TN*TN-TN)
350   C=C*C
355   VARH=VARH/C
356   SUM=SUM+VARH
357   SUMSQ=SUMSQ+VARH*VARH
360   PRINT,"VAR H =",VARH
362   100 CONTINUE
363   HBAR=SUM/40
364   SVAR=(SUMSQ-HBAR*SUM)/39
365   PRINT,"HBAR = ",HBAR
366   PRINT,"HVAR = ",SVAR
367   END
LIST OF REFERENCES


