CANADIAN THESES ON MICROFICHE

THÈSES CANADIENNES SUR MICROFICHE

National Library of Canada
Collections Development Branch

Canadian Theses on
Microfiche Service

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada
Direction du développement des collections

Service des thèses canadiennes
sur microfiche

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

Canada

NL 339 (r. 85/09)

COMPLEXITY AND THE MODIFIED FORMS OF THE TUKEY TEST FOR PAIR-WISE COMPARISONS

by

Donald A. Dickie

Thesis submitted to the School of Graduate Studies
and Research in partial fulfillment of Doctor of
Philosophy (Education)

UNIVERSITY OF OTTAWA

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

ABSTRACT

This study sought to examine the behaviour of average complexity (Shoffer 1981) in relation to experiment-wise error rate and number of rejections of null hypotheses under the condition of sample size disparity, variance heterogeneity, mean variability, and inverse and direct pairing. The Kramer (1956) method and the Games and Howell (1976) technique were used to represent those techniques dealing with sample size imbalance or heterogeneous variances, respectively.

The method involved the use of a Monte Carlo technique to examine the behaviour of all possible estimated pair-wise contrasts in a four group experimental design.

The results indicated that average complexity did not reflect changes apparent in experiment-wise error and average number of rejections of null hypotheses. Yet examination of the underlying pattern frequencies and the frequencies of significant pair-wise contrasts indicated a source of information valuable to the educational researcher.

It was concluded that the concept of complexity had merit, but its usefulness was camouflaged by the manner in which average complexity was calculated.

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my advisor, Dr. James E. Carlson, whose enthusiastic support and guidance enabled me to complete this thesis. I would also like to acknowledge the support and advice of Dr. Marvin Boss and Dr. Jean-Paul Dionne without which this study would not have been completed. Finally, I would like to express my thanks to Ms. Bev Greene, who taught me the difference between academic writing and creative and enjoyable writing. Her support, not only in the final production of this thesis, but also throughout the Doctoral program, allowed me to make the jump from theory to application.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

With the advent of computer technology, educational research has become increasingly complex. The use of the computer allows more information to be processed, and as a result the number and size of the experimental groups have increased proportionately.

Whenever an experiment involves the collection of data from more than two groups or in more than two conditions, a significant F test is automatically only the first step in the analytic procedure. If the overall hypothesis of equality of means is rejected, an experimenter still faces the problem of deciding which of the means is not equal. Unfortunately, classical methods such as the F test do not provide procedures for comparison of specific means with one another in a multi-group design. As a result, multiple comparison procedures were developed to permit comparison of each experimental mean with every other mean, thus allowing the researcher to identify the exact source of the significant differences.

Many multiple comparison procedures have been proposed in the literature. Games (1971a), O'Neil and Wetherill (1971), and Miller (1966, 1977) provide excellent reviews of the area. The nature of educational research places the educational researcher in a position where he/she must choose from a wide variety of multiple comparison techniques, each of which is suitable for certain conditions. Of specific interest is the situation in which a researcher wishes to examine all pair-wise comparisions in a given experimental design.

Tukey (1953) (cited in Kirk 1982) suggested a multiple range test that was called the wholly significant or honestly significant test. The joint confidence interval estimates of the quantities $\mu_i - \mu_j$ were

$$\bar{X}_i - \bar{X}_j \pm SR_{\alpha,k,N-k} \, S/n^{1/2} \tag{1.1}$$

where $\bar{X}_i$ denotes the sample mean for the ith treatment, $SR_{\alpha,k,N-k}$ is the $100(1-\alpha)$ point of the studentized range distribution of k normal variates and S^2 is the unbiased estimate of σ^2 based on N-k degrees of freedom,

$$S^2 = \frac{\sum_i^k \sum_j^n (X_{ij} - \bar{X}_{i.})^2}{N-k}, \tag{1.2}$$

$$N = \sum_{i=1}^k n_i . \tag{1.3}$$

Tukey's procedure assumes equal sample sizes--a necessary assumption in order for the sample means to have the same variance, σ^2/n. The procedure also assumes homogeneity among the k population variances.

Unfortunately, many educational studies cannot satisfy these assumptions: experimenters often find themselves with samples that are unpredictably unequal in size (due to subject loss) and this inequality suddenly prohibits use of Tukey's pair-wise comparison procedure.

Heterogeneity of variances also creates difficulties, in that parent populations from which the samples are drawn frequently have very different population variances; in other conditions, the estimated sample variances are simply not homogeneous. In both situations, the researcher is left with sample variance estimates which may be grossly heterogeneous and Tukey's original test is no longer applicable.

Since the formulation of the Tukey technique in 1953, modifications have been proposed for both the unequal n case and the heterogeneous variance situation. Modified Tukey procedures have been researched at length to compare their usefulness under varying conditions of sample size imbalance and/or variance heterogeneity.

The most popular criterion employed by these researchers has been the type I error rate (defined as the probability of rejecting the null hypothesis of no difference when it is in fact true); a second criterion, often found in the literature, has been the size of the confidence interval generated by the procedure for the pair-wise comparison . A third criterion--though not as popular as the two identified and one which is related to the width of the confidence interval--has been the comparison of power, or the probability of accepting the alternate hypothesis when it is in fact true.

These three criteria have formed the framework for the comparison of the various modified forms of the Tukey procedure, and as such, have provided researchers with guidelines for the selection of appropriate techniques for data analysis. Though the importance of these criteria cannot be understated, their use is limited to a judgement regarding the precision of the technique, or to the probability of making a type I error.

Of greater practical importance is a criterion related to interpretation of the results. A primary consideration for the educational researcher is his/her capacity to make inferences about differences among parent populations based upon the pattern of results: the lack of clear cut differences among sample means creates numerous interpretative problems for the researcher.

A new criterion has been proposed by Shaffer (1981), which deals specifically with the interpretation. In a k group experimental design, the population means fall into a specific configuration. This configuration may be represented by partitioning the k samples into subsets. Sample means within any given subset are not significantly different from each other, while comparison of two sample means from different subsets results in a significant estimated contrast. If as a result of a multiple comparison procedure, the k experimental groups could be formed into this configuration, the researcher is presented with very favourable conditions for interpretation. It is obvious that this situation does not always occur, and the possible number of clear cut configurations of the means is, of course, limited. Shaffer's criterion--called complexity--measures how close a given configuration of sample means is to an ideal configuration of means. Different multiple comparison techniques would create different frequencies of various mean configurations, thus resulting in different complexities. A comparison of various procedures is therefore possible using the concept of complexity.

The use of this new criterion does not negate the value of the previous criteria but rather helps in creating a more complete picture for the educational researcher; the researcher himself must weigh the relative importance of type I error rate, confidence interval width, and complexity, in his selection of a specific technique. In order to compare different criteria, relationships between complexity and the other criteria must be examined closely. The modifications of the Tukey technique present an ideal framework for consideration of complexity. Research indicates that with respect to type I error rates, the two classes of Tukey modifications (unequal sample size, and heterogeneous variance) behave differently under conditions of sample size disparity and unequal variances. The purpose of the present study is to examine how complexity changes under various conditions of sample size disparity and variance heterogeneity relative to changes in type I error rates.

Chapter I contains a review of the relevant literature dealing with the behaviour of the different modified Tukey techniques under varying conditions of sample size disparity and variance heterogeneity. Chapter II describes the procedure used in the present study. Chapter III presents the results of the study. Chapter IV follows with a discussion of the results and Chapter V, Conclusions, completes the study.

CHAPTER I

REVIEW OF LITERATURE

Due to the plethora of techniques presented in the literature, a survey of techniques will be considered first; following this, a discussion of the criteria for evaluating multiple comparison procedures, along with research comparing the techniques under various conditions will be presented.

MODIFICATIONS OF THE TUKEY TECHNIQUE FOR PAIR-WISE COMPARISONS

The modifications of the Tukey technique can be divided into two types: those dealing with unequal sample sizes and those dealing with unequal variances. Invariably the latter combine unequal sample sizes with unequal variances. The two types of techniques will be considered separately in the following sections.

TECHNIQUES DEALING WITH UNEQUAL SAMPLE SIZES

Kramer (1956)

A method for adapting the Tukey test to unequal sample sizes was proposed by Kramer (1956) and, according to Dunnett (1980a), was also proposed by Tukey (1953). Kramer's method replaces the n in (1.1) with the harmonic mean of the sample sizes involved in the contrast, to yield the following confidence interval estimate:

$$\overline{X}_i - \overline{X}_j \pm SR_{\alpha,k,N-k}\, s([n_i^{-1} + n_j^{-1}]/2)^{\frac{1}{2}} . \qquad (2.1)$$

The above confidence interval reduces to the Tukey test in the case of equal sample sizes. The right hand portion of the Tukey interval can be written as $SR_{\alpha,k,N-k}/\sqrt{2}$ times the standard error of the difference between two sample means, the latter becoming $s(n_i^{-1}+n_j^{-1})^{1/2}$, in the unequal sample case.

<u>Winer (1962), Miller (1966)</u>

Miller (1966) advised strongly against the use of Kramer's (1956) method and instead recommended that the median or average value of the sample sizes replace the n of Tukey's technique. This recommendation was repeated in Miller (1977).

Winer (1962, 1971) and Kirk (1968) have suggested the use of the harmonic mean of the sample sizes. In this case, the n is replaced by the following:

$$\tilde{n} = \frac{k}{(1/n_1 + 1/n_2 + \ldots + 1/n_k)} \qquad (2.2)$$

<u>Spjotvoll and Stoline (1973)</u>

Spjotvoll and Stoline (1973) presented a method they referred to as the T' method. This method differs from Kramer's method in that it uses the maximum value of the reciprocals of the two sample sizes involved in the comparison, rather than the sum of the reciprocals,

$$\bar{X} - \bar{X} \pm SAR_{\alpha,k,N-k}\, s(\max(n_i^{-1}, n_j^{-1}))^{1/2}, \qquad (2.3)$$

where $SAR_{\alpha,k,N-k}$ is the 100(1-α) point of the studentized augmented range distribution. Tables for this distribution are available for α = .01, .05, .1 and .2 in Stoline (1978).

Hochberg (1974)

The GT2 method was proposed by Hochberg (1974) and incorporates the studentized maximum modulus distribution,

$$\overline{X}_i - \overline{X}_j \pm smm_{\alpha,k^*,N-k}\, s(n_i^{-1}+n_j^{-1})^{1/2} , \qquad (2.4)$$

where $smm_{\alpha,k^*,N-k}$ is the 100(1-α) point of the studentized maximum modulus (or multivariate t with common correlation coefficient ρ = 0) distribution with

$$k^* = k\,(k-1)/2 \ . \qquad (2.5)$$

Tables for α = .01, .05, .1, and .2 and k = 3 to 20 were published by Stoline and Ury (1979). The justification for using the studentized maximum modulus is based on an inequality due to Sidak (1967).

Genizi and Hochberg (1978)

This method was specifically devised to be optimum within a class of procedures proposed by Hochberg (1975) that are known to achieve the desired error rate for the case where the sample sizes have only two distinct values $(n_1=n_2=n_3=\ldots=n_m=a, n_{k-m}=\ldots n_k=b)$,

$$\overline{X}_i - \overline{X}_j \pm SR_{\alpha,k,N-k}\, gs((n_i^{-1}+n_j^{-1})/2)^{1/2} , \qquad (2.6)$$

where g=1 for $n_i = n_j$ and g>1 if $n_i \neq n_j$ (size of g a function of sample sizes). Thus, the method gives limits that are identical to the Kramer (1956) method if two means with the same sample size are compared; if the two means have different sample sizes, the limits are wider. Dunnett (1980a) provided a table for implementing this procedure in the situation where $n_1 = a, n_2 = n_3 = \ldots n_k = b$

Gabriel (1978)

Gabriel, like Hochberg, incorporated the studentized maximum modulus distribution in his confidence interval.

$$\overline{X}_i - \overline{X}_j \pm smm_{\alpha,k^*,N-k}\, s\left((2n_i)^{-1/2} + (2n_j)^{-1/2}\right) \qquad (2.7)$$

Gabriel's procedure coincides with Hochberg's (1974) procedure when the n_i's are equal, and presents certain advantages if graphical comparisons between the means are desirable.

Scheffé (1953)

Scheffé's S method is applicable to the estimation of all contrasts of the form

$$\Psi = \sum_{i=1}^{k} c_i \mu_i \text{ where } \sum_{i=1}^{k} c_i = 0 \text{ for } \mu_1, \ldots, \mu_k \qquad (2.8)$$

When considering just pair-wise contrasts, the S method yields the following 100% (1 - α) joint confidence interval:

$$\overline{X}_i - \overline{X}_j \pm S\left[(K-1)(F_{\alpha,k-1,N-k})(n_i^{-1} + n_j^{-1})\right]^{1/2} \qquad (2.9)$$

where $F_{\alpha,k-1,N-k}$ is the 100 (1 - α) point of the F distribution with k-1 and N-k degrees of freedom.

Bonferroni Method

The Bonferroni inequality (Miller 1966) serves as the basis for the well-known and extensively used B method. The 100 (1 - α) confidence interval for $\mu_i-\mu_j$ is

$$\overline{X}_i-\overline{X}_j \pm \left(t_{\alpha/2k^*,N-k}\right) s \left(n_i^{-1}+n_j^{-1}\right)^{1/2} \qquad (2.10)$$

where $t_{\alpha,N-k}$ is the upper α point of the central t distribution with N-k degrees of freedom and k* is as defined in (2.5). This assumes that all pair-wise contrasts are to be examined....something which is not really necessary with the Bonferroni technique.

Dunn-Šidák Method

An improved B-type inequality presented by Šidák (1967) was used by Dunn (1974) to produce the conservative Dunn-Šidák method,

$$\overline{X}_i-\overline{X}_j \pm \left(t_{\alpha^*,N-k}\right) s \left(n_i^{-1}+n_j^{-1}\right)^{1/2} \qquad (2.11)$$

where $\alpha^*=\frac{1}{2}-\frac{1}{2}(1-\alpha)^{1/k^*}$ and where k* is defined as in (2.5). Special t tables for the Dunn-Šidák method were tabulated by Games (1977).

TECHNIQUES DEALING WITH UNEQUAL VARIANCES

Hochberg (1976)

The extended form of the Tukey test proposed by Spjøtvoll and Stoline (1973) was subsequently generalized to include heterogeneous variances by Hochberg (1976). The form reported here was modified by Keselman and Rogan (1978) to use the studentized range rather than the critical value originally proposed.

$$\overline{X}_i - \overline{X}_j \pm SAR_{\alpha,k,N-k} \max\left(s_i(n_i^{-1}), s_j(n_j^{-1})\right) \tag{2.12}$$

where

$$s_i^2 = \frac{\sum_j (X_{ij} - X_{i.})^2}{n_i - 1} \tag{2.13}$$

Games and Howell (1976)

Games and Howell (1976) have suggested adopting the Behrens-Fisher statistic with Welch's (1947) approximate t solution for the degrees of freedom. The Behrens-Fisher solution satisfactorily controls the rate of type I error on any one contrast when sample sizes and variances are unequal, and provides a powerful statistical test (Mehta and Srinivasan 1970, Wang 1971). The Games and Howell procedure is of the form

$$X_i - X_j \pm SR_{\alpha,k,\hat{\nu}_{ij}} \left(s_i^2(2n_i)^{-1} + s_j^2(2n_j)^{-1}\right)^{1/2} \tag{2.14}$$

where $\hat{\nu}_{ij}$ denotes the Welch (1938) and Satterthwaite (1946) approximate degrees of freedom formula,

$$\hat{\nu}_{ij} = \frac{\left(s_i^2(n_i)^{-1} + s_j^2(n_j)^{-1}\right)^2}{s_i^4(n_i^2\nu_i)^{-1} + s_j^4(n_j^2\nu_j)^{-1}} , \tag{2.15}$$

where

$$\nu_i = n_i - 1 \tag{2.16}$$

C Procedure

The C procedure is an extension of Cochran's (1964) method proposed by Dunnett (1980b). According to Dunnett, this procedure is the analog of Cochran's method for the case of k=2, in the same manner that the Games and Howell method is the analog for the use of Welch's approximate degrees of freedom for obtaining approximate confidence intervals for the difference between two means based on Student's t. Cochran (1964) developed his method (for the case of k=2) as an approximation for Sukhatme's (1938) solution to the Behrens-Fisher problem (Behrens 1964),

$$\overline{X}_i - \overline{X}_j \pm SR_{\alpha,k,\nu_{ij}^*}\left(s_i^2(2n_i)^{-1} + s_j^2(2n_j)^{-1}\right)^{1/2} \tag{2.17}$$

where

$$SR_{\alpha,k,\nu_{ij}^*} = \frac{SR_{\alpha,k,\nu_i} s_i^2(n_i)^{-1} + SR_{\alpha,k,\nu_j} s_j^2(n_j)^{-1}}{s_i^2(n_i)^{-1} + s_j^2(n_j)^{-1}} \tag{2.18}$$

which corresponds to the weighted average of Student's t proposed by Cochran (1964) but extended by Dunnett (1980b) to the case of $k > 2$.

Tamhane (1977, 1979)

Tamhane proposed the T2 method using the t distribution with a significance level based on an inequality proposed by Sidak (1967) and degrees of freedom based upon the approximation suggested by Welch (1938, 1947),

$$\bar{X}_i - \bar{X}_j \pm t_{\gamma,\hat{v}_{ij}} \left(s_i^2(n_i)^{-1} + s_j^2(n_j)^{-1}\right)^{1/2} \qquad (2.19)$$

where

$$\gamma = 1-(1-\alpha)^{1/k^*}, \qquad (2.20)$$

$$\hat{v}_{ij} = \frac{\left(s_i^2(n_i)^{-1} + s_j^2(n_j)^{-1}\right)^2}{s_i^4(n_i^2\nu_i)^{-1} + s_j^4(n_j^2\nu_j)^{-1}} \,. \qquad (2.21)$$

T3 Procedure

Dunnett (1980b) modified Tamhane's T2 procedure with the substitution of the studentized maximum modulus distribution for k* uncorrelated normal variates with $\hat{v}_{ij}$ degrees of freedom.

$$\bar{X}_i - \bar{X}_j \pm smm_{\alpha,k^*,\hat{v}_{ij}} \left(s_i^2(n_i)^{-1} + s_j^2(n_j)^{-1}\right)^{1/2} \qquad (2.22)$$

The use of the studentized maximum modulus distribution was based upon Sidak's (1967) uncorrelated t inequality on which the T2 procedure was also based.

In summary, fifteen statisical methods that could be used for pair-wise comparisons were presented. Three of the methods (Scheffé, Bonferroni, Dunn-Šidák) are not derivatives of the Tukey technique but were presented for comparison because of their prevalence in past research.

Of the remaining techniques, four from each of the classes dealing with sample size imbalance and variance heterogeneity are listed in Table 1 and 2 respectively. The techniques are categorized by the estimated contrast, the table value and the standard error term. As a result, the structure of the techniques may be compared as to where differences occur.

TABLE 1

COMPARISON OF TECHNIQUES DEALING WITH UNEQUAL SAMPLE SIZES

Technique	Estimated Contrast	Table Value	Standard Error Term
Kramer	$\overline{X}_i - \overline{X}_j$	$SR_{\alpha,k,N-k}$	$s([n_i^{-1}+n_j^{-1}]/2)^{1/2}$
Spjøtvoll and Stoline	$\overline{X}_i - \overline{X}_j$	$SAR_{\alpha,k,N-k}$	$s(\max(n_i^{-1},n_j^{-1}))^{1/2}$
Hochberg (1974)	$\overline{X}_i - \overline{X}_j$	$smm_{\alpha,k^*,N-k}$	$s(n_i^{-1}+n_j^{-1})^{1/2}$
Gabriel	$\overline{X}_i - \overline{X}_j$	$smm_{\alpha,k^*,N-k}$	$s((2n_i)^{-1/2}+(2n_j)^{-1/2})$

TABLE 2

COMPARISON OF TECHNIQUES DEALING WITH UNEQUAL VARIANCES

Technique	Estimated Contrast	Table Value	Standard Error Term
Games and Howell	$\overline{X}_i - \overline{X}_j$	$SR_{\alpha,k,\hat{\nu}_{ij}}$	$(s_i^2(2n_i)^{-1}+s_j^2(2n_j)^{-1})^{1/2}$
Hochberg (1976)	$\overline{X}_i - \overline{X}_j$	$SAR_{\alpha,k,N-k}$	$\max(s_i(n_i^{-1}),s_j(n_j^{-1}))$
C Procedure	$\overline{X}_i - \overline{X}_j$	SR_{α,k,ν_{ij}^*}	$(s_i^2(2n_i)^{-1}+s_j^2(2n_j)^{-1})^{1/2}$
T3 Procedure	$\overline{X}_i - \overline{X}_j$	$smm_{\alpha,k^*,\hat{\nu}_{ij}}$	$(s_i^2(n_i)^{-1}+s_j^2(n_j)^{-1})^{1/2}$

The techniques listed in Tables 1 and 2 seem to be the most popular forms based on the research to be discussed in the following sections. Type I error rate and robustness of the studentized range distribution will be considered before discussing research dealing with the specific forms of the modified Tukey technique.

TYPE I ERROR RATE

The concept of significance level has been a useful tool when dealing with a single difference between two means. Assessment of significance allows the researcher to adequately appreciate the probability of rejecting the null hypothesis when it in fact is true (type I error), thus tempering his/her conclusions with knowledge of the possible occurrence of a type I error.

The meaning of significance level becomes confused, however, when simultaneous statements about a number of different comparisons of means are made; this confusion may be due to the notion that significance level must be extended in several different directions when considering multiple comparisons. The necessity for controlling the type I error rate for all comparisons being made (either individually or in a group manner) further complicates the confusion.

O'Neill and Wetherill (1971), Balaam and Federer (1965), Steel (1961), and Hartley (1955) have proposed two bases which can be used for calculating error rates. Each inference is treated independently insofar as errors are concerned, to calculate an error rate per comparison. In a large experiment where many inferences are concerned, one would therefor expect typically more errors than in a smaller experiment. Viewed collectively, the error rate for the set of inferences can increase dramatically as the number of inferences is increased, thus making the probability that the whole set of inferences is correct, negligible.

Experiment-wise error rate has been defined as the probability that one or more erroneous conclusions will be drawn in a given experiment (Ryan 1959, O'Neill and Wetherill 1971). In this case, the experiment is divided into two classes: (a) those in which all conclusions are correct; and, (b) those in which some conclusions are incorrect. The experiment-wise error rate is the probability that a given experiment belongs to the latter category.

These probabilities (per comparison error rate and experiment-wise error rate) can be calculated as follows (O'Neill and Wetherill 1971):

$$\text{comparison error rate} = \frac{\text{number of erroneous inferences}}{\text{number of inferences}}$$

$$\text{experiment-wise error rate} = \frac{\text{number of experiments with one or more erroneous inferences}}{\text{number of experiments}} .$$

The two error rates are the same in a simple experiment with a single comparison, but they become more and more divergent as the number of comparisons per experiment increases. The problem becomes a decision as to which error rate should be kept under control and/or what compromises may be effected (Ryan 1959) (i.e., which error rate yields the best representation of the dependability of the data?).

When an experiment involves more than two groups, a complete null hypothesis refers to the hypothesis that all means were drawn from a single population. This is but one possibility of a set of all possible nulls dealing with any combination of means drawn from any combination of populations. For each different null, there is an error rate per comparison and per experiment-wise, which raises the question, which of these null hypotheses is used to define the error rate for the statistical test. Tukey (1953) (cited in Ryan 1959) proposed to define the error rate as the maximum value it attains under all possible null hypotheses. This would seem to be the most cautious approach to this problem.

Some procedures fix a given type I error rate per comparison; these are typically planned comparisons which control the error rate per contrast at the expense of having no control over the experiment-wise error rate. Post hoc contrasts on the other hand, control the experiment-wise error rate but often sacrifice power (Swaminathan and Gifford 1978). Multiple comparison procedures suggested by Scheffé, Tukey, Newman-Keuls and Duncan, control the experiment-wise error rate to differing degrees.

As a criterion for the evaluation of multiple comparison procedures, it would seem that most authors prefer the experiment-wise error rate. Einot and Gabriel (1975), Petrinovich and Hardyck (1969), Ramsey (1978), Ryan (1959) and Welsch (1977) each advocated the use of experiment-wise control of type I errors. Hummel and Sligo (1971), Miller (1966), Morison (1976), Ramsey (1980) and Timm (1975) suggested the use of experiment-wise type I error rates in multivariate experiments.

When using the experiment-wise type I error rate for simultaneous multiple comparisons, it should be remembered that the probability is related to a collective group of comparisons, and not a single one which may be of interest to the researcher. This concern was most aptly stated by Cox (1965):

> "The fact that a probability can be calculated for the simultaneous correctness of a large number of statements does not usually make that probability relevant for the measurement of uncertainty of one of the statements. If we are directly interested in a single statement about the vector parameter the probability of simultaneous correctness would however be appropriate. The practical usefulness of the multiple comparison techniques lies in giving a conservative bound for the effect of selection rather than in giving an "exact" solution."

Thus, the probability statement of the type I error applies to all multiple comparisons that may be performed by the experimenter and not just those of interest.

The use of experiment-wise type I error rate has been popular amongst researchers who have compared different multiple comparison techniques. Stoline (1981) has placed this criterion under the label of conservativeness. If α is equal to the nominal type I error probability, then a conservative procedure is one in which the confidence interval has a probability greater than 1 - α (ie. the actual probability of a type I error (experiment-wise) is less than the nominal level α). A liberal procedure, on the other hand, has a probability for its confidence interval less than 1 - α (the actual experiment-wise error probability exceeds the nominal α level).

Stoline (1981) continued to suggest three other considerations for the evaluation of multiple comparison procedures; the second criterion, optimality, deals with the size of the confidence interval. An optimum multiple comparison procedure produces the narrowest confidence intervals, the result being that more precise estimations of the possible values for the contrast are possible.

A more practical criterion is the convenience of the procedure. Convenience was evaluated in terms of its simplicity (easy to use), availability of tables and computer processing expenses.

The last criterion, robustness, considers the situation when the assumption of homogeneity of variances is violated and is discussed in the following section with regard to experiment-wise errors and power.

ROBUSTNESS OF THE STUDENTIZED RANGE DISTRIBUTION

The robustness of the studentized range statistics has been evaluated under conditions of variance heterogeneity and non-normality; the former case will be considered in more detail in a subsequent section. At present, only general evidence with respect to variance heterogeneity will be presented with slightly more emphasis being placed on the effects of non-normality.

Ramseyer and Tcheng (1973) and Brown (1974) present simulation evidence showing that the studentized range distribution generally shares the robustness properties of the central F distribution with respect to variance heterogeneity and non-normality (Scheffé 1959, ch. 10). Brown (1974) studied the effects of heterogeneity and non-normality of the population distribution on the sampling distribution of the studentized range: moderate heterogeneity of variance was found not to seriously affect the experiment-wise error rate. The probability of exceeding the nominal alpha level tended to increase with the number of groups examined. This was confirmed by Ringland (1983) who found that the maximum modulus and range distributions were ultimately non-robust as the number of groups was increased. The Tukey method was found to be fairly conservative and stable when contrasting fewer than five groups, whereas the Scheffé technique was generally conservative and stable when many comparisons were involved.

Brown (1974) also discovered that the studentized range distribution is affected by extreme non-normality. In this situation, the sampling distribution becomes more sharply peaked than under normality, and the tail probabilities are reduced as a result. Thus, the application of the studentized range method in the presence of non-normality may result in a multiple comparison procedure that is more conservative than anticipated (i.e., an actual error rate less than the nominal α level).

Ramseyer and Tcheng (1973) reported similar results in that the studentized range statistic, like the t and the F, withstood violations of the homogeneity of variance and non-normality assumptions when type I error rate was the criterion. Violations of the normality assumption using exponential and rectangular distributions resulted in rates systematically, but, negligibly, below nominal levels.

Hence, it could be concluded that the studentized range distribution is relatively robust under conditions of non-normality. In situations where non-normality is extreme, the range distribution generates critical values that are more conservative than in the normal case. The effect of heterogeneity of variance will depend greatly on the use of the specific multiple comparison method. The phenomenon will be discussed after a discussion of the research on techniques developed for unequal sample size. subsequently.

RESEARCH DEALING WITH THE CASE OF UNEQUAL SAMPLE SIZES

Based on the preceding section on multiple comparison procedures, it is evident that the modified Tukey or alternative techniques can be subdivided into those dealing with the case of unequal sample sizes and those dealing with unequal variances. This section will address the former condition.

The following techniques deal with unequal n's and are essentially derived from the Tukey method for equal sample sizes. They employ either the studentized range distribution, the studentized augmented range distribution or the studentized maximum modulus distribution. The techniques are as follows: Kramer (1956), Winer (1962), Miller (1966), Spjøtvoll and Stoline (1973), Hochberg (1974), Genizi and Hochberg (1978), and Gabriel (1978). The Scheffé (1953), Bonferroni and Dunn-Šidák methods are not derivations from Tukey but are included for the purposes of comparison.

Since most of these multiple comparison procedures can be expressed as studentized range statistics (Einot and Gabriel 1975) or alternatively as Student t statistics when the comparisons are pair-wise (Games 1971a), a comparison between their critical values and/or mathematical modifications can be made. Stoline (1981) has shown that:

$$\frac{SR_{\alpha,k,N-k}}{\sqrt{2}} \leq smm_{\alpha,k^*,N-k} \leq t_{\alpha^*,N-k} \qquad (2.23)$$

for the methods proposed by Kramer (1956), Hochberg (1974) and Dunn-Šidák. The remaining terms in the determination of the confidence intervals are equivalent for all three methods (i.e., $s(n_i^{-1}+n_j^{-1})^{\frac{1}{2}}$). Thus, it can be inferred that Kramer's method will produce confidence intervals as narrow as, if not narrower than, either those of Hochberg (1974) or Dunn-Šidák.

Stoline (1981) has also shown that

$$SR_{\alpha,k,N-k} \leq SAR_{\alpha,k,N-k} \tag{2.24}$$

and

$$\frac{n_i^{-1}+n_j^{-1}}{2} \leq \max(n_i^{-1},n_j^{-1}) \tag{2.25}$$

Therefore, Kramer's method will be as optimum as (if not more so) the method of Spjøtvoll and Stoline (1973).

The method proposed by Genizi and Hochberg (1978) will reduce to Kramer's method if the two sample sizes are equal; if this situation does not exist, the constant "g" will be greater than one (Dunnett 1980a). The confidence interval will also exceed that of the Kramer technique as a result.

Keselman, Murray and Rogan (1976), Keselman, Toothaker and Shooter (1975), Rogan, Keselman and Breen (1977) and Smith (1971) focused on assessing the rates of type I error and the sensitivity of the methods proposed by Winer (1962), Miller (1966) and Kramer (1956). Smith (1971) compared the three procedures and varied the disparity between group sizes for a 3:1 difference. Of the three procedures, Kramer's consistently provided experiment-wise error probabilities in close agreement with corresponding nominal probabilities. Kramer's method usually provided the lowest error probabilities, while Miller's (1966) technique had the largest. As the group size imbalance became more pronounced, Miller's method became more liberal.

Dunnett (1980a) has suggested that the methods of Winer (1962) and Miller (1966) suffer from the disadvantage that they provide allowances of fixed length for all comparisons, regardless of sample size. Hence, some of the individual comparisons would be more likely to exceed their allowances than others, and comparison error rates would therefore not be constant.

Keselman, Murray and Rogan (1976) extended Smith's (1971) study to include a sample size disparity of 40:1, as well as investigating the effects of the number of groups and the nominal significance levels. The results indicated that experiment-wise error probabilities seldom exceeded their nominal significance levels by more than 1%. The Kramer estimates were generally less than nominal α levels, while the harmonic means (Winer 1962) estimates were usually larger than the true alphas.

The results of these studies demonstrate that unequal group sizes did not greatly increase experiment-wise error probabilities of the Tukey test when either the Kramer or Winer modifications were employed. These authors have suggested the use of the Kramer (1956) method as opposed to the harmonic mean because the experiment-wise error probabilities were consistently lower than those of Winer's method. When power was considered (Keselman, Toothaker and Shooter 1975), the data did not favour either procedure.

Methods proposed by Spjøtvoll and Stoline (1973), Hochberg (1974), and Genizi and Hochberg (1978) have been shown analytically to guarantee that their experiment-wise error rates are equal to or less than α for any configuration of sample sizes. They differ, however, in their degree of conservativeness. Large variations in sample size from the equal sample size case resulted in the techniques generating different experiment-wise error probabilities. Dunnett (1980a) showed that Kramer's (1956) method had error probabilities which fluctuated around the .05 level; these were usually less for small variations in sample sizes. Larger sample size imbalance resulted in the Kramer method remaining conservative but close to nominal levels, while the harmonic mean method (Winer 1962) had experiment-wise error rates excessively large. The conservative nature of the methods of Spjøtvoll and Stoline (1973) and Genizi and Hochberg (1978) (relative to that of Kramer (1956)), was shown in the resulting smaller rates of experiment-wise error. The method of Gabriel (1978) also had smaller error rates than Kramer's except when the sample size imbalance was extreme; in this case Gabriel's method had shown a liberal tendency.

From Dunnett (1980a), it is apparent that for certain cases k > 3, the Kramer method is conservative for one way analysis of variance designs ranging from modest to severe sample size imbalance. If the Kramer method is liberal, it is only for analysis of variance designs where slight imbalance exists; in all such cases, the degree of liberality is minimal.

Stoline (1981) performed a comparison of the ratios of confidence interval lengths for ranges of k from 3 to 20 and degrees of freedom between 20 and ∞. The Kramer method exhibited large improvements over the Bonferroni for different nominal α levels. Improvement was also shown over the Hochberg (1974) technique. A limited comparison was performed with the method of Spjøtvoll and Stoline (1973) which showed improvements by the Kramer method to be more dramatic than when compared to the Bonferroni method. Similar results were also reported for a comparison with the Dunn-Šidák method.

Stoline (1981), like Dunnett (1980a), concluded that the Kramer method was to be preferred over most other methods producing as it did narrower confidence intervals than the Bonferroni, Dunn-Šidák, Hochberg (1974) and Spjøtvoll and Stoline (1973) methods for most imbalanced designs.

Gabriel (1978) reported a Monte Carlo study comparing the techniques of Hochberg (1974) and Spjøtvoll and Stoline (1973) with his own method (Gabriel 1978) over varying sample sizes. The Gabriel method was found to be conservative except for extreme imbalance of sample sizes. The Hochberg (1974) technique was only slightly more conservative than the Gabriel method, whereas the technique of Spjøtvoll and Stoline was much more conservative unless the imbalance was quite small. Though all three methods were conservative when k was large, the pattern of their dependence on imbalance was not the same. Gabriel's method became less conservative and even liberal as the imbalance increased, while Hochberg's (1974) method was unaffected by the degree of sample size imbalance. The technique of Spjøtvoll and Stoline was least conservative for small imbalance but became highly conservative as imbalance increased. The pattern of experiment-wise errors described above was supported by the studies of Stoline and Ury (1979), Stoline (1978) and Dunnett (1980a).

It should be noted that the Gabriel method can produce shorter confidence intervals than those of Hochberg (1974), especially in those cases where α is small and a large sample size imbalance exists (Gabriel 1978). This is the only situation in which the Gabriel method seriously challenges the Kramer method. As pointed out by Stoline (1981), this is also the situation in which the Gabriel method is liberal. Therefore, the Kramer (1956) method is still to be preferred.

The Genizi and Hochberg (1978) method is a technique specifically tailored for the occasion when only two different sample sizes exist in a one way layout design. Stoline (1981) has shown that the Spjøtvoll and Stoline (1973) method is a special case of the Genizi and Hochberg technique. Genizi and Hochberg (1978) cite evidence showing that their procedure yields confidence intervals not longer than Spjøtvoll and Stoline for all pair-wise comparisons, with some confidence intervals being strictly shorter. Stoline (1981) and Dunnett (1980a) report strong evidence supporting the superiority of the Kramer method over the Genizi and Hochberg technique with respect to confidence interval length and experiment-wise error probabilities.

Keselman and Rogan (1978) studied the experiment-wise error rates of the methods of Scheffé (1953), Spjøtvoll and Stoline (1973), and Kramer (1956). In the homogeneous variance condition, all methods kept their error rates at or below the nominal level. The Kramer technique deviated the least from the nominal α while Scheffé's method had the smallest error rates for small and medium sample size disparity. In the case of large sample size imbalance, the method of Spjøtvoll and Stoline was the most conservative.

Ury (1976) and Ury and Wiggins (1975) compared the methods of Scheffé (1953), Dunn-Šidák, Hochberg (1974), and Spjøtvoll and Stoline (1973). Both studies found the Spjøtvoll and Stoline technique to be preferable when the sample size disparity was small, but as the imbalance increased, the methods of Dunn-Šidák, Hochberg (1974) or Scheffé (1953) were preferable. Though the Kramer method was not examined, previous studies had shown its superiority over the Dunn-Šidák, Hochberg (1974) and Spjøtvoll and Stoline (1973) methods.

In conclusion, the procedure preferred by most researchers with regard for conservativeness and optimality, was the method of Kramer (1956). Experiment-wise errors for Kramer's method are below but close to nominal levels, and the method results in the narrowest confidence intervals. The second most preferred procedure would seem to be that of Hochberg (1974), not because its confidence interval lengths or degree of conservativeess are better than other procedures, but because it is more consistent across a wide disparity of sample sizes. The techniques of Gabriel (1978) and Spjøtvoll and Stoline (1973) would be next in preference, though their type I error probabilities change with the degree of sample size imbalance. Though the Genizi and Hochberg (1978) technique competes favourably with the others, it is limited to the condition when only two sample sizes exist. The other procedures, though on occasion preferable, do not control type I error probabilities or confidence interval lengths to the same extent as the other methods do in the case of pair-wise contrasts with unequal sample sizes and homogeneous variances.

RESEARCH DEALING WITH THE CASE OF UNEQUAL VARIANCES

When the variances of the populations which were sampled are not equal, a pooled estimate of the common variance (as used in the methods of the previous section), is inappropriate. The techniques considered in this section use a variance estimate formed from the individual sample variances of the form:

$$s_i^2(n_i)^{-1}+s_j^2(n_j)^{-1} \qquad (2.26)$$

The methods to be considered in this section stem from the research of Hochberg (1976), Games and Howell (1976), Dunnett (1980b) and Tamhane (1977, 1979).

Knowledge that the means being compared come from two different populations with different standard deviations creates a problem regarding how to combine the variances. The Behrens-Fisher solution (Behrens 1964) was one possible solution which has been incorporated into some of the techniques to be discussed.

Howell and Games (1974) compared the robustness of three multiple comparison procedures (multiple t tests, Tukey (1953), Scheffé (1953)) to the violation of the homogeneous variance assumption. Three different standard error estimates were investigated, namely: (a) the conventional test using $\sqrt{\frac{2msw}{n}}$ (where msw = mean square within); (b) the standard error of the traditional t test; and (c) the standard error of the Behrens-Fisher t' statistic. The procedures were highly robust to variance heterogeneity if the Behrens-Fisher standard error estimate was employed. The procedures incorporating the mean square within error estimate were considerably less robust, and produced major distortions in many of the individual contrast power curves. This consideration was not altered by the increase in common sample size. Games and Howell (1976) have demonstrated the inappropriateness of a pooled estimate of the common variance when the assumption of homogeneity of variance is violated. The use of a non-pooled error term provided increased control over experiment-wise error probabilities for all contrasts as well as producing an acceptable power curve.

The majority of research dealing with the case of heterogeneity of variance has also considered the case of equal and unequal sample sizes. Three conditions may be specified as a result: (a) the case of equal sample sizes and heterogeneous variances; (b) the case of unequal sample sizes and heterogeneous variances where small sample sizes are paired with small variances (direct pairing); and (c) the case of unequal sample sizes and heterogeneous variances where small samples sizes are paired with large variances (inverse pairing). Each of the three situations involving equal sample sizes, direct pairing and inverse pairing must be considered separately, since they create different results with the various multiple comparison procedures.

After examining the performance of the F test under conditions of direct and inverse pairing, Box (1954) suggested that a major determinant of the degree of bias was the coefficient of variation of variances (C).

$$C = \frac{\left[\sum_{i=j}^{k}\left(\sigma_i^2-\bar{\sigma}^2\right)^2/k\right]^{1/2}}{\bar{\sigma}^2} \qquad (2.27)$$

The coefficient of variation of variances has been suggested as a criterion for comparison across conditions of different degrees of variance heterogeneity (Howell and Games 1973, Rogan, Keselman and Breen 1977, Keselman and Rogan 1978). Tamhane (1979) also utilized the coefficient of variation of variances in his study, but used the ratio of variance and sample size instead of the variance alone.

$$\mathrm{Var}(X_i) = \sigma^{2'} = \frac{\sigma_i^2}{n_i} \tag{2.28}$$

$$V' = \frac{\left[\sum_{i=1}^{k} (\sigma^{2'} - \bar{\bar{\sigma}}^{2'})^2 / k\right]^{1/2}}{\bar{\bar{\sigma}}^{2'}} \tag{2.29}$$

Box (1954) demonstrated that the F test becomes conservatively biased under the condition of direct pairing, while under inverse pairing, the F test exhibits a liberal bias. Similar results were recorded by Petrinovich and Hardyck (1969) who studied the type I and II error rates of seven pair-wise comparison methods (including Scheffé, Tukey, Duncan, Newman-Keuls and multiple t tests). The researchers concluded that, inverse pairing also caused increases in the type I error probabilities. Petrinovich and Hardyck went on to state that such conditions as unequal sample size, unequal variances and non-normal populations make little difference to the Tukey or Scheffé tests, except as these conditions also affect the obtained F values.

The results of Petrinovich and Hardyck (1969) have been questioned on two counts. Keselman and Toothaker (1973) criticize the manner in which the data were collected, stating that the counting procedure used masked the correspondence between the maximum contrast and the analysis of variance F test. Keselman and Toothaker (1973) conducted a similar monte carlo study and concluded that following a significant F test, the Tukey test was as sensitive for detecting the maximum difference in the set of k means as the Scheffé method. Games (1971b) criticized the results generated by the inverse pairing condition as being an artifact of their method of increasing variances. The multiplication of the initial variance of the comparison populations effectively decreased the distance between the means, relative to the mean square error.

Howell and Games (1973) concluded that the Tukey test (Winer 1962) and the t test show the same response when the homogeneous variance assumption is violated. Both methods were robust when the sample sizes were equal but exhibited either a conservative or liberal bias when heterogeneous variances were paired with unequal samples sizes.

Keselman and Toothaker (1974) reported that the robustness of the Tukey (Winer 1962) and Scheffé statistics was related not just to the type of assumption violation, but to the number of contrasts being considered; this result was later supported by Ringland (1983). The results indicate that the experiment-wise error rates increased within the unequal variance condition versus the equal variance condition and were larger for inverse pairing than for direct pairing. Keselman and Toothaker (1974) concluded that the Tukey statistic could be judged as robust a statistic as that of Scheffé.

Keselman (1976) studied the effect of various combinations of unequal sample size, unequal variance, and non-normality on the probability of the type II error for the Kramer (1956) modification of the Tukey test. After examining a wider range of conditions than Petrinovich and Hardyck (1969) and Keselman, Toothaker and Shooter (1975), Keselman concluded that the power of the Kramer test was not appreciably affected by most of the treatments.

These studies invariably confronted unequal sample size conditions when comparing the Tukey test under violations of the variance homogeneity assumption. Since the original Tukey test is not applicable to this situation, researchers were forced to use a modification, and most often this consisted of one of the Kramer (1956), Winer (1962) or Miller (1966) techniques.

Keselman, Toothaker and Shooter (1975) examined the Kramer (1956) and Winer (1962) unequal n forms of the Tukey test. Type I error and correct decision rates were evaluated under the restriction that the analysis of variance F test be significant at $\alpha = .05$. The data indicated that both methods suffered when unequal variances were paired with unequal sample sizes, and yet both techniques had the same sensitivity for detecting real mean differences. The inverse pairing condition caused the observed experiment-wise error probabilities to exceed the nominal α level for both the Winer and Kramer forms of the Tukey test.

The techniques of Kramer (1956), Miller (1966) and Winer (1962) were compared in a study reported by Rogan, Keselman and Breen (1977). The results supported the Kramer method, as it consistently resulted in empirical experiment-wise error rates deviating less from the nominal significance level than the other two methods. In the equal sample size condition, the experiment-wise error was highly inflated when the variances were highly divergent; this would support the promotion of a non-pooled estimate of the error variance. The results agree with those of Petrinovich and Hardyck (1969), Howell and Games (1973), and Keselman and Toothaker (1974), in that the harmonic mean method provided conservative experiment-wise error rates in the direct pairing condition and liberal rates in the inverse pairing condition. A similar pattern was exhibited by the Kramer method and was supported by Keselman and Rogan (1978).

Keselman and Rogan (1978) and Keselman, Games and Rogan (1979) performed a study comparing the following techniques: Kramer (1956), Spjøtvoll and Stoline (1973), Scheffé (1953), Hochberg (1976), Games and Howell (1976), Gabriel (1978), and Hochberg (1974). When unequal sample sizes were directly paired with unequal variances, all tests except the Games and Howell procedure produced experiment-wise error rates that were less than the nominal level. The Kramer, and Games and Howell procedures generated experiment-wise errors closest to the nominal level, though only the Games and Howell procedure was within one unit of sampling variability of the nominal level. The Scheffé procedure produced the most conservative experiment-wise error estimates when the sample size disparity was small, and when variance heterogeneity was small or medium. As the variances became more divergent, Hochberg's (1976) technique was the most conservative in the small sample size disparity condition. When the sample size disparity was medium or large, the techniques of Scheffé, and Spjøtvoll and Stoline were comparable and provided more conservative experiment-wise error estimates than Hochberg (1976).

In the inverse pairing condition, the Kramer method produced liberal experiment-wise error rates which became more liberal as the sample size and variance heterogeneity disparity increased. The Scheffé procedure was able to maintain its error rates below nominal levels when the degree of variance divergence was not large and when the sample size imbalance was small. With these exceptions, the technique became very liberal. The smallest empirical error rates were produced by the method of Hochberg (1976), which kept its error rates low through all conditions but became more conservative as the variance heterogeneity increased. Of all the procedures examined, Games and Howell (1976) was the only one not affected by inverse pairing of variances and sample sizes and produced error rates that were closest to the nominal level, a result supported by Games and Howell (1976). A comparison of the power rates of the Hochberg (1976) and Games and Howell (1976) procedures supported the latter as the more powerful. In the inverse pairing condition, Kramer's method became very biased, demonstrating the inappropriateness of a pooled estimate of common variance when the assumption of variance homogeneity is violated.

Tamhane (1979) reported a study comparing the techniques of Spjøtvoll and Stoline (1973), Hochberg (1974, 1976), Tamhane (1977) and Games and Howell (1976). The techniques of Spjøtvoll and Stoline, and Hochberg (1974) were found to be fairly robust unless the divergence in the variances was extreme. Of the procedures created for the unequal variance situation, the Games and Howell procedure tended to be liberal without any pattern, but it also produced the shortest confidence interval lengths of all procedures examined. The Tamhane (1977) procedure was conservative and supplied confidence interval widths only slightly larger tha the Games and Howell procedure.

It would appear that Tamhane (1979) narrowed the choice of procedures down to Games and Howell (1976) and Tamhane (1977). Dunnett (1980b) examined these two procedures, along with two others which he devised--T3, an extension of Tamhane's (1977) T2 procedure, and the C procedure, an extension of a method proposed by Cochran (1964).

Dunnett (1980b) examined only the direct pairing condition, with large variances being paired with large sample sizes. These conditions were examined over degrees of freedom ranging up to infinity (i.e., the known variance case). The T2 and T3 methods were both conservative and became identical when the variances were known. In the case of finite degrees of freedom case, the T3 procedure was less conservative than the T2 procedure and generated shorter confidence interval widths. This would be anticipated given the tighter inequality (Dunnett 1980b). The empirical experiment-wise error probabilities for the Games and Howell procedure decreased as the variance became more divergent. In the case of the equal variance case (or those close to it), the error rate appeared to increase slightly as the degrees of freedom became smaller.

In practical situations, most researchers would choose n_i's so that σ_i^2/n_i would be approximately the same in each sample. As a result, variation in variance from the equal variance case would tend to be small. This is the situation wherein the Games and Howell procedure becomes liberal. As the variances become more divergent, the empirical experiment-wise error estimates become more conservative.

The C procedure and the Games and Howell procedure are seen to be identical in the known variance case. For the finite variance case, the C method is more conservative than Games and

Howell. This is not a surprising development given that the degrees of freedom for the Games and Howell method is between ν_i and ν_j, whereas the degrees of freedom for the C procedure may be as high as $\nu_i + \nu_j$. The C method also provides narrower confidence intervals than T3 for large degrees of freedom.

Dunnett (1980b) reported that the T3 procedure would be preferable when the degrees of freedom are small. In these instances, the Games and Howell method is slightly liberal and the C procedure much more conservative than the T3 method. For large degrees of freedom, it would seem that the Games and Howell procedure should be preferred, since its error rates are the closest to the nominal level without becoming liberal. In the extreme variance heterogeneity situation, the Games and Howell procedure would also be preferable regardless of degrees of freedom.

In summary, it would seem that of the techniques designed for the unequal variance case, the method of Games and Howell (1976) provides empirical experiment-wise error rates which are closer to the nominal levels and the shortest confidence interval widths. In situations where sample size disparity is small and the variance heterogeneity is also small, this method may become liberal, in which case the T3 procedure is to be preferred. Hochberg's (1976) procedure would seemto be the next preferred method. Unfortunately, Dunnett (1980b) did not consider this method in his comparison with the C and T3 procedures. Note also that Dunnett (1980b) did not consider the inverse pairing condition of variances and sample sizes for it is this condition which seems to cause major problems in using the various multiple comparison procedures.

COMPLEXITY

The multiple comparison procedures presented thus far have been compared with respect to their experiment-wise type I error rates and the lengths of their confidence intervals. Though the performance of the multiple comparison procedures in terms of these statistical criteria is important, it may not be of primary interest to the researcher. A principle concern of the researcher is the interpretation of the results in the sense of portrayal of a clear picture of the relationships amongst the sample means. A criterion proposed by Shaffer (1981), called complexity, deals with this consideration.

To simplify the discussion, the following notation will be used. Consider a four group one-way analysis of variance design. Let each sample mean be represented by the appropriate numeral 1, 2, 3 or 4. Let each pair-wise contrast be represented by the pair of numerals corresponding to the populations involved in the contrast. Thus, the contrast $\overline{X}_1 - \overline{X}_2$ can be represented by (12). In the four group design, there are six possible pair-wise contrasts:

$$(12) \quad (13) \quad (14) \quad (23) \quad (24) \quad (34) \qquad (2.30)$$

Finally, if a pair-wise contrast has reached the level of significance, this will be represented by an asterisk following the appropriate pair-wise notation. For example, in the following display

$$(12)^* \quad (13) \quad (14)^* \quad (23) \quad (24) \quad (34) \qquad (2.31)$$

the estimated contrasts $(\overline{X}_1 - \overline{X}_2)$ and $(\overline{X}_1 - \overline{X}_4)$ were found to be significant while the other four estimated contrasts were found to be not significant.

Consider the six pair-wise contrasts as forming a set of six elements; from this set, subsets ranging in size from zero elements to six elements may be chosen. The number of combinations for the various sizes of subsets can be calculated as follows:

$$\binom{6}{0}, \quad \binom{6}{1}, \quad \binom{6}{2}, \quad \binom{6}{3}, \quad \binom{6}{4}, \quad \binom{6}{5}, \quad \binom{6}{6}. \qquad (2.32)$$

These calculations yield a total of 64 possible combinations of the six elements for subsets of all sizes.

Consider that the specific subset selected is in fact the set of pair-wise contrasts which has reached the level of significance. It can now be seen that there are 64 possible combinations of estimated pair-wise contrasts reaching the level of significance, with the number of significant contrasts ranging from zero to six.

If a specific subset of significant estimated contrasts is illustrated in the above notation, then the resulting display could be seen as a comparison of the estimated contrasts that are significant and those that are not. For example, consider the following display:

(12)* (13) (14)* (23) (24)* (34) (2.33)

In this case, the estimated pair-wise contrasts (12), (14), and (24) are significant, while the remaining three estimated contrasts are not. This display of significant versus non-significant contrasts has been defined by Shaffer (1981) as a pattern. [illegible]er again the four populations represented by the numerals 1, 2, 3 or 4. A partition of the populations is defined as the subdivision of the set of four populations into mutually-exclusive subsets based upon significant differences of some sample statistic (i.e., the sample mean). In this case, there are five different forms that the partition may take, corresponding to one subset, two subsets, three subsets, or four subsets. The partitions are indicated as follows:

[1, 2, 3, 4] (2.34)
[1] [2, 3, 4]
[1, 2] [3, 4]
[1] [2] [3, 4]
[1] [2] [3] [4]

The above partitions indicate only the form of the partition. For example, [1] [2, 3, 4] and [2] [1, 3, 4] are the same type of partition, though the arrangement of the sample means is different.

A partition, as described above, and inferred from the samples, is the ideal situation that a researcher would desire. As a result of the division of the samples into mutually-exclusive subsets, interpretation based upon the relationships among sample means is relatively straightforward.

There is obviously a specific relationship between the partition and the pattern. When a pattern creates one of the above partitions, it is defined as a simple pattern. For example, the pattern of

$$(12)^* \quad (13)^* \quad (14)^* \quad (23)^* \quad (24)^* \quad (34) \tag{2.35}$$

yields the partition of

$$[1] \quad [2] \quad [3, 4] \tag{2.36}$$

Only certain patterns can yield unequivocal partitions of the sample means. By far, the majority of possible patterns do not yield partitions of the sample means, thus increasing the difficulty of interpretation. Consider the following pattern:

$$(12) \quad (13)^* \quad (14)^* \quad (23) \quad (24)^* \quad (34) \tag{2.37}$$

$\overline{X}_1$ is significantly different from $\overline{X}_3$ and $\overline{X}_4$ and $\overline{X}_2$ is significantly different from $\overline{X}_4$. Yet, $\overline{X}_2$ was found not to be significantly different from $\overline{X}_1$ or $\overline{X}_3$. The apparent inconsistency of $\overline{X}_2$ prohibits a partition of the sample means. If (23) was also declared significant, then a partition of the form of [1, 2] [3, 4] would occur.

A simple pattern can therefore be defined as a pattern of significant and non-significant estimated contrasts which yield a partition of the sample means into mutually-exclusive subsets. The sample means within each subset are not significantly different from each other. Sample means from different subsets result in estimated contrasts that are significant.

Shaffer (1981) uses the simple pattern as a reference point for measuring complexity. Because the simple pattern yields a partition of the sample means, it is given a complexity of zero. All other patterns are given a complexity relative to the distance from a simple pattern. Complexity of a pattern is equal to the minimum number of non-significant contrasts which must be declared significant in order to transform the original pattern into a simple pattern. The pattern cited above, (2.37), has a complexity of one. The pattern cited in (2.33) has a complexity of two since (23) and (13) must be declared significant in order to yield a simple pattern.

As a result, each of the 64 possible combinations of pair-wise contrasts has a complexity indicative of the distance from a simple pattern. These 64 patterns may be further categorized into pattern types. Pattern types have two characteristics. First, they must have the same number of significant contrasts and they must have the same complexity.

The 64 possible patterns (from the 4 sample case) may be reduced down to eleven pattern types, five of which are simple patterns. The eleven pattern types are listed in Table 3 with their appropriate complexities and the number of null hypotheses that are rejected. The latter is, of course, equivalent to the number of pair-wise contrasts that are significant.

Table 3 illustrates an important trend: complexity decreases as the number of rejections of null hypotheses increases. This is understandable, since the chance of achieving a partition of the sample means is greater with more significant contrasts. Three of the simple patterns illustrated in Table 3 involve four or more significant contrasts.

Since the population means must fall into one of the partitions, complexity offers a method of comparing patterns of different multiple comparison procedures with respect to their distance from a simple pattern. Thus, a preferred procedure would be characterized by a relatively low complexity and a high number of rejections of the null hypotheses.

It is interesting to note that complexity is defined as the number of non-significant estimated contrasts that must be made significant rather than those that are significant being made non-significant. In some cases, the distance from a simple

TABLE 3

PATTERN TYPES FOR THE 4 SAMPLE CASE

Pattern Type	Significant Pair-wise Contrasts						Complexity	Number of Rejections
1	No pair-wise contrasts significant						0	0
2	Any single pair-wise contrast						2	1
3	(12)*	(13)	(14)	(23)	(24)	(34)*	2	2
	(12)	(13)*	(14)	(23)	(24)*	(34)		
	(12)	(13)	(14)*	(23)*	(24)	(34)		
4	Any other two pair-wise contrasts						1	2
5	(12)*	(13)*	(14)*	(23)	(24)	(34)	0	3
	(12)*	(13)	(14)	(23)*	(24)*	(34)		
	(12)	(13)*	(14)	(23)*	(24)	(34)*		
	(12)	(13)	(14)*	(23)	(24)*	(34)*		
6	(12)*	(13)*	(14)	(23)*	(24)	(34)	2	3
	(12)	(13)*	(14)*	(23)	(24)	(34)*		
	(12)	(13)	(14)	(23)*	(24)*	(34)*		
	(12)*	(13)	(14)*	(23)	(24)*	(34)		
7	Any other three pair-wise contrasts						1	3
8	(12)	(13)*	(14)*	(23)*	(24)*	(34)	0	4
	(12)*	(13)	(14)*	(23)*	(24)	(34)*		
	(12)*	(13)*	(14)	(23)	(24)*	(34)*		
9	Any other four pair-wise contrasts						1	4
10	Any five pair-wise contrasts						0	5
11	All pair-wise contrasts significant						0	6

* significant contrast

pattern would be shorter if significant estimated contrasts were declared non-significant as opposed to non-significant estimated contrasts becoming significant.

Shaffer (1981) does not give any reason as to why complexity was defined in this manner. The implication is made, however, that the situation of estimated contrasts being falsely declared significant is less prevalent. As a result, complexity was measured on the assumption that some estimated contrasts were mistakenly found to be non-significant.

It should be noted that Shaffer's (1981) article deals with a more general formulation of complexity. In this case, the concept of complexity has been refined to deal only with pair-wise contrasts of means. Originally, Shaffer's formulation dealt with more complex contrasts. In this manner, a pattern was seen as a series of tests beginning with the overall test of all experimental groups and continuing with progressively smaller subsets of all possible sizes and combinations. Within this framework, Shaffer was able to compare the performance of range, F and gap procedures. Since the present study deals solely with range statistics, only certain contrasts may be studied. For the sake of simplicity, Shaffer's theory was refined to consider pair-wise comparisons. This modification did not involve a change in the complexities of the different patterns, though it did change how they are displayed. The number of rejections of null hypotheses consisted only of pair-wise rejections and not those of other subsets. Since this created a uniform decrease across all patterns, the relative difference among patterns did not change.

Shaffer (1981) performed a Monte Carlo study to examine the usefulness of complexity in distinguishing among three different statistical procedures: range, gap and F statistic. Shaffer employed a three group (three means) design that could be represented by the following five pattern types (and their respective complexities):

1. no subsets rejected (complexity = 0)
2. 123 rejected (complexity = 2)
3. 123 and 1 pair rejected (complexity = 1)
4. 123 and 2 pairs rejected (complexity = 0)
5. all subsets rejected (complexity = 0)

As was mentioned previously, Shaffer considered more subsets than those involving just pair-wise contrasts.

The two different partitions examined by Shaffer were [1,2] [3] and [1] [2] [3]. In the first partition, two means were equal and different from the third and, in the second partition, the three means were equally spaced. The constants ranged in value from .2 to 6 for the first partition and .2 to 7.2 for the second.

Shaffer calculated the average number of rejections and the average complexity as two dependent variables. The average number of rejections is the average of the number of rejections of the null hypothesis within each pattern type weighted by the pattern frequency. Average complexity was calculated in a similar manner by replacing the number of rejections within each pattern type by the respective pattern complexity.

Results from Shaffer's simulations indicated that as the difference between group means increased, the average number of rejections increased. It was also apparent that average complexity for the range and F procedures was higher than the gap procedure across all the conditions cited. It was interesting to note that when mean variability was partitioned into low, middle and upper thirds, average complexity had higher values in the middle third. This was due to the relatively large frequencies of pattern type 3. Shaffer was able to conclude that average complexity was able to discriminate among the range, gap and F procedures.

Several limitations of Shaffer's study should be considered. Since average complexity is based upon pattern complexities which have discrete values and a very limited range (ie., 0, 1, 2), the calculation of an "average" is questionable. This is further complicated by the fact that pattern type 2 does not occur under the F and range statistics, and only to a small extent (less than a proportion of .01) in the gap procedure. Consequently, average complexity is based solely on the frequency of pattern type 3 since the other patterns have zero complexity. This limitation exists because of the small number of patterns in a three group design. The range of possible patterns and values of complexity is increased in the four group design (Table 3).

SUMMARY

The survey of literature has described the behaviour of the various forms of the Tukey test. With respect to experiment-wise error rates, it has shown that the unequal sample size and unequal variance forms of the Tukey test behave differently under conditions of sample size disparity, variance heterogeneity and direct versus inverse pairing. This information is valuable for the educational researcher who is constantly confronted with either samples of varying sizes and/or different variances, and who is also concerned about the probability of type I error rates.

Complexity could provide knowledge about the various multiple comparison procedures along a new dimension, hitherto not investigated by statistical researchers. Specifically, complexity compares the various forms of the Tukey test with the interpretability of their results (i.e., how close are the patterns generated by the technique to a simple pattern?).

The purpose of the present study is to examine the behaviour of complexity under the conditions of sample size disparity, variance heterogeneity, mean variability and direct/inverse pairing. These conditions are studied with respect to the unequal and unequal variance forms of the Tukey test. The four group design was chosen because it yielded a larger range of pattern types as well as five different partitions to study. It was decided to use a Monte Carlo study for several reasons.

Since the behaviour of average complexity was to be compared to type I error rates, it was decided to replicate as closely as possible those studies by Keselman and Rogan et al. Though it was possible to predict general increases and decreases in pattern frequency, the behaviour of average complexity across all conditions could not be predicted. The large number of conditions to be studied and the use of a four group design necessitated the aid of a computer simulation technique.

The results of this study should yield a more complete picture of the performance of various forms of the Tukey test, thus enabling the educational researcher to better select the most appropriate method for analysis of his/her data. Inherent in the formation of various patterns, is the examination of the frequency of individual estimated contrasts which are found to be significant. This will provide information as to the behaviour of various contrasts under the conditions being studied.

CHAPTER II

PROCEDURE

Four independent normal populations were generated with mean zero and standard deviation one, using an IMSL program called GGNPM which employs a subroutine based on an algorithm designed by Box and Muller (1958).

From each population, the respective samples were drawn to correspond to one of the sample size disparity conditions of small, medium or large disparity. The small sample size disparity condition had a disparity between the smallest and largest sample sizes of 42.3%; the medium and large disparity conditions had disparities of 100% and 182.3% respectively. The sample sizes in the present study were chosen so that their sizes and disparities were similar to those studied by Keselman and Rogan (1978), the total being kept constant for all three cases. The sample sizes for each population are listed in Table 4.

TABLE 4

SAMPLE SIZES DRAWN UNDER EACH DISPARITY CONDITION

Sample Size Disparity	Population (1)	(2)	(3)	(4)	Total
Small	26	27	34	37	124
Medium	21	26	35	42	124
Large	17	23	36	48	124

Each sample score was multiplied by the appropriate standard deviation to create the various variance heterogeneity conditions. The variance heterogeneity conditions were determined using the coefficient of variation (Box, 1954). The values of 0 (homogeneous), .2 (small heterogeneity), .6 (medium heterogeneity) and 1.0 (large heterogeneity) characterized the four conditions and were similar to those employed by Keselman and Rogan (1978). The respective variances for each population under the four conditions are listed in Table 5.

TABLE 5

VARIANCES USED FOR EACH LEVEL OF COEFFICIENT OF VARIATION

Coefficient of Variation (C)	Population (1)	(2)	(3)	(4)
.0	1.0	1.0	1.0	1.0
.2	.73	.92	1.08	1.27
.6	.30	.80	1.03	2.0
1.0	.01	.06	1.58	2.35

In the direct pairing condition, the smallest sample size was drawn from the population with the smallest variance, the next smallest sample size from the population with the next smallest variance and so on for the four samples. This condition is illustrated in Tables 4 and 5. In the inverse pairing condition, the smallest sample size was drawn from the population with the largest variance, the next smallest sample size from the population with the second largest variance, etc., for the remaining samples. This condition would be illustrated in Table 5 if the largest variances were associated with population 1, the next largest with population 2 and so on for the remaining variances.

A constant was added to each sample score to create the various partitions under study. The constants were determined using a procedure proposed by Cohen (1969) which generates small, medium and large mean variabilities based upon the range from smallest to largest sample mean. The constants used for each partition are listed in Table 6. In order to create the partitions, Δ_1 was added to population 1, Δ_2 to population 2, Δ_3 to the third population and Δ_4 to population 4.

Cohen's (1969) procedure was used to determine the value of the constants for the fourth partition where the four population means are equally distributed over the range. The constants for the other partitions were calculated using the constant from the fourth partition so that the distance between distinguishable populations was kept constant across all partitions.

For each of the four samples under each condition, sample means, sample variances, and pooled variances were calculated.

$$\bar{X}_i = \frac{\Sigma X_i}{n_i}$$

$$s_i^2 = \frac{\sum_{j=1}^{n_i} (X_{ij}-\bar{X}_i)^2}{n_i-1}$$

$$s^2 = \frac{\sum_{i=1}^{4} \sum_{j=1}^{n_i} (X_{ij}-X_{i.})^2}{N-4} \quad \text{where} \quad N = \sum_{i=1}^{4} n_i$$

TABLE 6

MEAN VARIABILITY FOR EACH PARTITION

Partition		Mean Variability	
	Small	Medium	Large
I [1] [2,3,4]	$\Delta_1 = 0$ $\Delta_i = .08933$ for $i = 2,3,4$	$\Delta_1 = 0$ $\Delta_i = .22333$ for $i = 2,3,4$	$\Delta_1 = 0$ $\Delta_i = .35733$ for $i = 2,3,4$
II [1,2] [3,4]	$\Delta_1 = \Delta_2 = 0$ $\Delta_i = .08933$ for $i = 3,4$	$\Delta_1 = \Delta_2 = 0$ $\Delta_i = .08933$ for $i = 3,4$	$\Delta_1 = \Delta_2 = 0$ $\Delta_i = .35733$ for $i = 3,4$
III [1][2] [3,4]	$\Delta_1 = 0$ $\Delta_2 = \frac{1}{2}\Delta_3$ $\Delta_i = .17866$ for $i = 3,4$	$\Delta_1 = 0$ $\Delta_2 = \frac{1}{2}\Delta_3$ $\Delta_i = .44666$ for $i = 3,4$	$\Delta_1 = 0$ $\Delta_2 = \frac{1}{2}\Delta_3$ $\Delta_i = .71466$ for $i = 3,4$
IV [1][2][3][4]	$\Delta_1 = 0$ $\Delta_2 = \frac{1}{3}\Delta_4$ $\Delta_3 = \frac{2}{3}\Delta_4$ $\Delta_4 = .268$	$\Delta_1 = 0$ $\Delta_2 = \frac{1}{3}\Delta_4$ $\Delta_3 = \frac{2}{3}\Delta_4$ $\Delta_4 = .67$	$\Delta_1 = 0$ $\Delta_2 = \frac{1}{3}\Delta_4$ $\Delta_3 = \frac{2}{3}\Delta_4$ $\Delta_4 = 1.072$
V [1, 2, 3, 4]	$\Delta_i = 0$ for $i = 1,2,3,4$	$\Delta_i = 0$ for $i = 1,2,3,4$	$\Delta_i = 0$ for $i = 1,2,3,4$

The six possible estimates of the pair-wise contrasts were then calculated:

Estimated contrast (12): $X_1 - X_2$
(13): $X_1 - X_3$
(14): $X_1 - X_4$
(23): $X_2 - X_3$
(24): $X_2 - X_4$
(34): $X_3 - X_4$

Seven statistical methods were compared under each condition. These were Kramer (1956), Spjotvoll and Stoline (1973), Hochberg (1974), Gabriel (1978), Games and Howell (1976), T3 procedure and the C procedure (Dunnett, 1980b). The necessary critical values were extracted directly from the appropriate tables, or were estimated using linear interpolation (Stoline and Ury 1979, Stoline 1978, Harter 1960). An alpha level of .05 was used for all critical values.

Within each condition, 1000 simulations were performed. Each simulation involved the calculation of the seven modified forms of the Tukey test and a recording of which of the six possible contrasts were found to be significant under each method. Three dependent variables were calculated on the basis of this information for each form of the Tukey Test method and under each condition.

Type I experiment-wise error was calculated under all partitions except partition IV which did not permit the occurrence of an experiment-wise error. The frequency of pairwise contrasts rejecting the null hypothesis when it was true (at $\alpha = .05$) was tabulated. The specific contrasts representing an experiment-wise error under each partition are listed in Table 7.

Experiment-wise error was tabulated in such a way that one or more incidences of an error within any given simulation were recorded as only one occurrence. Thus, experiment-wise error was calculated in the following manner:

TABLE 7

TYPE I ERROR CONTRASTS IN EACH PARTITION

Partition				
I	II	III	IV	V
[1] [2,3,4]	[1,2] [3,4]	[1][2] [3,4]	[1][2][3][4]	[1,2,3,4]
(23)	(12)	(34)		(12)
(24)	(34)			(13)
(34)				(14)
				(23)
				(24)
				(34)

$$\text{Experiment-wise error} = \frac{\sum \text{number of simulations for which at least one error occurred}}{\text{number of simulations}}.$$

The frequency of each of the 11 pattern types proposed by Shaffer was also tabulated under each condition. This information and that of the complexities of each pattern (Table 3) were used to calculate the average complexity (Shaffer 1981).

$$\text{Average Complexity} = \frac{\sum \text{(freq. of pattern)(complexity of pattern)}}{\text{number of simulations}}$$

The frequency of each pattern type was also used to calculate the average number of rejections occurring under each condition (see Table 3) (Shaffer 1981).

$$\text{Average number of rejections} = \frac{\sum \text{(freq. of pattern)(no. of rej.)}}{\text{number of simulations}}$$

As well as calculating the above dependent variables, three other sources of information were also tabulated under each condition. The frequency of occurrence of significant contrasts was recorded for each of the six contrasts, as was the frequency of occurrence of each of Shaffer's eleven pattern types. The frequency of occurrence of the 64 possible combinations of the six contrasts was also recorded under the prospective eleven pattern types.

CHAPTER III

RESULTS

The purpose of the result section is to describe the general trends of the data as exhibited by experiment-wise error rates, average number of rejections and average complexity for the methods dealing with sample size imbalance and variance heterogeneity.

The seven modified forms of the Tukey method were divided into two classes. The methods proposed by Kramer (1956), Spjøtvoll and Stoline (1973), Hochberg (1974) and Gabriel (1978) deal with the case of unequal sample sizes. The techniques of Games and Howell (1976) and the T3 and C procedures of Dunnett (1980b) involve the situation of unequal variance as well as unequal sample sizes.

The results of the simulations across all conditions indicated that these two classes of techniques reacted differently. Within each class, however, the individual methods demonstrated similar trends. With respect to experiment-wise error rates, Spjøtvoll and Stoline's (1973) method was slightly more conservative than the others in its class, and the method of Games and Howell (1976) slightly more liberal than the T3 or C procedures. With respect to the average number of rejections, the methods of Kramer (1956) and Games and Howell (1976) exceeded the other techniques in their respective classes. Despite these differences, the trend across various conditions was replicated (though not necessarily to the same degree) by the other techniques within the same category. Subsequently, it was decided to select the Kramer (1956) and the Games and Howell (1976) methods to illustrate the trends exhibited by the two classes of modified Tukey techniques.

The results will be described under four of the independent variables: partition, sample size disparity, variance heterogeneity and mean variability. The results of the direct and inverse pairing conditions will be described for each of the three variables. On each of the following figures (1 through 6), the curves represent the three different levels of sample size disparity, the horizontal axis indicates the levels of variance heterogeneity and the vertical axis measures the magnitude of each of the three dependent variables.

PARTITION

Kramer Method

In the direct pairing condition, the Kramer method experienced approximately the same type I error rates across all partitions. The average number of rejections increased as the number of distinct sets increased across the partitions. Average complexity also appeared to increase in a manner similar to the one observed for average number of rejections.

In the inverse pairing condition, the highest type I error rates occurred in partition V, with partitions I and II exhibiting similar trends but with lower error rates. Partition

III had error rates which decreased to zero. This particular case will be discussed in a later section. The average number of rejections increased with the number of distinct sets as did average complexity, but the differences were less marked and depended to some extent upon the other dependent variables (see Appendix II).

Games and Howell Method

In the direct pairing condition, the type I error rates decreased as the number of distinct sets increased. Likewise, the average number of rejections increased. Average complexity also evidenced a slight increase but not to the same extent as average number of rejections.

In the inverse pairing condition, type I error rates did not show any great differences across the various partitions. Average number of rejections and average complexity both increased with an increase in the number of distinct sets.

MEAN VARIABILITY

Kramer Method

For partition III, direct pairing, experiment-wise error rates for the Kramer method under the small, medium and large mean variability conditions are shown in Fig. 1. Experiment-wise error rates did not evidence great changes as mean variability increased in the direct pairing condition. Similar results were also found in partitions I and II. The average number of rejections of the null hypotheses and average complexity increased in value as mean variability increased. (Fig. 2, Fig. 3).

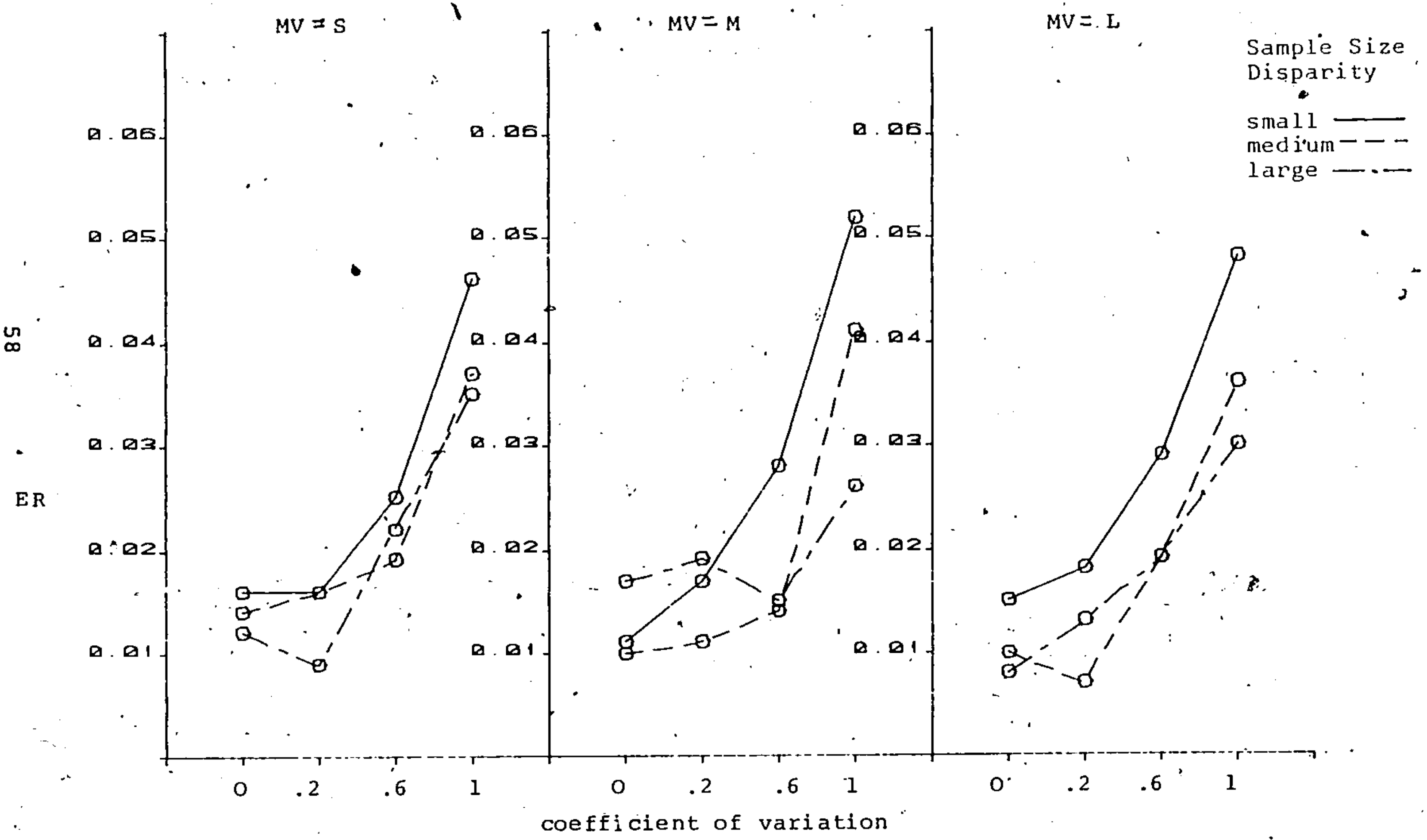
In the inverse pairing condition, experiment-wise error rate was again found to exhibit little change as mean variability increased (Fig. 4). The average number of rejections and average complexity increased with increased mean variability (Fig. 5, Fig. 6). In general, the average number of rejections reached higher values in the inverse pairing condition than in the direct pairing condition when the comparison was made within respective partitions (see Appendix II). (Note that Fig. 1, 2, 3 come from partition III and Fig. 4, 5, 6 from partition II, and thus should not be compared.) With respect to average complexity, similar ranges of values were found in both direct and inverse pairing.

Games and Howell Method

The results for the Games and Howell method exhibited similar trends to those of the Kramer method. In the direct pairing condition, there was little change in experiment-wise error rates as mean variability increased (Fig. 7). The average number of rejections and average complexity increased in value as mean variability increased (Fig. 8, Fig. 9).

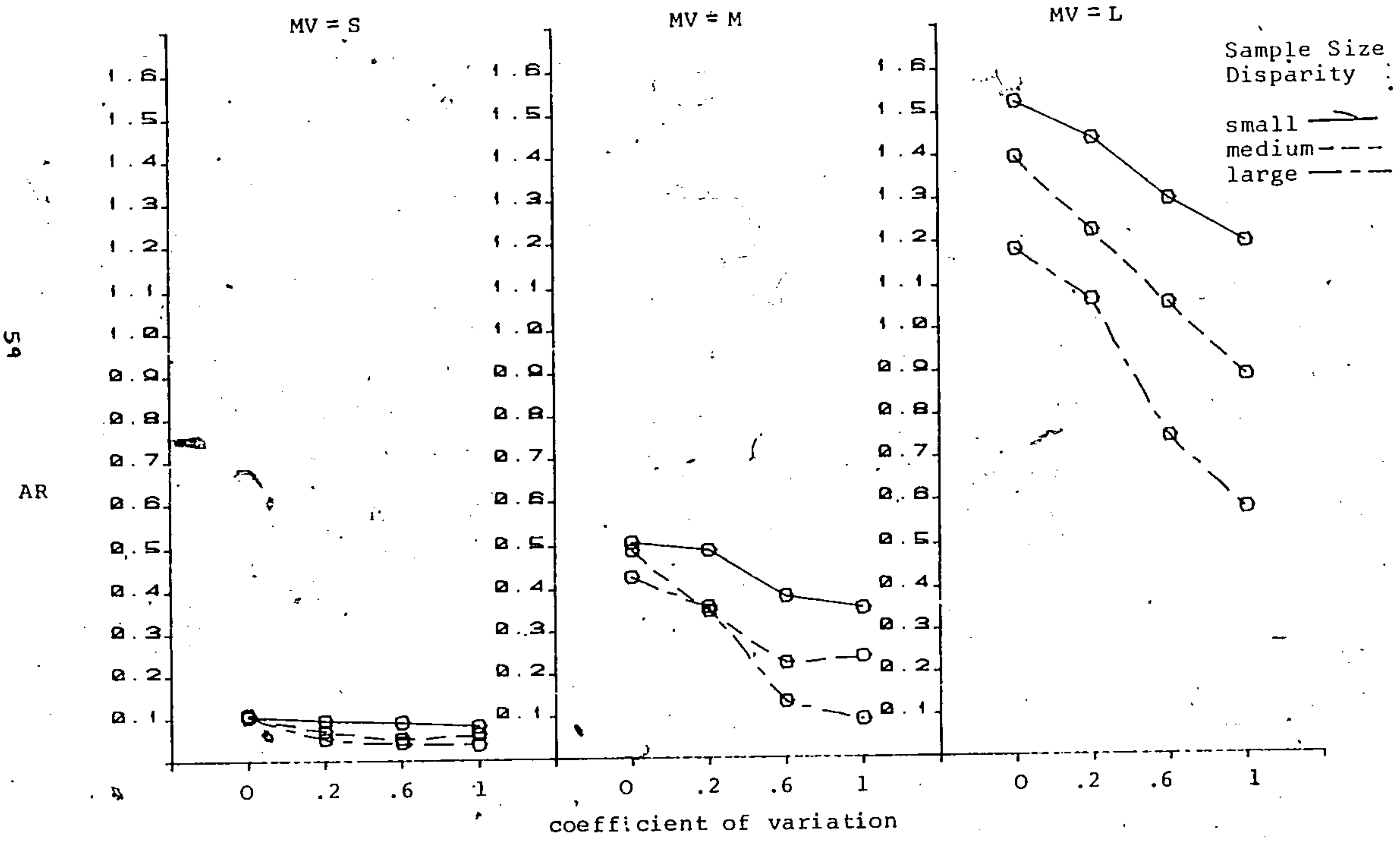
In the inverse pairing condition, experiment-wise error rates again showed little change with increased mean variability (Fig. 10), whereas both average number of rejections and average complexity increased (Fig. 11, Fig. 12). In general, a higher

Figure 1: Type I error rates, Kramer method, partition III, direct pairing



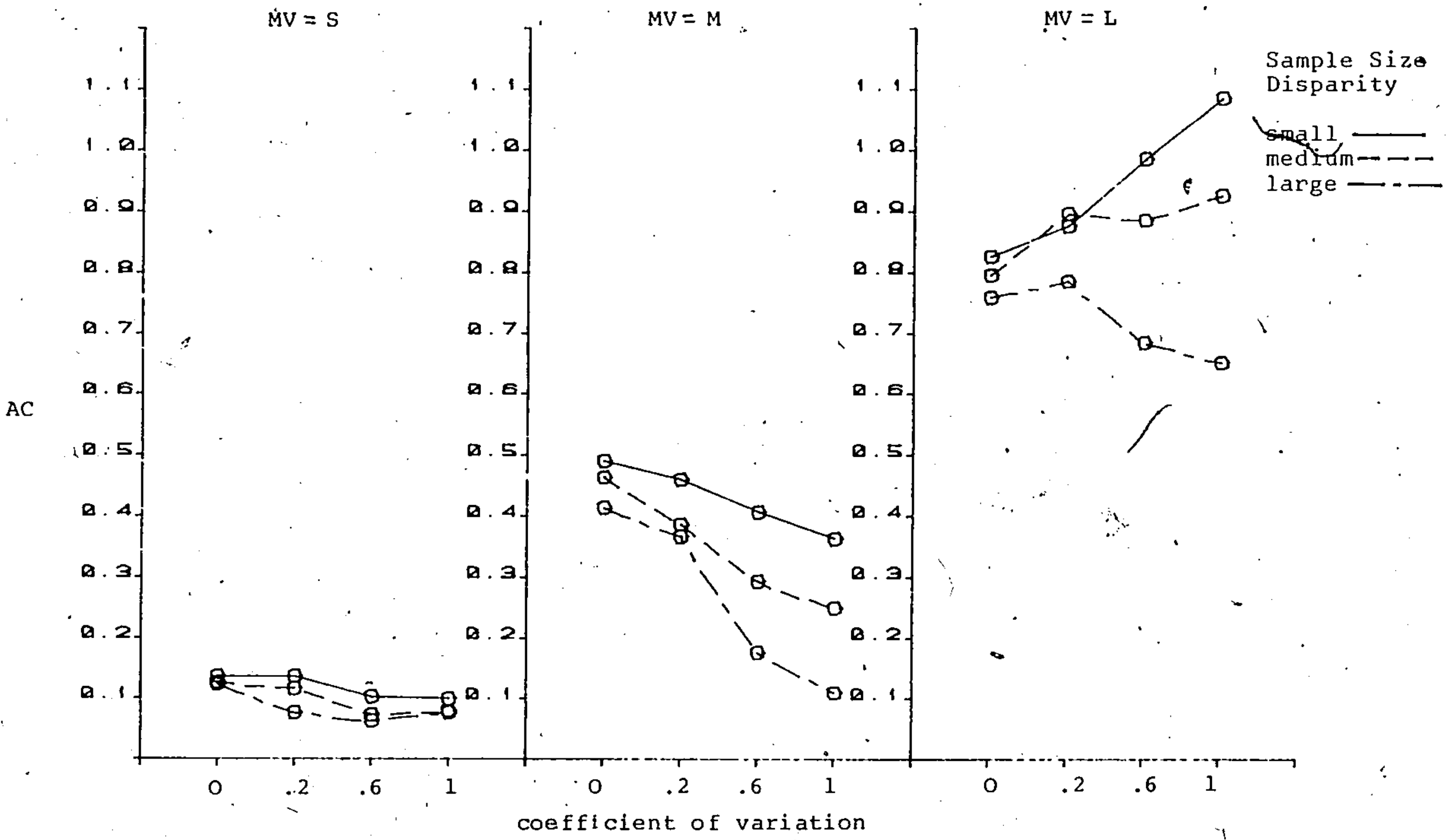
MV-mean variability, ER-experiment-wise error rate, S-small, M-medium, L-large

Figure 2: Average number of rejections, Kramer method, partition III, direct pairing



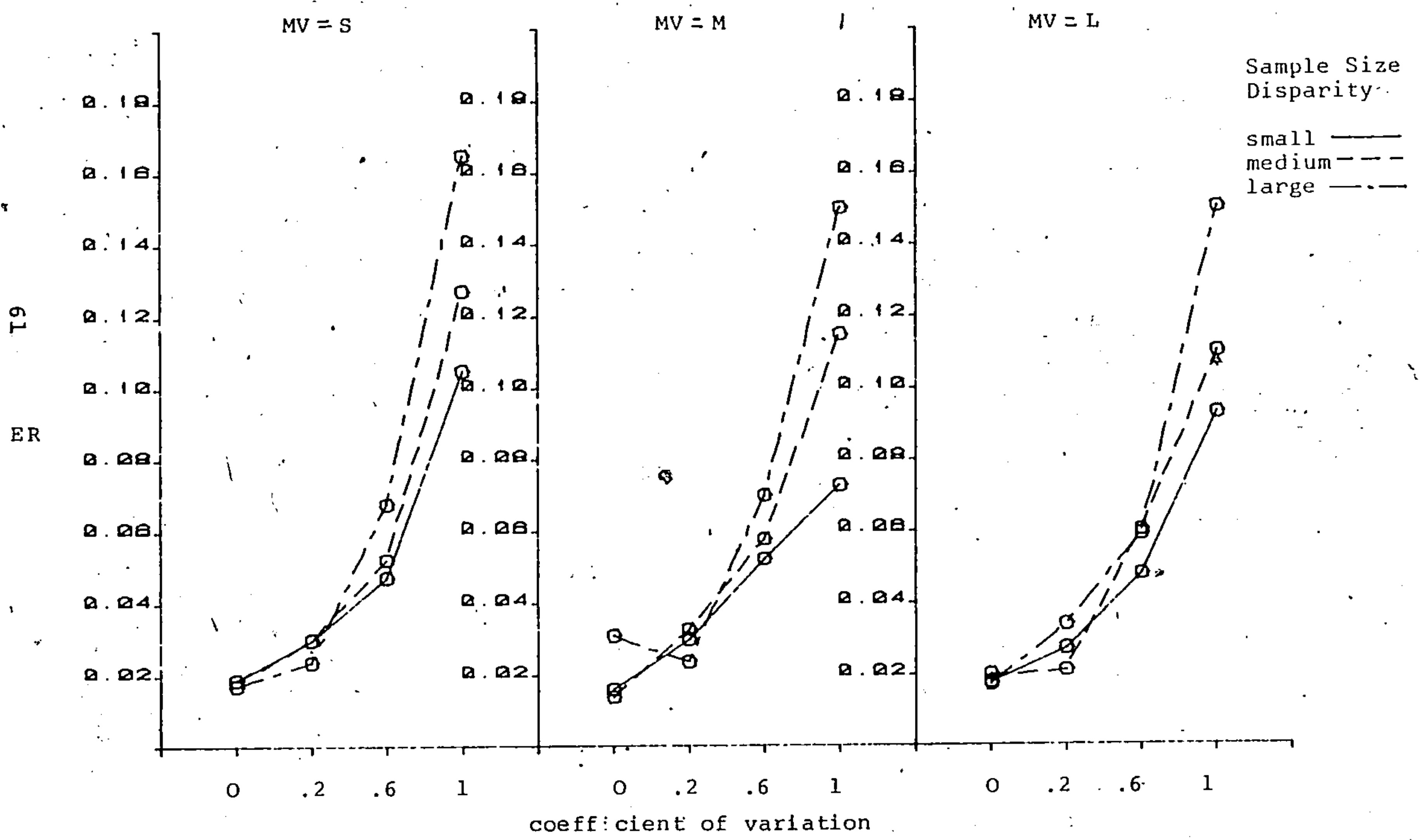
MV-mean variability, AR-average number of rejections, S-small, M-medium, L-large

Figure 3: Average complexity, Kramer method, partition III, direct pairing



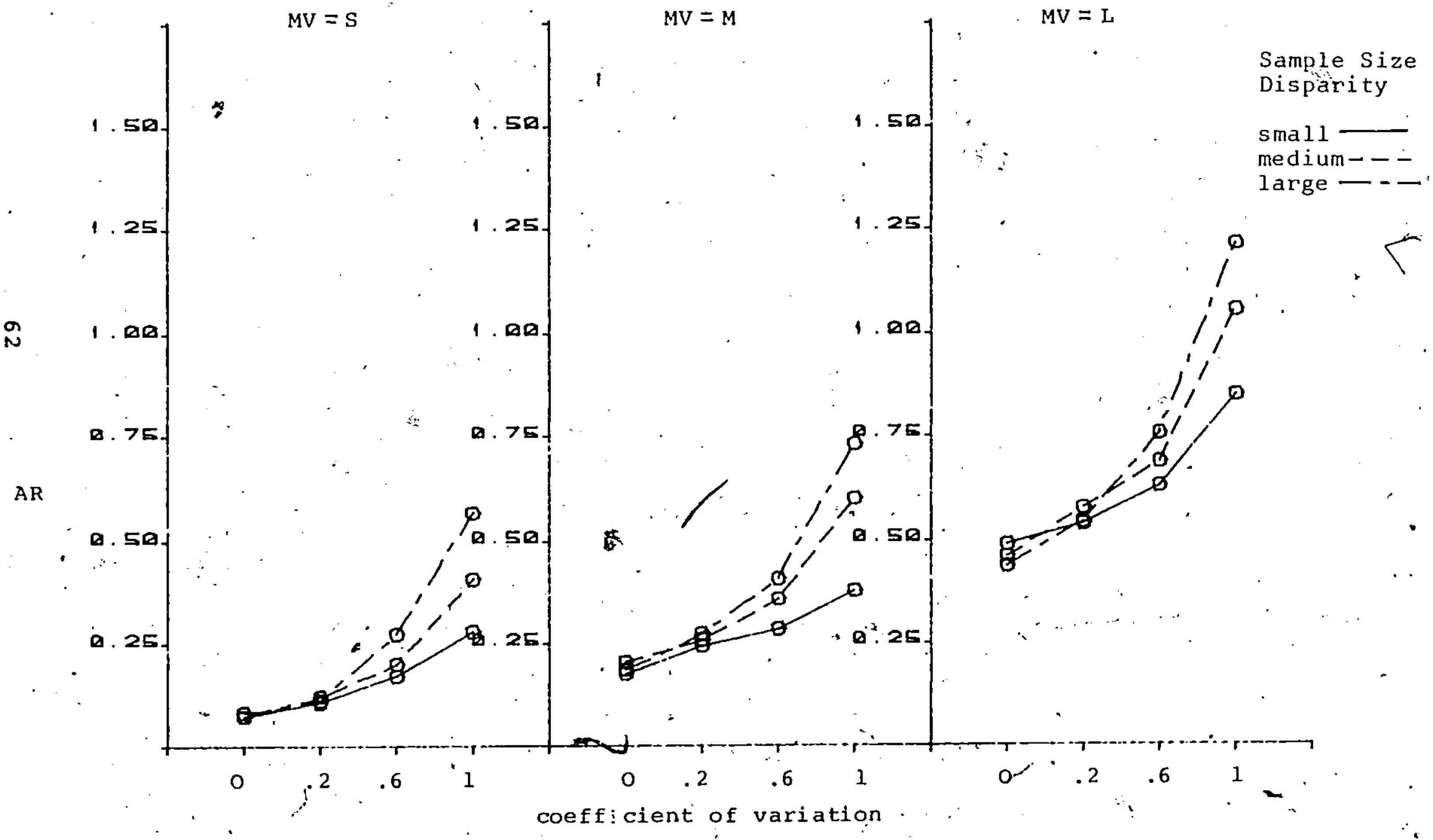
MV-mean variability, AC-average complexity, S -small, M-medium, L-large

Figure 4: Type I error rates, Kramer method, partition II, inverse pairing



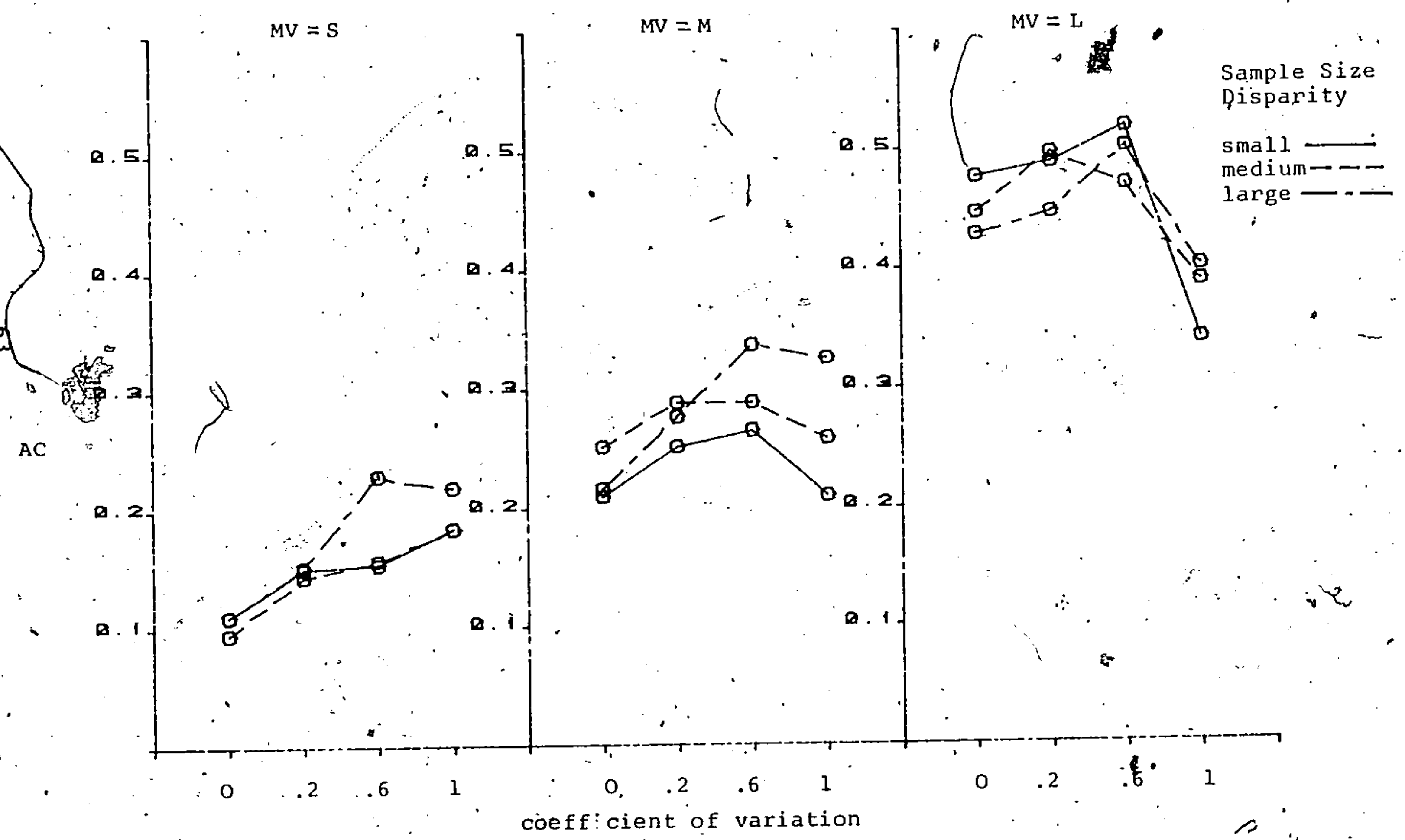
MV-mean variability, ER-experiment-wise error rate, S-small, M-medium
L-large

Figure 5: Average number of rejections, Kramer method, partition II, inverse·pairing



MV-mean variability, AR-average number of rejections, S-small, M-medium, L-large

Figure 6: Average complexity, Kramer method, partition II, inverse pairing



MV-mean variability, AC-average complexity, S-small, M-medium, L-large

average number of rejections of the null hypotheses was achieved under the direct pairing condition than in the inverse pairing condition (Appendix II), a development which was opposite to that found for the Kramer method. This finding did not hold true for average complexity across all partitions.

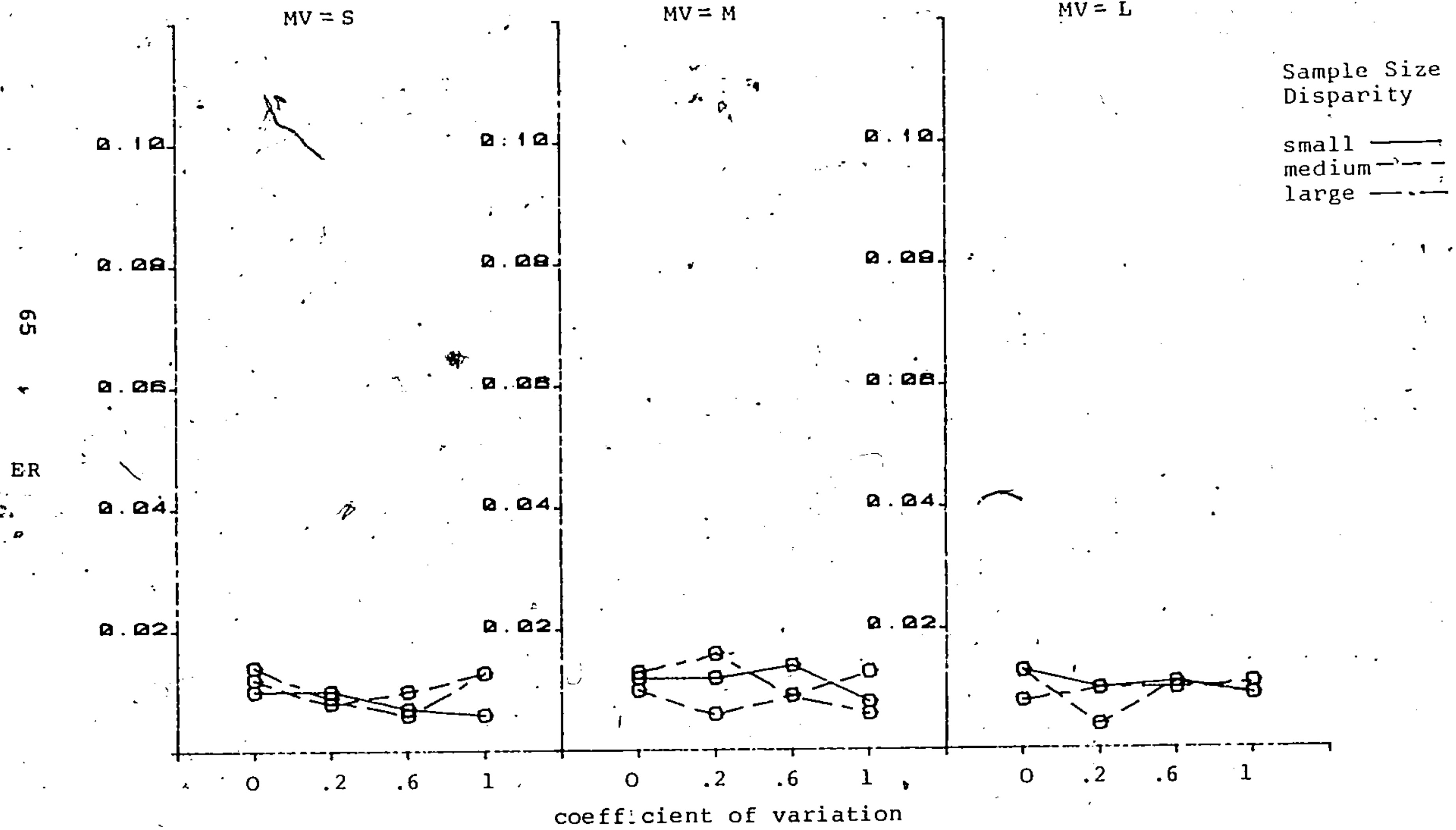
SAMPLE SIZE DISPARITY

It should be noted that there was a strong relationship between the effect of sample size disparity and that of variance heterogeneity; in some cases, this relationship was magnified as mean variability increased. Sample size disparity and variance heterogeneity become increasingly inter-related as the results are described.

Kramer Method

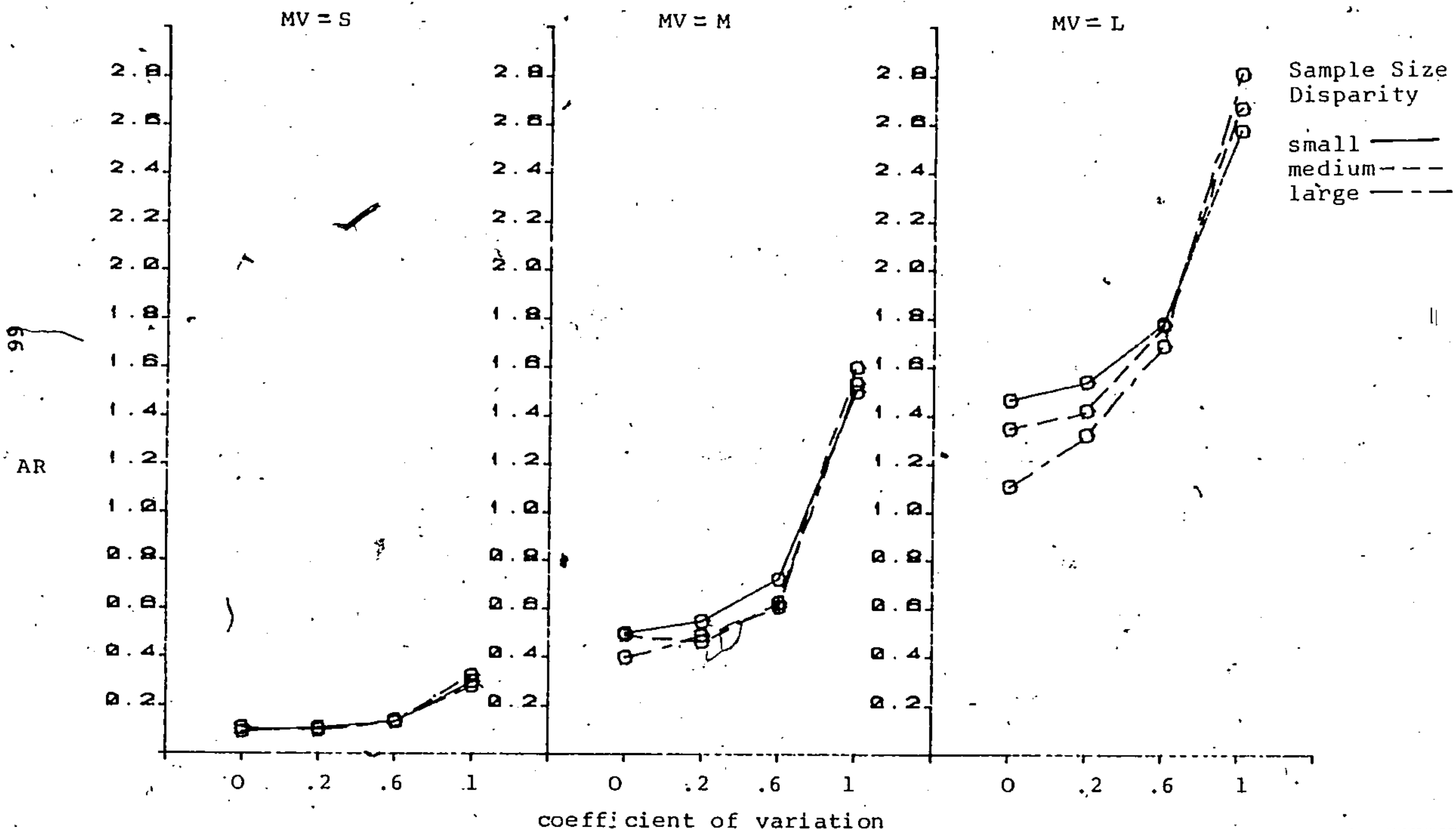
In the direct pairing condition, all three levels of sample size disparity had similar experiment-wise error rates when the variance heterogeneity was small or not present (C = 0, .2). As did the variances, so did the error rates of the various sample size disparity levels become more divergent; when the variance heterogeneity was maximum (C = 1.0), the small sample size disparity had a larger error rate than the medium, which in turn, was larger than the large disparity condition. The separation amongst the sample size disparity levels increased as mean variability increased (Fig. 1).

Figure 7: Type I error rates, Games and Howell method, partition III, direct pairing



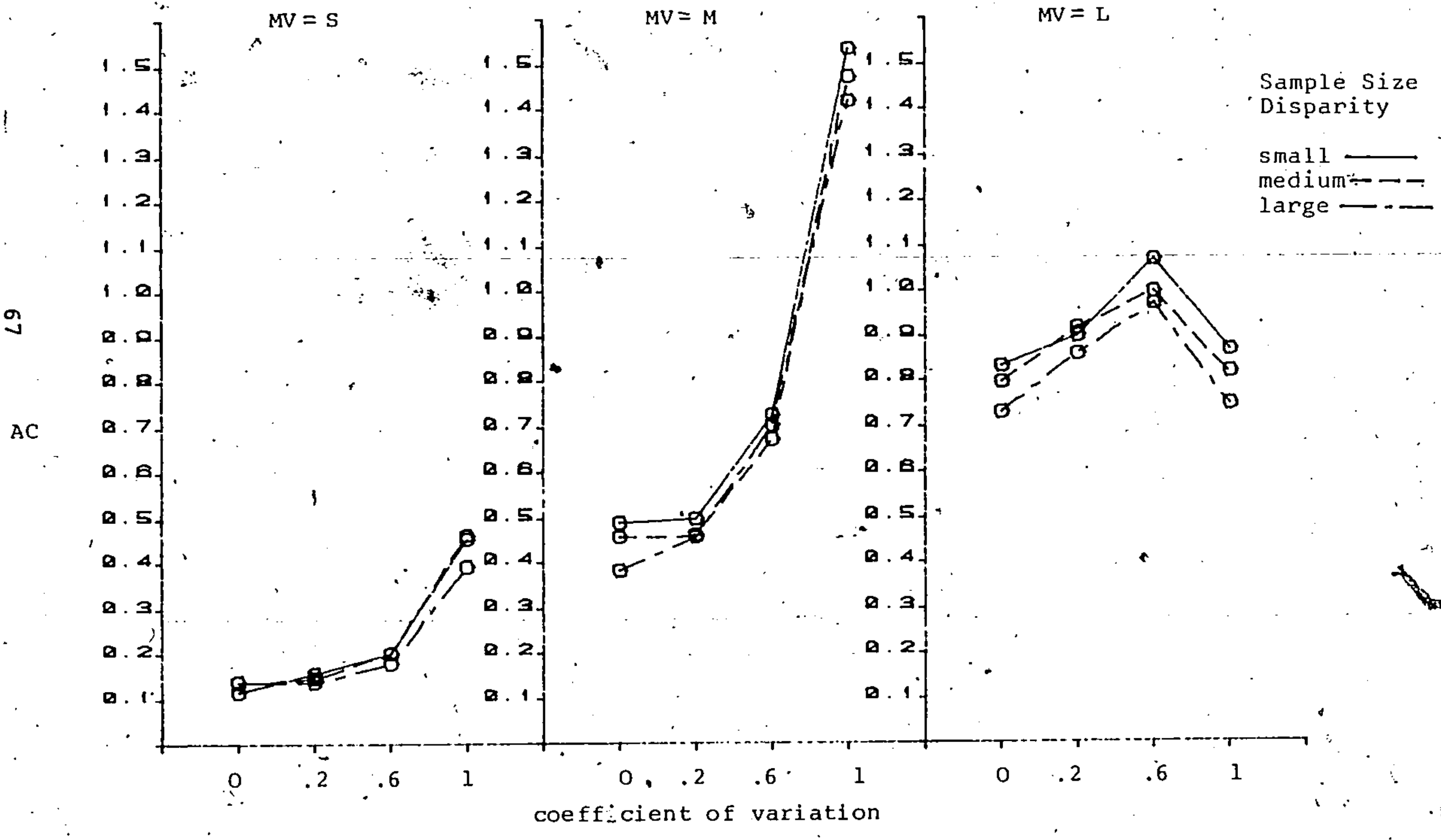
MV-mean variability, ER-experiment-wise error rate, S-small, M-medium, L-large

Figure 8: Average number of rejections, Games and Howell method, partition III, direct pairing



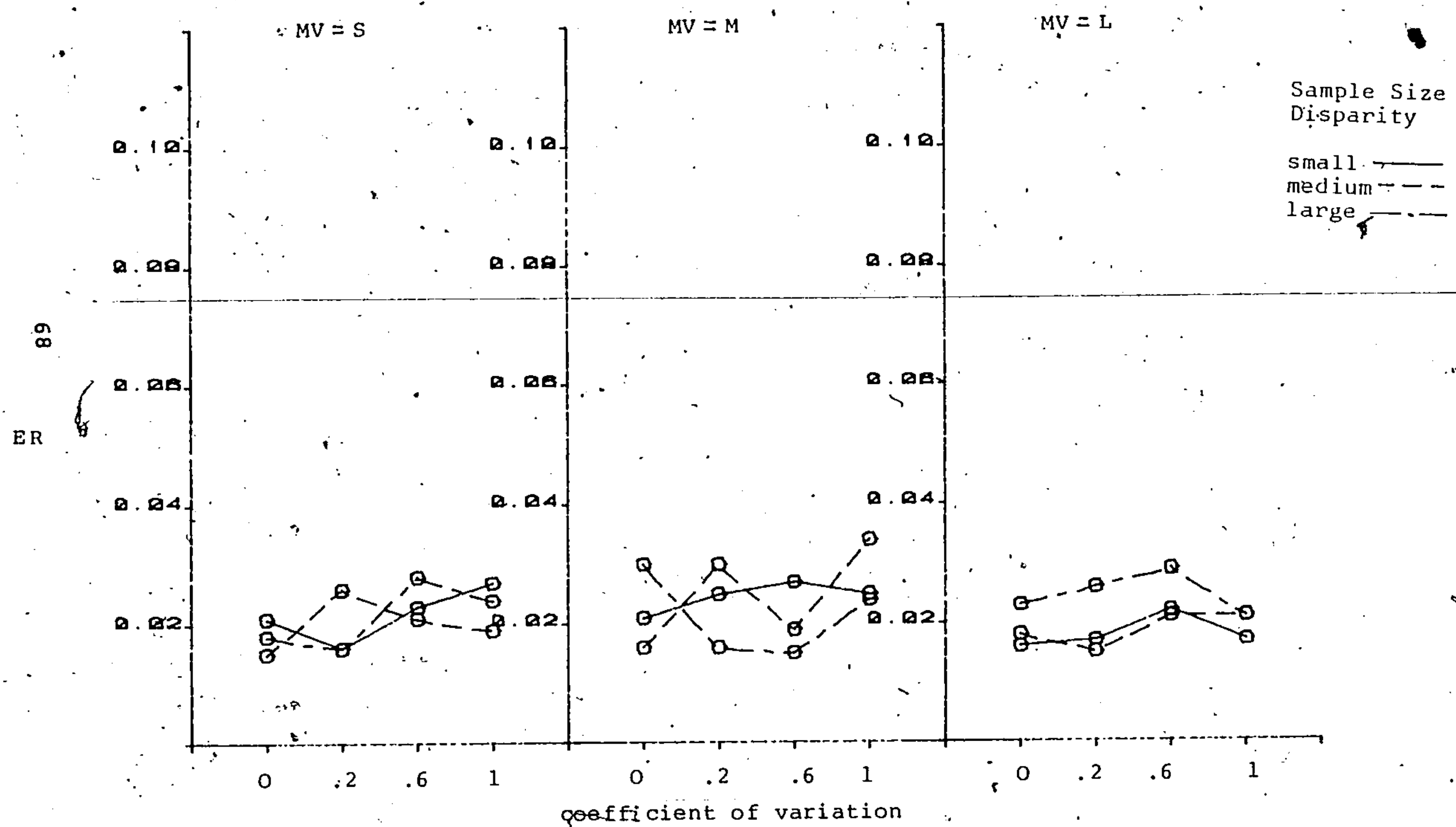
MV-mean variability, AR-average number of rejections, S-small, M-medium, L-large

Figure 9: Average complexity, Games and Howell method, partition III, direct pairing



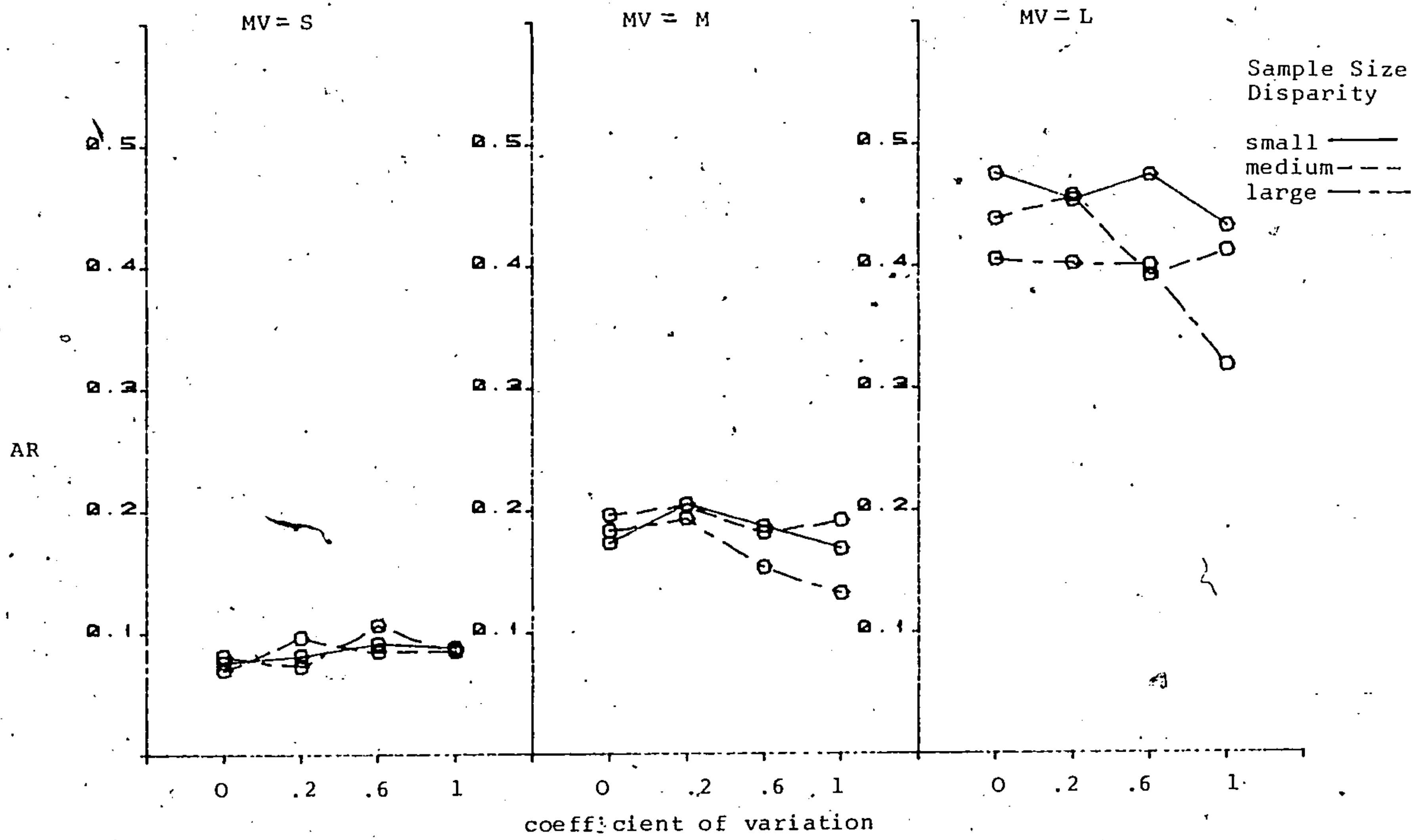
MV-mean variability, AC-average complexity, S-small, M-medium, L-large

Figure 10:. Type I error rates, Games and Howell method, partition II, inverse pairing



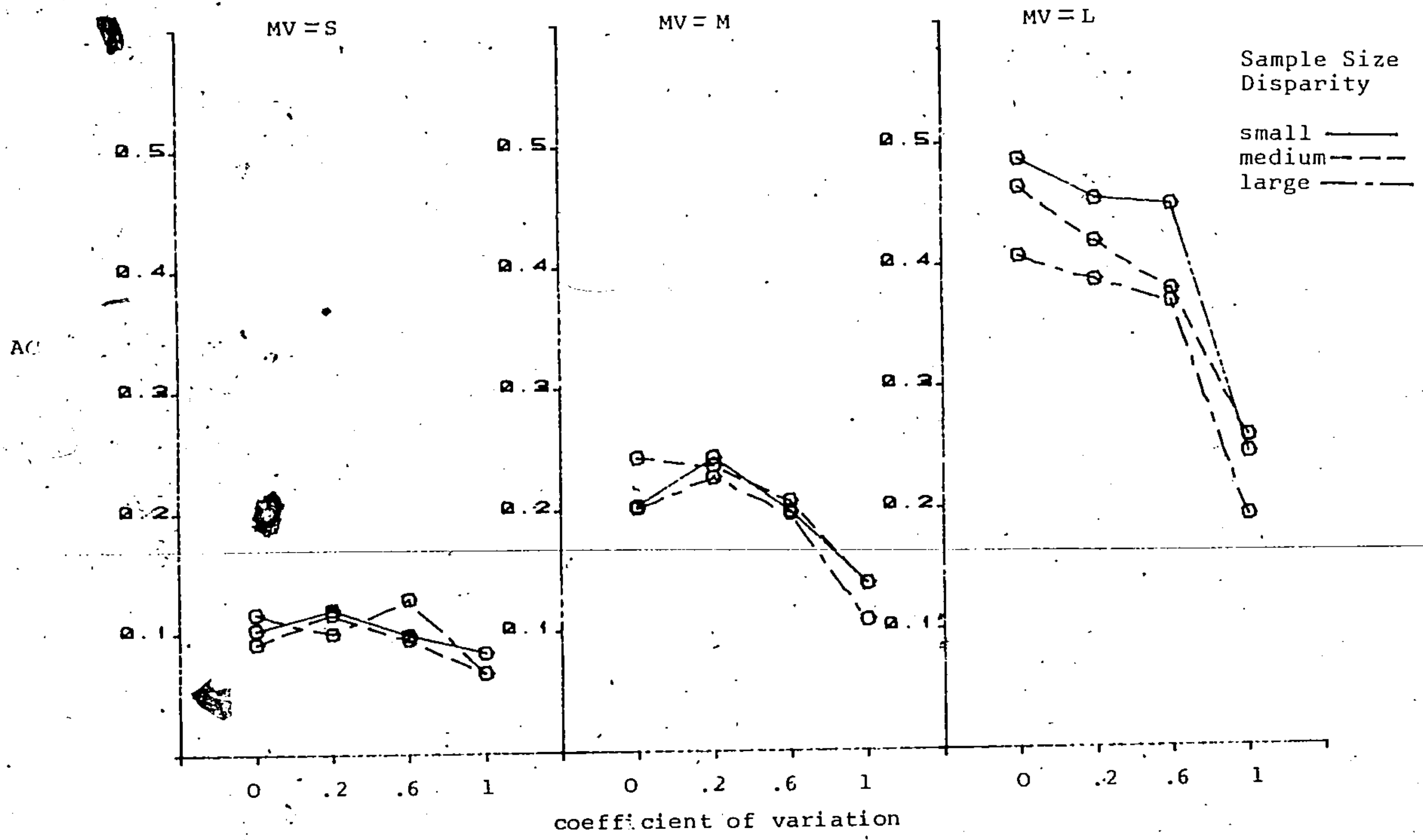
MV-mean variability,ER-experiment-wise error rate,S-small,M-medium, L-large

Figure 11: Average number of rejections, Games and Howell method, partition II, inverse pairing



MV-mean variability, AR-average number of rejections, S-small, M-medium, L-large

Figure 12: Average complexity, Games and Howell method, partition II, inverse pairing



MV-mean variability, AC-average complexity, S-small, M-medium, L-large

A similar trend was evident in the average number of rejections of the null hypotheses (Fig. 2) (Appendix II). The small sample size disparity level had a greater average number of rejections than the medium and large conditions. This fact became more evident as variance heterogeneity increased and mean variability increased. This trend was consistent across all partitions and was related to the increased size of the pooled error term which became large as sample size disparity increased.

Average complexity varied in the same manner as average number of rejections. In most instances (except Partition II & IV, mean variability large, see Appendix II), the small sample size disparity condition had larger average complexities followed by the medium and then the large conditions. This effect increased as variance heterogeneity and mean variability increased (Fig. 3).

In the inverse pairing condition, variance heterogeneity and mean variability in most cases acted to increase the separation amongst the sample size disparity levels as their values increased. This effect was similar to that found in the direct pairing condition but was not as strong, especially when considering average complexity. The major difference between the direct and inverse pairing conditions lies in the relative order of sample size disparity levels. In the inverse pairing condition, the large sample size disparity level had greater experiment-wise error rates and average number of rejections than the medium and small. This was the reversed order from that which was described under the direct pairing condition (Fig. 4, Fig. 5). Average complexity showed less separation than the other two dependent variables amongst the sample size disparity levels. The most common trend seemed to favour the one found in the average number of rejections and experiment-wise error (large > medium > small), but this was not true in all conditions.

Games and Howell Method

The most notable trend of the Games and Howell method was the lack of separation amongst the sample size disparity levels. On the whole, the curves representing the various sample size disparity levels overlapped and lay close together in both the direct and inverse pairing conditions. Though there were exceptions (Partition III & IV, inverse pairing), the Games and Howell method did not appear to be affected by sample size disparity in the same manner as was the Kramer method. This trend was evident in all three dependent variables (Fig. 7, 8, 9, 10, 11, 12)(Appendix II).

VARIANCE HETEROGENEITY

Kramer Method

Experiment-wise error rates increased as variance heterogeneity increased in the direct pairing condition. In most instances, the small sample size condition increased at a greater rate than the medium or large disparity conditions. The differential rate of increase of the sample size disparity levels seemed in part a function of variance heterogeneity, and in part, mean variability (Fig. 1).

The average number of rejections decreased as variance heterogeneity increased. A similar differential effect of variance heterogeneity ~~and mean~~ variability was also found (Fig. 2), such that the small sample size disparity level decreased less rapidly than the medium or large conditions.

Average complexity exhibited a similar differential trend with respect to the sample size disparity leels though, not as pronounced in all conditions. But whereas average number of rejections and experiment-wise error rates (for Kramer) consistently decreased or increased as variance heterogeneity increased, this was not true for average complexity. In some instances--usually when mean variability was large--average complexity would increase when previously it had decreased as variance heterogeneity increased (Fig. 3).

In the inverse pairing condition, experiment-wise error rates increased as variance heterogeneity increased. Again, a differential effect of sample size disparity existed such that error rate for the large disparity condition increased at a rate greater than for the medium or the small conditions. This trend was in the opposite order to that for the direct pairing condition (Fig. 4). The one exception to this trend was partition III; this partition will be considered in a later section of the discussion. In this partition, the experiment-wise error rate decreased to zero for all the sample size disparity levels as variance heterogeneity increased.

Similar trends were demonstrated with average numbers of rejections (Fig. 5), which also increased as variance heterogeneity increased.

Average complexity exhibited an inconsistency similar to that found in the direct pairing condition (Fig. 6). It would seem, however, that the point at which its direction changes is in some way related to increases in mean variability and increased variance heterogeneity. Both these factors bring about increased numbers of rejections of the null hypotheses which could be the common factor affecting complexity.

Games and Howell Method

Experiment-wise error rates were low and stable or decreasing as variance heterogeneity increased in the direct pairing condition (Fig. 7). It would seem that variance heterogeneity had a much smaller effect on the Games and Howell method than it did on the Kramer method.

The average number of rejections was stable (Partition III) or increased as variance heterogeneity increased. The largest increase in number of rejections appeared as the disparity in variances changed from C = .6 to C = 1.0 (Fig. 8).

Average complexity seemed to follow the same trend as the number of rejections (Partition I), but it would, under certain conditions (large mean variability) change direction suddenly as though it had reached a critical point of abutment (Fig. 9).

In the inverse pairing condition, experiment-wise error rates were stable or slightly decreasing as variance heterogeneity increased (Fig. 10).

In partitions I, II, III, and V, the average number of rejections was stable or decreasing as variance heterogeneity increased (Fig. 11). In partition IV, this trend continued until C = .6. When C was equal to 1.0, the number of rejections increased sharply (Appendix II).

Under these conditions, average complexity followed the same trends as that of the average number of rejections. In partitions I, II, III, and , average complexity was stable or decreasing (Fig. 12). In partition IV, complexity increased as variance heterogeneity changed from C = .6 to C - 1.0.

CHAPTER IV

DISCUSSION

The discussion is divided into three sections, each pertaining to the three dependent variables, average number of rejections, experiment-wise error; and, average complexity. This organization of the text should not be construed as a separation amongst the variables as they are strongly inter-related, but rather as a means of describing their relationship.

AVERAGE NUMBER OF REJECTIONS

As mean variability increased, the average number of rejections increased. This was true for both the Kramer and the Games and Howell methods. This finding was a predictable result. With the increase in mean variability, the distinguishable populations (i.e., those with different population means) became more widely separated. Consequently, there was a greater chance that random samples selected from these populations would reflect these differences and thus produce a greater number of significant contrasts.

The increases in the average number of rejections were also a function of the number of possible significant estimated contrasts within each partition.' Due to the structure of the partition, only certain contrasts indicated a true difference between population means. Partition I had three such estimated contrasts ((12) (13) (14)); partition II had four ((13) (14) (23) (24)); partition III had five ((12) (13) (14) (23)

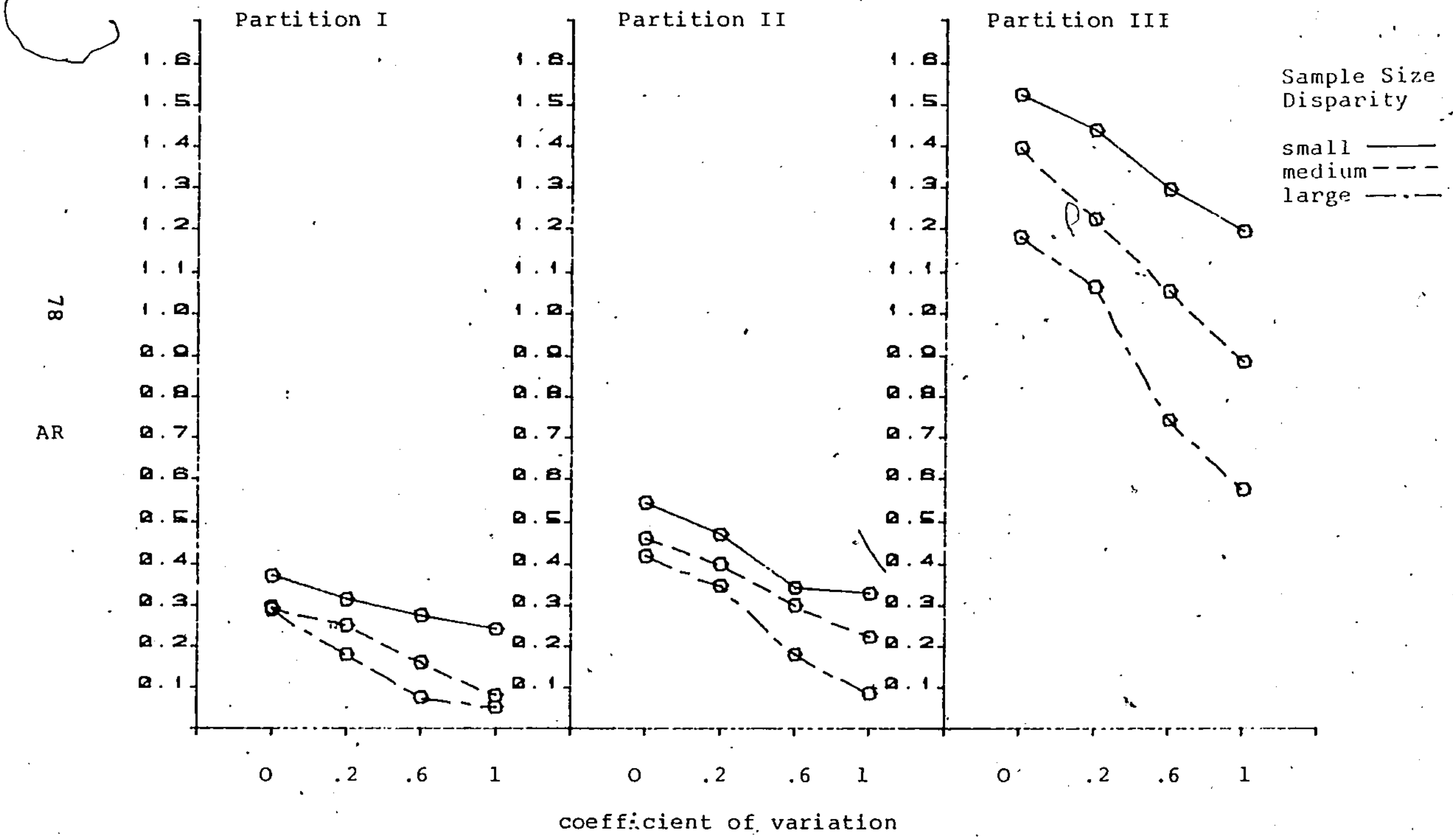
(24)) and partition IV, all six estimated pair-wise contrasts, indicating true differences. As the number of possible true pair-wise contrasts increased, so did the average number of rejections. This finding was common in both the direct and inverse pairing conditions and for both Kramer, and Games and Howell methods. Fig. 13 shows the data for the Kramer method across the first three partitions when mean variability was large.

Provided the rejections are not due to an increase in type I error rates, an increase in the average number of rejections of the null hypotheses is beneficial to the researcher: quite simply, the increase provides more information regarding the partitioning of the population means into mutually-exclusive subsets. As such, it is interesting to consider the frequency with which different contrasts become significant under the various conditions.

The partitions in this study were structured in such a way that the distance between consecutive populations was kept constant across the partitions. If the distance between two consecutive and distinguishable populations is Δ, then a distance measure can be calculated between each pair of means for each contrast within each partition (Table 8).

If other factors were equal, then the frequency of significant contrasts should be directly related to the distance between the means involved in the contrasts. Those with a greater distance should have larger frequencies while those with the same

Figure 13: Average number of rejections, Kramer method, direct pairing, mean variability large



MV-mean variability,AR-average number of rejections distance should have approximately the same frequency. Tables 9 through 16 list the contrast frequencies for the four partitions when the mean variability was large and the sample size disparity was small or large for the Kramer method and for the Games and Howell method, under both direct and inverse pairing conditions.

That the distance between the means involved in the contrast accounted in part for the size of the various frequencies is apparent. Table 9 shows that the contrasts with the largest distances also have the largest frequencies in partitions III and IV. This trend was consistent within the direct pairing condition for both the Games and Howell and Kramer methods. It was also evident in the inverse pairing condition (Tables 13 through 16).

TABLE 8

DISTANCE BETWEEN PAIRS OF MEANS FOR EACH PARTITION

Contrast	Partition I	Partition II	Partition III	Partition IV	Partition V
(12)	1Δ	0	1Δ	1Δ	0
(13)	1Δ	1Δ	2Δ	2Δ	0
(14)	1Δ	1Δ	2Δ	3Δ	0
(23)	0	1Δ	1Δ	1Δ	0
(24)	0	1Δ	1Δ	2Δ	0
(34)	0	0	0	1Δ	0

All factors being equal, it was expected that those contrasts involving the same intermean distance should have similar frequencies. This was true to some extent for the Kramer method in the homogeneous situation (C = 0) when the sample size disparity was small. Increases in either sample size disparity or variance heterogeneity resulted in a divergence amongst the frequencies of the various contrasts (Tables 9, 10).

TABLE 9

FREQUENCY OF SIGNIFICANT ESTIMATED CONTRASTS: KRAMER METHOD, MEAN VARIABILITY LARGE, SAMPLE SIZE DISPARITY SMALL

DIRECT PAIRING

		Coefficient of Variation			
Partition	Contrast	0	.2	.6	1.0
I	(12)*	102	81	28	0
	(13)*	107	89	41	52
[1] [2,3,4]	(14)*	121	106	88	75
	(23)	14	6	3	1
	(24)	10	11	26	9
	(34)	13	17	30	55
II	(12)	14	4	0	0
	(13)*	119	100	39	55
	(14)*	135	109	74	78
[1,2] [3,4]	(23)*	119	115	89	58
	(24)*	142	123	114	87
	(34)	16	17	22	46
III	(12)*	119	76	42	0
	(13)*	580	557	479	465
	(14)*	584	542	535	532
[1][2] [3,4]	(23)*	111	119	76	58
	(24)*	111	123	132	94
	(34)	15	18	29	48
IV	(12)*	112	63	27	0
	(13)*	587	548	488	465
	(14)*	950	937	911	910
[1][2][3][4]	(23)*	132	96	79	54
	(24)*	593	548	520	549
	(34)*	134	163	139	207

Total number of simulations per condition = 1000

* Contrasts which represent a true difference between population means.

The divergence apparent in the contrast frequencies was the result of the differential effect of the interaction of variance heterogeneity and sample size disparity on the dependent variables. With the Kramer method, the average number of rejections decreased as variance heterogeneity increased, and the rate of decrease was much larger for the large sample size disparity condition than the small condition (direct pairing). A comparison of Tables 9 and 10 shows that the small sample size disparity condition had a greater number of rejections than the large sample size disparity condition. As sample size disparity increased, the number of estimated contrasts reaching significance decreased proportionally for all six contrasts. Increases in variance heterogeneity had a differential effect, causing a greater decrease in the size of the frequencies in the large sample size disparity condition. It also caused some estimated contrasts reaching significance to decrease in frequency and others to increase, even though the distance between the population means was the same.

In the direct pairing condition, the smallest sample size was paired with the smallest variance. Thus, samples drawn from population 1 had the smallest n and the smallest variance; those from population 2 had the next smallest and similarily for populations 3 and 4. Consequently, the estimated contrast (12) involved samples that had the smallest sample sizes and the smallest variances, whereas the estimated contrast (34) had the largest n's and variances. The estimated contrast (23) had the

TABLE 10

FREQUENCY OF SIGNIFICANT ESTIMATED CONTRASTS: KRAMER METHOD, MEAN VARIABILITY LARGE, SAMPLE SIZE DISPARITY LARGE

DIRECT PAIRING

		Coefficient of Variation			
Partition	Contrast	0	.2	.6	1.0
I	(12)*	69	37	8	0
	(13)*	95	51	10	9
[1] [2,3,4]	(14)*	96	66	28	14
	(23)	12	9	2	0
	(24)	7	7	9	0
	(34)	8	8	18	27
II	(12)	8	5	0	0
	(13)*	84	58	6	3
	(14)*	99	66	27	11
[1,2] [3,4]	(23)*	96	92	46	11
	(24)*	114	113	86	38
	(34)	13	10	15	21
III	(12)*	62	37	11	0
	(13)*	434	381	250	215
	(14)*	446	452	330	253
[1][2] [3,4]	(23)*	113	79	52	31
	(24)*	119	99	81	45
	(34)	8	13	19	30
IV	(12)*	65	46	5	0
	(13)*	429	414	269	200
	(14)*	900	859	836	789
[1][2][3][4]	(23)*	113	93	66	21
	(24)*	607	525	442	398
	(34)*	176	141	125	162

Total number of simulations per condition = 1000

* Contrasts which represent a true difference between population means.

medium sample sizes and variances. Because each of these contrasts involves consecutive populations, the distance between their respective populations was the same in partition IV. In the other partitions, the estimated contrast (34) involved populations with the same mean.

A comparison of the contrast frequencies under the homogeneous variance condition (C = 0) for partition IV in Tables 9 and 10 shows the effect of sample size disparity. Contrast (1) has a lower frequency than contrast (23) which, in turn, was lower than contrast (34). The difference in frequencies was magnified as sample size disparity increased. It would seem that the term in the Kramer method causing the difference was related to sample sizes. The Kramer confidence interval can be stated as:

$$\overline{X}_i - \overline{X}_j \pm SR_{\alpha.k.N-k}\, s\left([n_i^{-1}+n_j^{-1}]/2\right)^{1/2} . \qquad (5.1)$$

When n_i and n_j were small, the right side of (5.1) was correspondingly large, thus increasing the size of the confidence interval: as a result, fewer estimated contrasts had confidence intervals that did not contain zero (eg. (12)). On the other hand, when n_i and n_j were large, the right side of (5.1) was smaller, thus creating a smaller confidence interval and an increased chance of significance (as seen for contrast (34)).

As sample size disparity increased, the difference between any two sample sizes also increased. As a result, the term $\left([n_i^{-1}+n_j^{-1}]/2\right)^{1/2}$ would also increase in size, thus increasing the size of the confidence interval. The subsequent increase brought about an overall decrease in the number of estimated contrasts attaining significance. It also increases the difference in size between the various $\left(\left[n_i^{-1}+n_j^{-1}\right]/2\right)^{1/2}$ terms for consecutive samples (ie. the contrasts (12) (23) and (34)). Consequently, there was greater differentiation in the large sample size disparity condition for these three contrasts frequencies, as opposed to the small sample size disparity condition.

A similar trend to that described previously for the Kramer method was found for the Games and Howell method (Table 11 and 12). The confidence interval for the Games and Howell method also relies to some extent upon sample sizes,

$$\overline{X}-\overline{X} \pm SR_{\alpha,k,\hat{\nu}_{ij}}\left(s_i^2(2n_i)^{-1}+s_j^2(2n_j)^{-1}\right)^{1/2} \qquad (5.2)$$

A similar argument may also be used here: an increase in sample sizes results in a smaller confidence interval and, therefore, a greater chance of having a confidence interval not containing zero. An increase in sample size disparity results in a larger confidence interval, thus decreasing the number of times an estimated contrast would have a confidence interval without zero.

As variance heterogeneity increased in the direct pairing condition, the frequency of occurrence of certain significant contrasts using the Kramer confidence interval decreased. The frequencies of contrasts (12), (13), (14), (23), (24) decreased in both the small and large disparity conditions. Contrast (34) usually increased in frequency of occurrence. Increased variance heterogeneity leads to an increase in the size of the observed pooled variance term in (5.1). As a result, the confidence interval would increase in size and would result in a decreased frequency of occurrence of significant contrasts. This explanation would be suitable for contrasts (12), (13), (14), (23) and (24), but did not explain the behaviour of the estimated contrast (34). This trend was exhibited in the three sample size disparity levels and occurred in all partitions to some extent regardless of whether the contrasts indicated a true difference or were experiment-wise errors. In the case where they were experiment-wise errors, the size of the frequencies was much smaller, though the trend was still present. The combination of small sample sizes and large variance heterogeneity resulted in contrast (12) achieving a frequency of zero when C = 0 in Figure 9 and 10.

In the direct pairing condition, the Games and Howell method had a similar trend, albeit in the opposite direction. In partition IV in Tables 11 and 12, the frequencies of the first five contrasts increased and the sixth contrast (34) decreased as variance heterogeneity increased. When particular contrasts indicated experiment-wise errors, the trend was still present--though the frequencies were much smaller. This trend was so strong that the Games and Howell method was able to distinguish contrast (12) as a true difference 100% of the time when C = 1.0. This was greater than the frequency of occurrence of significant contrasts with larger differences between their population means.

TABLE 11

FREQUENCY OF SIGNIFICANT ESTIMATED CONTRASTS: GAMES & HOWELL METHOD, MEAN VARIABILITY LARGE, SAMPLE SIZE DISPARITY SMALL

DIRECT PAIRING

		Coefficient of Variation			
Partition	Contrast	0	.2	.6	1.0
I	(12)*	102	124	185	1000
	(13)*	105	116	191	163
[1] [2,3,4]	(14)*	120	119	112	101
	(23)	14	8	13	6
	(24)	9	10	14	12
	(34)	18	10	12	9
II	(12)	17	10	11	7
	(13)*	112	135	200	168
	(14)*	133	117	96	112
[1,2] [3,4]	(23)*	124	121	141	158
	(24)*	141	107	78	117
	(34)	14	11	6	10
III	(12)*	119	116	184	1000
	(13)*	560	624	799	734
	(14)*	561	570	572	583
[1][2] [3,4]	(23)*	99	111	119	145
	(24)*	112	114	95	113
	(34)	13	10	11	9
IV	(12)*	102	108	179	1000
	(13)*	562	602	791	718
	(14)*	942	944	918	931
[1][2][3][4]	(23)*	130	108	115	148
	(24)*	571	520	433	593
	(34)*	126	112	77	68

Total number of simulations per condition = 1000

* Contrast representing a true difference between population means.

TABLE 12

FREQUENCY OF SIGNIFICANT ESTIMATED CONTRASTS: GAMES & HOWELL METHOD, MEAN VARIABILITY LARGE, SAMPLE SIZE DISPARITY LARGE

DIRECT PAIRING

		Coefficient of Variation			
Partition	Contrast	0	.2	.6	1.0
I	(12)*	69	74	134	1000
	(13)*	83	94	173	175
[1] [2,3,4]	(14)*	87	109	140	163
	(23)	7	12	9	21
	(24)	5	8	11	7
	(34)	8	4	12	10
II	(12)	6	10	13	11
	(13)*	81	106	173	176
	(14)*	93	99	131	141
[1,2] [3,4]	(23)*	88	103	121	169
	(24)*	115	116	103	140
	(34)	16	8	11	2
III	(12)*	65	81	135	1000
	(13)*	411	492	729	771
	(14)*	409	532	602	712
[1][2] [3,4]	(23)*	110	110	123	169
	(24)*	102	97	89	157
	(34)	8	10	10	11
IV	(12)*	60	92	130	1000
	(13)*	392	492	750	753
	(14)*	865	897	951	978
[1][2][3][4]	(23)*	113	107	146	176
	(24)*	587	547	459	665
	(34)*	170	110	95	84

Total number of simulations per condition = 1000

* Contrast representing a true difference between population means.

The trend, depicted in the Games and Howell method, may, in part, be explained by the standard errors of the contrast. Hays and Winkler (1971) proposed that the standard error of a pairwise contrast (when the population variances were known) be considered as follows:

$$\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^{1/2} \qquad (5.3)$$

Tables 13 and 18 list the standard errors for each of the sample size disparity levels in the direct and inverse pairing conditions. Table 13 shows that the standard errors for the first five estimated contrasts decreased or remained stable as variance heterogeneity increased. The amount of decrement became larger as sample size disparity increased. Contrast (12) when C = 1.0, had a standard error much smaller than the other contrasts. The large difference between the standard error for contrast (12) and the other contrasts was due to the large range of variances employed in the C = 1.0 condition. The two smallest variances were very much smaller than the largest two. Thus when they were paired together, the standard error was greatly reduced. This may account for the high frequencies of occurrence as a significant contrast in partitions I, III and IV. The standard errors of the contrasts (13) (14) (23) and (24) also decreased, while the frequencies of the significant contrasts increased correspondingly. The standard error for contrast (34) increased as variance heterogeneity increased; corresponding decreases in the frequency of significant estimated contrasts also occurred to a greater extent in the small sample size disparity than in the large.

TABLE 13

STANDARD ERRORS OF THE ESTIMATED CONTRASTS

DIRECT PAIRING

Sample Size Disparity	Contrasts	Coefficient of Variation			
		0	.2	.6	1.0
Small	(12)	.2748	.2493	.2029	.0511
	(13)	.2605	.2446	.2045	.2165
	(14)	.2559	.2498	.2561	.2528
	(23)	.2578	.2566	.2448	.2207
	(24)	.2531	.2615	.2893	.2564
	(34)	.2375	.2571	.2904	.3316
Medium	(12)	.2934	.2649	.2123	.0528
	(13)	.2760	.2562	.2091	.2138
	(14)	.2673	.2550	.2488	.2375
	(23)	.2589	.2574	.2454	.2178
	(24)	.2495	.2562	.2800	.2365
	(34)	.2289	.2472	.2776	.3180
Large	(12)	.3198	.2880	.2290	.0565
	(13)	.2943	.2701	.2151	.2109
	(14)	.2822	.2634	.2435	.2226
	(23)	.2669	.2646	.2518	.2156
	(24)	.2536	.2578	.2765	.2271
	(34)	.2205	.2376	.2651	.3047

The Kramer method in the inverse pairing condition had similar trends to the Games and Howell method in the direct pairing condition (Tables 14 and 15). The first five estimated contrasts ((12) (13) (14) (23) (24)) increased in frequency of occurrence while contrast (34) decreased in occurrence. In the small and large sample size disparity conditions, the frequency of contrast (34) decreased to zero in most of the partitions. This trend had become the exact opposite to that found in the direct pairing condition.

TABLE 14

FREQUENCY OF SIGNIFICANT ESTIMATED CONTRASTS: KRAMER METHOD, MEAN VARIABILITY LARGE, SAMPLE SIZE DISPARITY SMALL

INVERSE PAIRING

Partition	Contrast	Coefficient of Variation 0	.2	.6	1.0
I	(12)*	88	115	134	216
	(13)*	92	117	165	199
[1] [2,3,4]	(14)*	111	135	144	214
	(23)	10	19	5	13
	(24)	9	16	3	10
	(34)	13	2	0	0
II	(12)	11	24	47	93
	(13)*	110	112	169	213
	(14)*	116	123	172	215
[1,2] [3,4]	(23)*	121	128	130	159
	(24)*	122	145	106	164
	(34)	8	3	1	0
III	(12)*	97	124	161	230
	(13)*	555	546	554	612
	(14)*	578	568	572	621
[1][2] [3,4]	(23)*	120	134	100	156
	(24)*	120	125	97	174
	(34)	15	5	1	0
IV	(12)*	103	139	166	233
	(13)*	579	565	590	611
	(14)*	938	934	942	941
[1][2][3][4]	(23)*	129	113	116	133
	(24)*	592	587	633	645
	(34)*	151	119	94	0

Total number of simulations per condition = 1000

* Contrast representing a true difference between population means

TABLE 15

FREQUENCY OF SIGNIFICANT ESTIMATED CONTRASTS: KRAMER METHOD, MEAN VARIABILITY LARGE, SAMPLE SIZE DISPARITY LARGE

INVERSE PAIRING

		Coefficient of Variation			
Partition	Contrasst	0	.2	.6	1.0
I	(12)*	77	112	178	278
	(13)*	87	107	196	290
[1] [2,3,4]	(14)*	103	120	201	305
	(23)	11	24	14	50
	(24)	10	18	13	54
	(34)	6	5	2	0
II	(12)	11	26	57	150
	(13)*	83	122	187	273
	(14)*	100	121	185	292
[1,2] [3,4]	(23)*	111	125	163	236
	(24)*	124	137	161	261
	(34)	6	8	2	0
III	(12)*	76	113	183	296
	(13)*	421	474	518	644
	(14)*	444	516	558	660
[1][2] [3,4]	(23)*	109	151	137	255
	(24)*	113	147	145	286
	(34)	12	6	2	0
IV	(12)*	83	97	167	267
	(13)*	442	477	499	638
	(14)*	876	893	875	906
[1][2][3][4]	(23)*	104	128	138	250
	(24)*	595	635	688	770
	(34)*	172	187	195	76

Total number of simulations per condition = 1000

* Contrast representing a true difference between population means

The Games and Howell method in the inverse pairing condition demonstrated a similar trend to the Kramer method in the direct pairing condition and the opposite trend to that exhibited by the Games and Howell technique in the direct pairing condition. In this situation, the frequency of the first five estimated contrasts ((12) (13) (14) (23) (24)) decreased while that of the sixth ((34)) increased as variance heterogeneity increased (Table 16 and 17). The standard errors of the contrast for the inverse pairing condition are listed in Table 18. The trend was opposite to that found in the direct pairing condition.

A similar argument could be used to explain the changes in the frequency of the contrasts. An increased standard error would bring about a decrease in the frequency of occurrence of that particular contrast.

While the Games and Howell trends could be explained easily by the standard errors of the contrast, use of a pooled error term complicated the explanation with the Kramer method. In the direct pairing condition, the large sample sizes were paired with the largest variances, and as such, were the contributing factor to the pooled error term. As the variance heterogeneity increased, the variances associated with the large sample sizes became larger, thus causing a corresponding increase in the pooled error term.

TABLE 16

FREQUENCY OF SIGNIFICANT ESTIMATED CONTRASTS: GAMES & HOWELL METHOD, MEAN VARIABILITY LARGE, SAMPLE SIZE DISPARITY SMALL

INVERSE PAIRING

		Coefficient of Variation			
Partition	Contrasst	0	.2	.6	1.0
I	(12)*	82	78	63	50
	(13)*	99	81	50	86
[1] [2,3,4]	(14)*	107	120	77	93
	(23)	9	13	6	8
	(24)	9	13	9	7
	(34)	11	9	16	12
II	(12)	10	10	9	13
	(13)*	105	82	68	76
	(14)*	120	102	89	74
[1,2] [3,4]	(23)*	117	109	120	131
	(24)*	116	144	175	133
	(34)	7	7	13	4
III	(12)*	90	88	58	44
	(13)*	532	476	346	374
	(14)*	566	525	394	376
[1][2] [3,4]	(23)*	116	119	97	121
	(24)*	128	128	174	133
	(34)	13	12	18	5
IV	(12)*	96	87	64	45
	(13)*	556	482	375	350
	(14)*	924	902	859	773
[1][2][3][4]	(23)*	122	101	119	104
	(24)*	574	584	739	552
	(34)*	137	156	283	1000

Total number of simulations per condition = 1000

* Contrast representing a true difference between population means

TABLE 17

FREQUENCY OF SIGNIFICANT ESTIMATED CONTRASTS: GAMES & HOWELL METHOD, MEAN VARIABILITY LARGE, SAMPLE SIZE DISPARITY LARGE

INVERSE PAIRING

		Coefficient of Variation			
Partition	Contrast	0	.2	.6	1.0
I	(12)*	71	54	44	30
	(13)*	82	67	54	44
[1] [2,3,4]	(14)*	89	65	49	43
	(23)	8	20	11	12
	(24)	9	16	17	12
	(34)	7	6	11	8
II	(12)	14	12	18	13
	(13)*	77	70	53	57
	(14)*	86	86	46	53
[1,2] [3,4]	(23)*	103	103	123	95
	(24)*	116	117	149	92
	(34)	9	14	11	8
III	(12)*	63	61	46	40
	(13)*	392	345	215	222
	(14)*	409	392	242	219
[1][2] [3,4]	(23)*	102	115	90	102
	(24)*	110	130	138	108
	(34)	10	11	11	19
IV	(12)*	75	54	48	31
	(13)*	405	324	214	215
	(14)*	826	799	558	527
[1][2][3][4]	(23)*	101	107	104	102
	(24)*	562	573	663	500
	(34)*	174	205	311	1000

Total number of simulations per condition = 1000

* Contrast representing a true difference between population means

TABLE 18

STANDARD ERRORS OF THE ESTIMATED CONTRAST

INVERSE PAIRING

Sample Size Disparity	Contrasts	Coefficient of Variation 0	.2	.6	1.0
Small	(12)	.2748	.2981	.3392	.3859
	(13)	.2605	.2755	.3169	.3036
	(14)	.2559	.2619	.2916	.3011
	(23)	.2578	.2589	.2483	.2455
	(24)	.2531	.2444	.2151	.2425
	(34)	.2375	.2163	.1779	.0451
Medium	(12)	.2934	.3194	.3672	.4155
	(13)	.2760	.2946	.3436	.3371
	(14)	.2673	.2790	.3200	.3349
	(23)	.2589	.2604	.2499	.2500
	(24)	.2495	.2427	.2162	.2470
	(34)	.2289	.2090	.1732	.0442
Large	(12)	.3189	.3409	.4030	.4549
	(13)	.2943	.3166	.3740	.3740
	(14)	.2822	.2999	.3520	.3721
	(23)	.2669	.2693	.2589	.2653
	(24)	.2536	.2493	.2259	.2625
	(34)	.2206	.2019	.1687	.0433

In the inverse pairing condition, the larger sample sizes were paired with the smaller variances: as the variance heterogeneity increased, the variances associated with the large sample sizes decreased-- reducing the size of the pooled error term. The changes in the pooled error term could account for the changes in the frequency of occurrence of the first five contrasts in the direct and inverse pairing conditions. The changes do not explain the behaviour of the sixth contrast (34).

Partition V was structured such that the four samples were drawn from populations with the same means, and as a consequence, none of the contrasts indicated a true difference between population means. This makes the difference between the means the same for all contrasts. In this situation, it is simpler to examine the trends in the frequencies of the contrasts, since they can now be considered equally. Tables 19 and 20 list the frequencies of the six contrasts for the three levels of sample size disparity within the direct and inverse pairing conditions. Similar trends to those described in the other four partitions are observed in the fifth partition. In the direct pairing condition, the sixth contrast increased in frequency as variance heterogeneity increased, and to a greater extent in the small sample size disparity condition than the large. In the inverse condition, the sixth contrast decreased in frequency of occurrence, while the other five contrasts increased in frequency to a greater extent in the large sample size disparity condition.

EXPERIMENT-WISE ERROR RATE

Experiment-wise type I error rates exhibited similar trends to those found by previous researchers (Keselman and Rogan 1978, Dunnett 1980a, b). In the direct pairing condition with the Kramer method, the error rates increased as variance heterogeneity increased. The small sample size disparity condition increased to a greater extent than the large. In the inverse pairing condition, error rates again increased as variance heterogeneity increased, but to a greater extent in the large sample size disparity condition. The Games and Howell technique was not greatly affected by sample size disparity and maintained relative low and stable error rates as variance heterogeneity increased in most cases.

TABLE 19

FREQUENCY OF SIGNIFICANT ESTIMATED CONTRASTS: KRAMER METHOD

PARTITION V - DIRECT PAIRING

Sample Size Disparity	Contrasts	Coefficient of Variation 0	.2	.6	1.0
Small	(12)	9	6	0	0
	(13)	15	4	0	0
	(14)	10	11	7	6
	(23)	7	4	12	0
	(24)	6	14	7	7
	(34)	16	22	25	39
Medium	(12)	12	3	0	0
	(13)	11	7	0	0
	(14)	11	7	0	1
	(23)	13	8	2	0
	(24)	10	7	10	4
	(34)	13	19	22	40
Large	(12)	11	6	0	0
	(13)	13	4	0	0
	(14)	9	4	2	0
	(23)	7	11	4	0
	(24)	15	8	7	1
	(34)	15	13	17	26

Total number of simulations per condition = 1000

TABLE 20

FREQUENCY OF SIGNIFICANT ESTIMATED CONTRASTS: KRAMER METHOD

PARTITION V - INVERSE PAIRING

Sample Size Disparity	Contrasts	Coefficient of Variation 0	.2	.6	1.0
Small	(12)	12	29	32	95
	(13)	10	22	28	45
	(14)	9	15	14	48
	(23)	11	15	13	16
	(24)	13	15	5	15
	(34)	8	5	1	0
Medium	(12)	6	22	44	116
	(13)	12	23	53	72
	(14)	12	15	41	81
	(23)	6	17	14	26
	(24)	13	13	8	24
	(34)	15	5	1	0
Large	(12)	8	37	66	139
	(13)	9	21	74	87
	(14)	14	19	70	109
	(23)	14	9	19	41
	(24)	15	7	4	55
	(34)	14	7	3	0

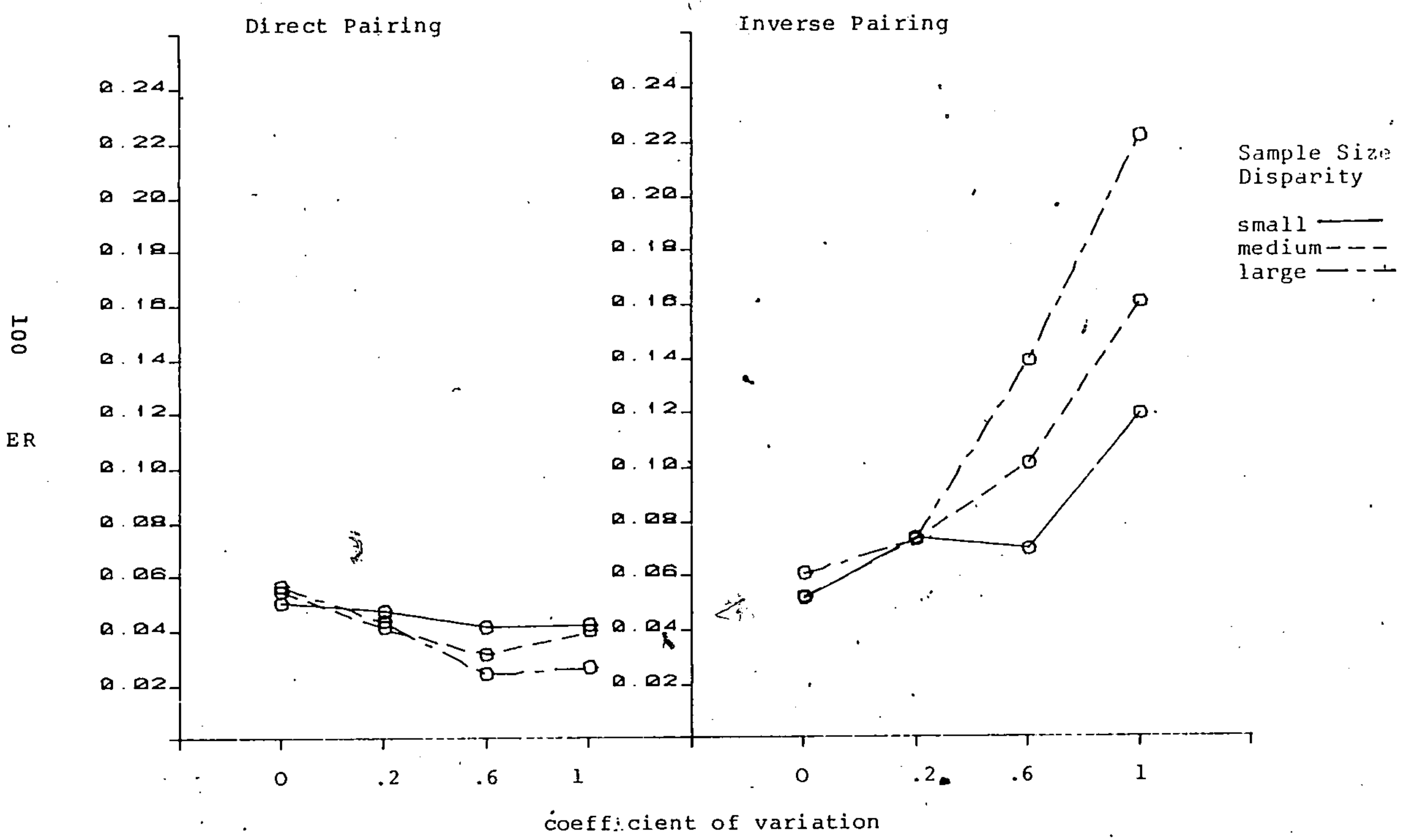
Total number of simulations per conditon = 1000

Though the trends were similar between the present research and that previously done, there were differences in the sizes of the error rates. This was due to the fact that previous researchers have always collapsed error rates across the partitions studied, or they have studied only one or two specific partitions. The present research demonstrates a wide range of error rates across the various partitions (Appendix II).

Two specific partitions should be considered. Fig. 14 and 15 show the experiment-wise error rates for the Kramer, and Games and Howell techniques under both the inverse and direct pairing conditions for partition V. This partition, structured such that all four samples were drawn from populations with the same mean, was frequently used in previous research to generate experiment-wise error rates under various conditions. Partition V yielded error rates similar in size and direction to those cited in such studies as Keselman and Rogan (1978) and Dunnett (1980a, b).

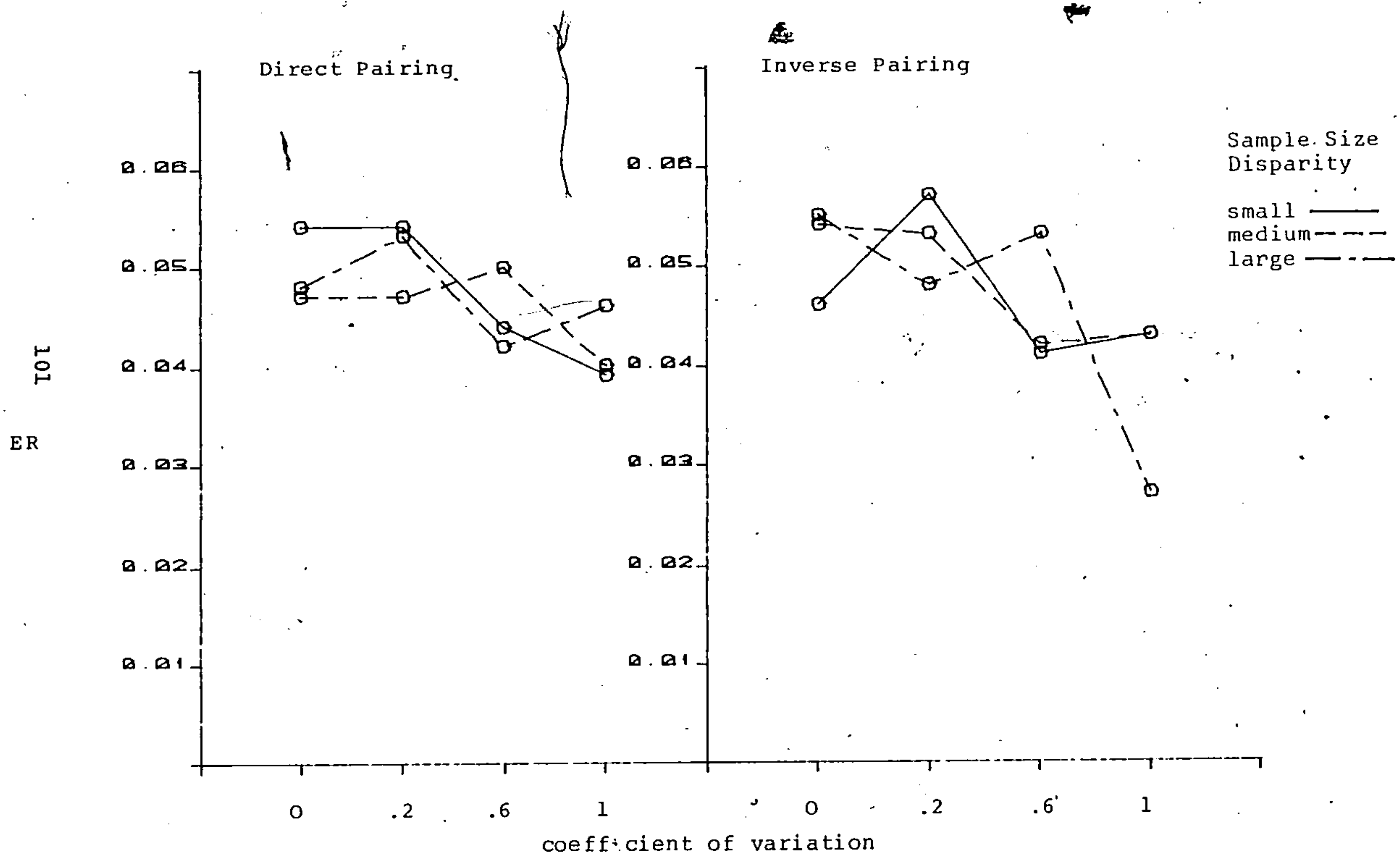
Experiment-wise error rates and the average number of rejections are directly related when examining those contrasts which indicate an error rather than a true difference. Thus, the trends described in the previous section are relevant to the discussion of error rates. In partitions I, II and III, the estimated contrasts representing experiment-wise errors were (23) (24) (34); (12) (34), and (34) respectively. Of specific interest were the trends which caused an increase in frequency of (34) and decreases in frequency of (12) in the direct pairing condition for the Kramer method, and the reversal of this trend in the inverse pairing condition. As such, contrast (34) was the major contributor to the error rate in the direct pairing condition. In the inverse pairing condition, the other error contrasts were significant contributors. This situation created the unexpected decrease in error rates in the third partition in the inverse pairing condition (Fig. 16). Partition III's error rates were due solely to estimated contrast (34) which decreased in frequency of occurrence in the inverse pairing condition.

Figure 14: Type I error rates, Kramer method, partition V



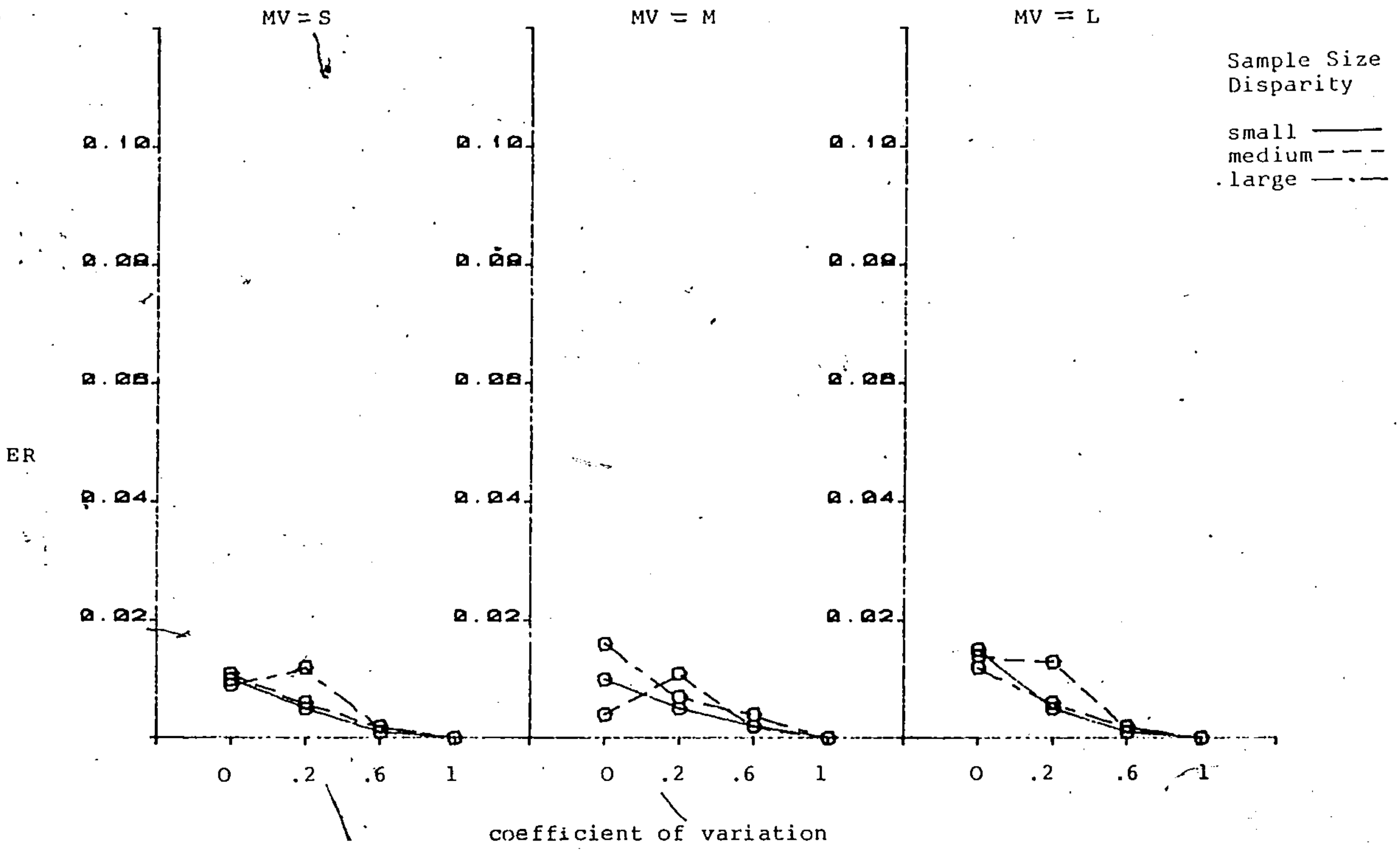
MV-mean variability,ER-experiment-wise error rate,S-small,M-medium, L-large

Figure 15: Type I error rates, Games and Howell method, partition V



MV-mean variability,ER-experiment-wise error rate,S-small,M-medium, L-large

Figure 16: Type I error rates, inverse pairing, Kramer method, partition III



MV-mean variability,ER-experiment-wise error rate,S-small,M-medium, L-large

The Games and Howell method experienced similar trends with respect to contrasts representing experiment-wise errors, but the changes were much less pronounced. The changes that afflicted the Kramer method were more controlled by the Games and Howell technique in the sense that the Games and Howell method did not reflect the wide variation in error rates, but neither did it experience zero error rates. Partition III in the inverse pairing condition had low error rates, though not as low as the Kramer method.

Experiment-wise error rates did not change substantially as mean variability increased for either the Games and Howell or the Kramer techniques. It was expected that the error rates would decrease as the distance between the parent populations increased. It was possible that the decrease in error rates was masked by the changes in error rates due to sample size disparity and variance heterogeneity. Keselman and Rogan (1978) also found that mean variability was unrelated to experiment-wise error rates.

The trends in the average number of rejections of the null hypotheses described in the previous section were evident in the experiment-wise error data. The same arguments which accounted for increases in the number of rejections also explain the changes in experiment-wise errors with respect to sample size disparity, variance heterogeneity and direct and inverse pairing.

AVERAGE COMPLEXITY.

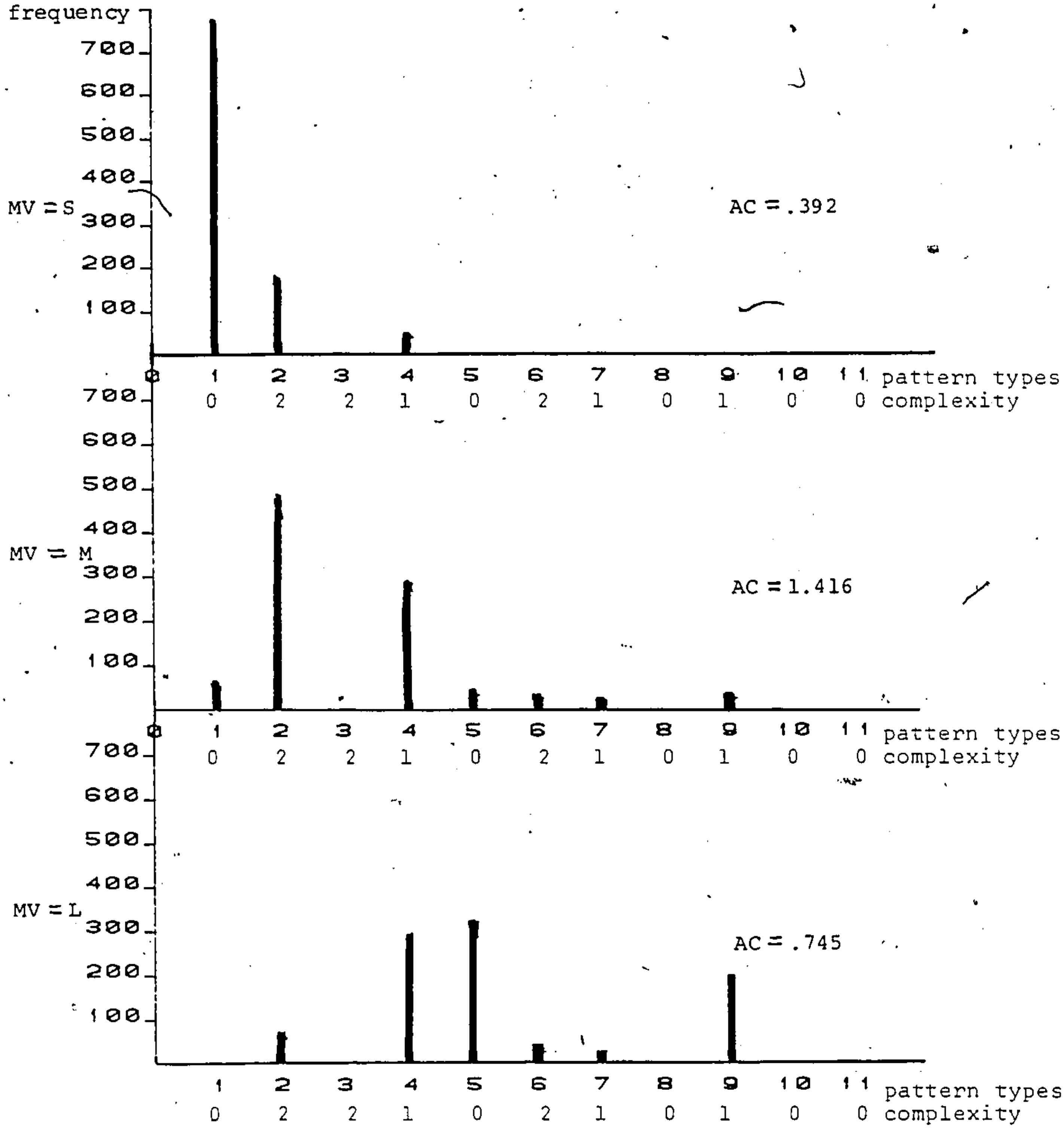
Shaffer's (1981) concept of complexity employed the categorization of various rejections of the null hypotheses into patterns. In the four group design, there are sixty-four possible combinations of the six pair-wise contrasts. Following Shaffer's criteria, the sixty-four patterns have been divided into eleven pattern types, each type containing patterns with the same complexity and the same number of rejections (Table.3).

Observation of Table 3 indicates that as pattern types change from lower order to higher order, complexity decreases. It is also apparent that the average number of rejections increases as the patterns change from lower to higher order. Since one of the criteria for distinguishing amongst patterns is number of rejections, it is clear that average complexity is directly related to the average number of rejections. As a result, the trends in the data that were discussed previously should also be apparent in the average complexity.

As mean variability increases, average complexity increases (Fig. 3, 6, 9, 12). (That the trend was not consistent in all cases is shown in Fig. 9, where the medium mean variability reachs higher average complexities than the large mean variability). Nonetheless, the average number of rejections shows a continuous increase as mean variability increases (Fig. 8). This apparent inconsistency can be explained by the relative increases and decreases in the frequency of occurrence of various pattern types (Appendix IV). Fig. 17 shows the shift in the frequency of occurrence of the eleven pattern types as mean variability increased for the same method and partition described in Fig. 8 and 9 (Games and Howell method, Partition III). The frequencies shifted from the lower order patterns to the higher order patterns. In the medium mean variability (which experienced a large increase in average complexity when variance heterogeneity was maximum (C - 1.0)), pattern types 2 and 4 had large increases in frequency. The complexities of these two patterns were 2 and 1 respectively. In the small mean variability condition, the major pattern type was one which had a complexity of zero. The change in the complexities of the most frequent patterns resulted in similar changes in the average complexity. In the large mean variability condition, the major contributors were pattern types 4, 5 and 9, with complexities of 1, 0 and 1. Average complexity decreased.

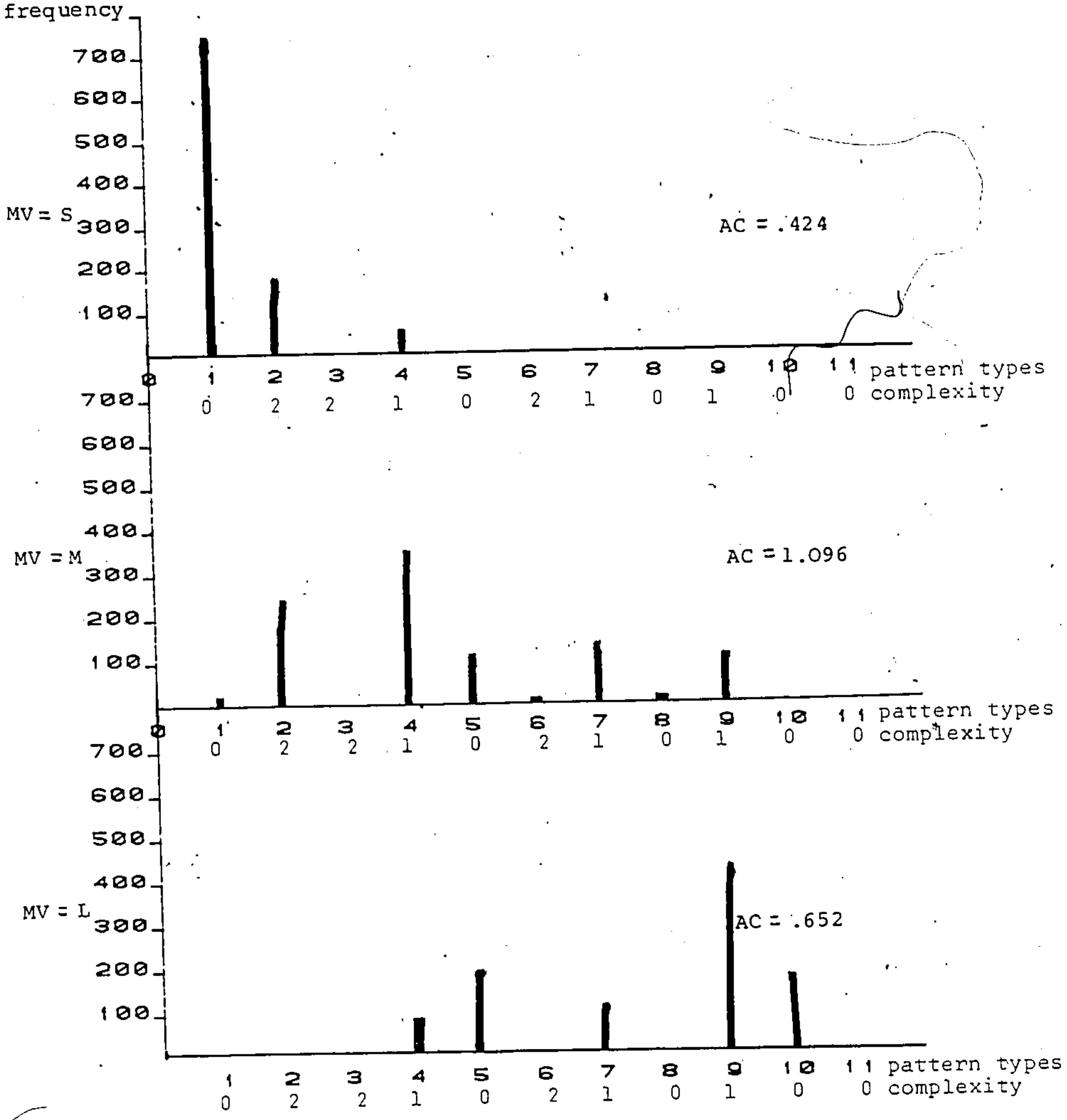
A similar, but more dramatic effect was evident in partition IV direct pairing for the Games and Howell method. Because the shift towards the higher order patterns was greater in Fig. 18 than Fig. 17, the average complexities were lower. This result concurred with the finding that average number of rejections increase as the partitions increase. Due to the fact that higher order pattern types had lower complexities, average complexity decreased. The changes in average complexity are directly related to the frequency of various patterns.

Figure 17: Average complexity, Games and Howell method, partition III, variance heterogeneity (C = 1.0), sample size disparity (large), direct pairing



MV-mean variability, AC-average complexity, S-small, M-medium, L-large

Figure 18: Average complexity, Games and Howell method, partition IV, variance heterogeneity (C = 1.0), sample size disparity (large), direct pairing



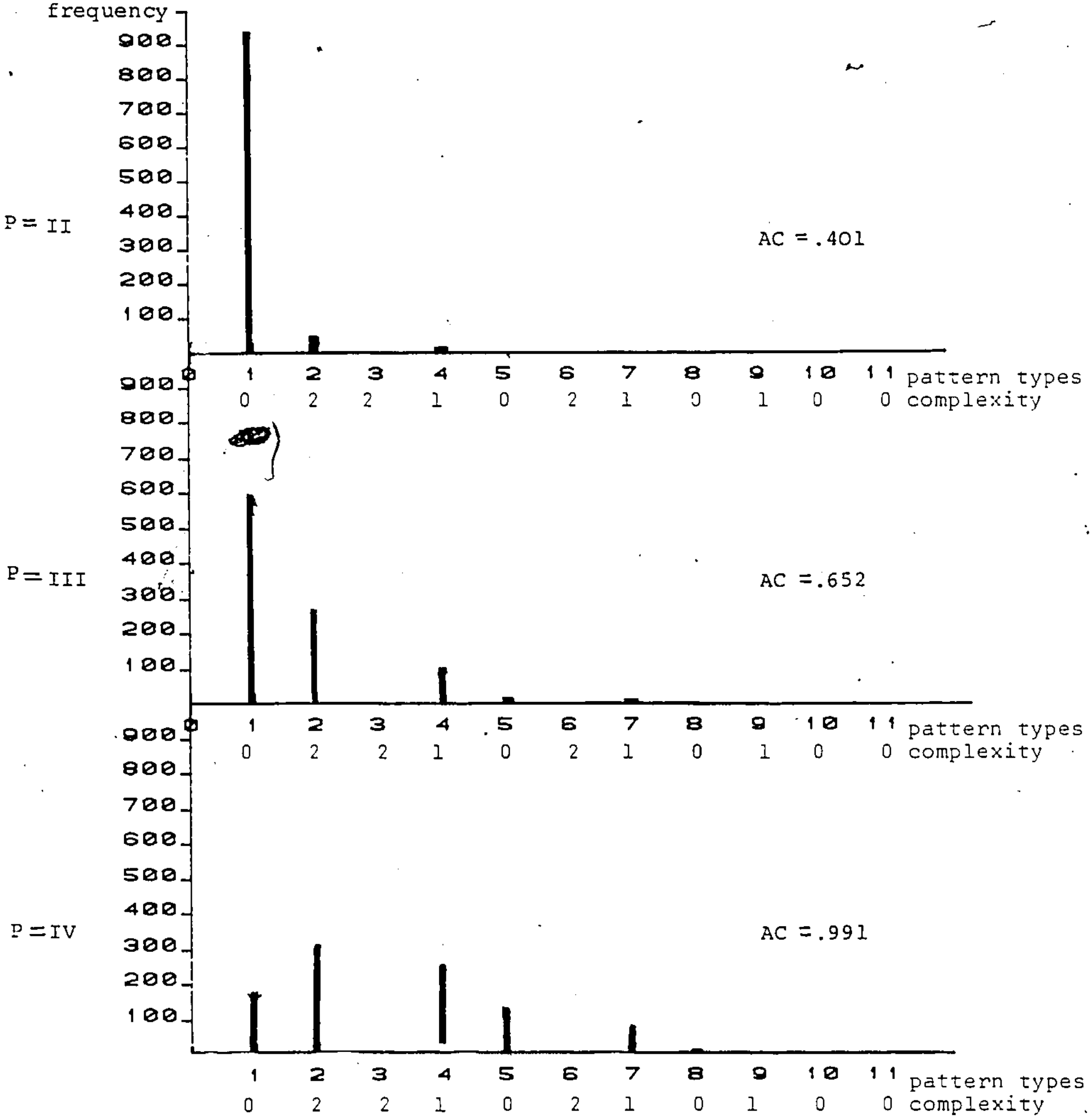
MV-mean variability, AC-average complexity, S-small, M-medium, L-large

The change to lower average complexities depended upon the involvement of higher order patterns. If the lower order patterns increased in frequency, then average complexity increased. Pattern type 1 with a complexity of zero represented a pool of possible significant contrasts from which the frequeies of the other patterns were removed or replaced. Fig. 19 shows how average complexity increased as partitions increased because lower order patterns were the main contributors. It is now apparent that average complexity is often misleading when compared to changes in pattern frequency.

The differential effect created by the interaction of sample size disparity and variance heterogeneity on the performance of the Kramer method in both the direct and inverse pairing conditions was evident with average complexity, but the effect was not as strong as that experienced for the average number of rejections. In the direct pairing condition, increased sample size disparity caused a shift towards the lower order patterns--especially pattern type I (Complexity = 0)--and a resulting decrease in average complexity. In the homogeneous variance situation (Fig. 20), the effect was present but not nearly as marked as when variance heterogeneity reached its maximum value (Fig. 21).

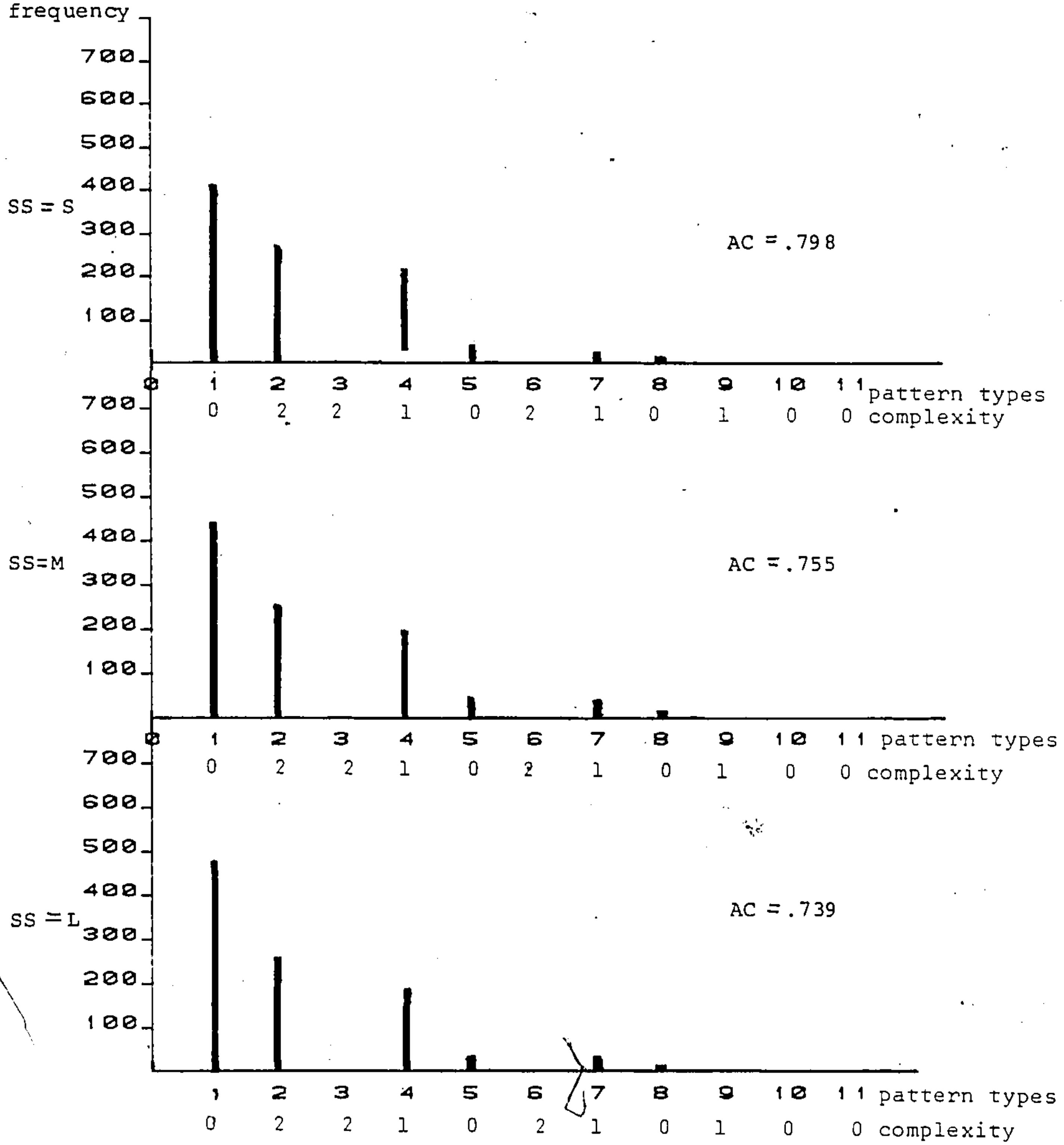
In the inverse pairing condition, the differences among the average complexities for the different sample size disparity levels were smaller than in the direct pairing condition, yet, the pattern type frequencies indicated a change in trend. In the

Figure 19: Average complexity, Kramer method, mean variability (large), variance heterogeneity (C = 1.0), sample size disparity (large), direct pairing



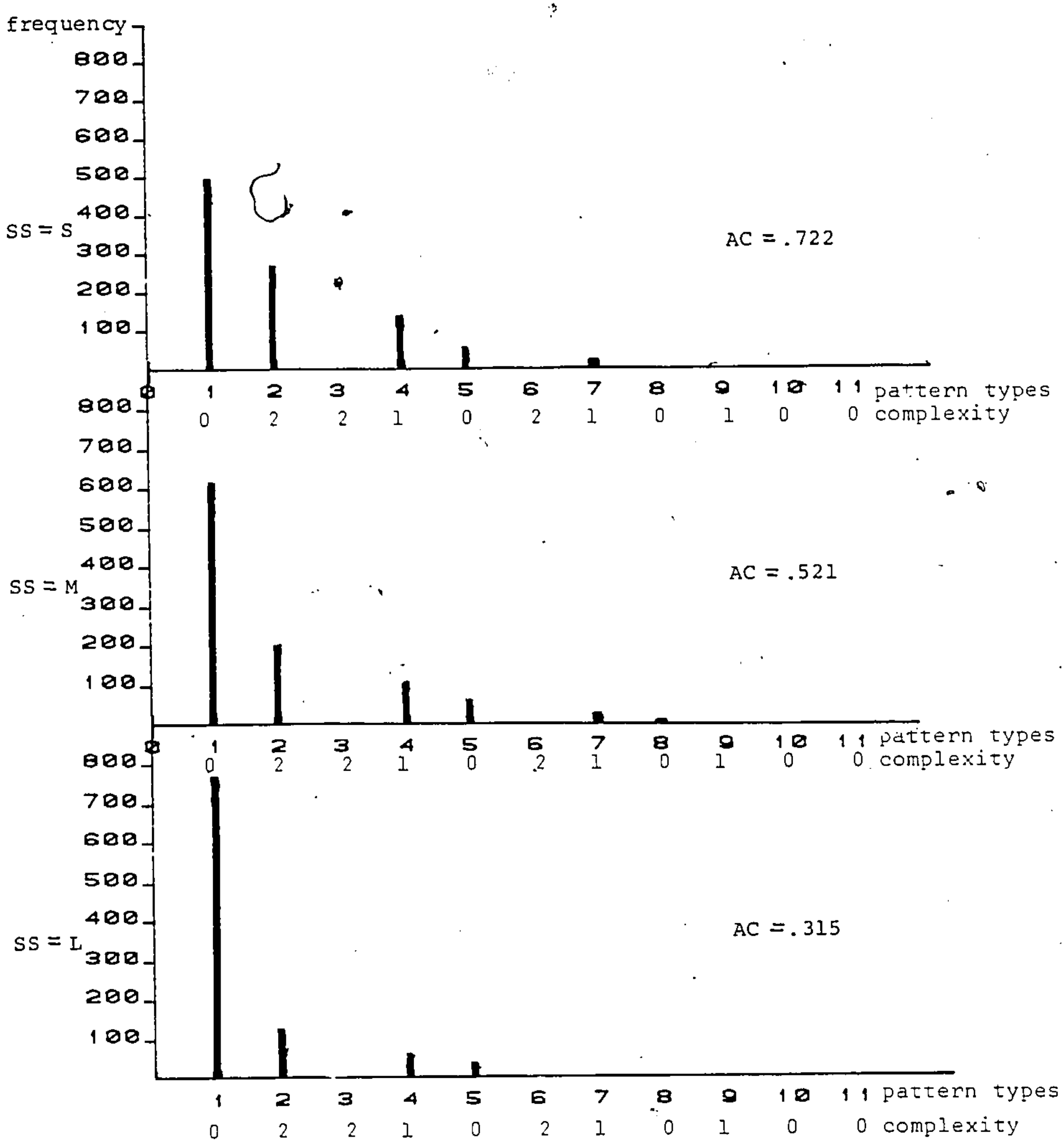
P-partition , AC-average complexity

Figure 20: Average complexity, Kramer method, partition IV, mean variability (medium), variance heterogeneity (C = 0), direct pairing



SS-sample size disparity, AC-average complexity, S-small, M-medium, L-large

Figure 21: Average complexity, Kramer method, partition IV, mean variability (medium), variance heterogeneity (C = 1.0), direct pairing

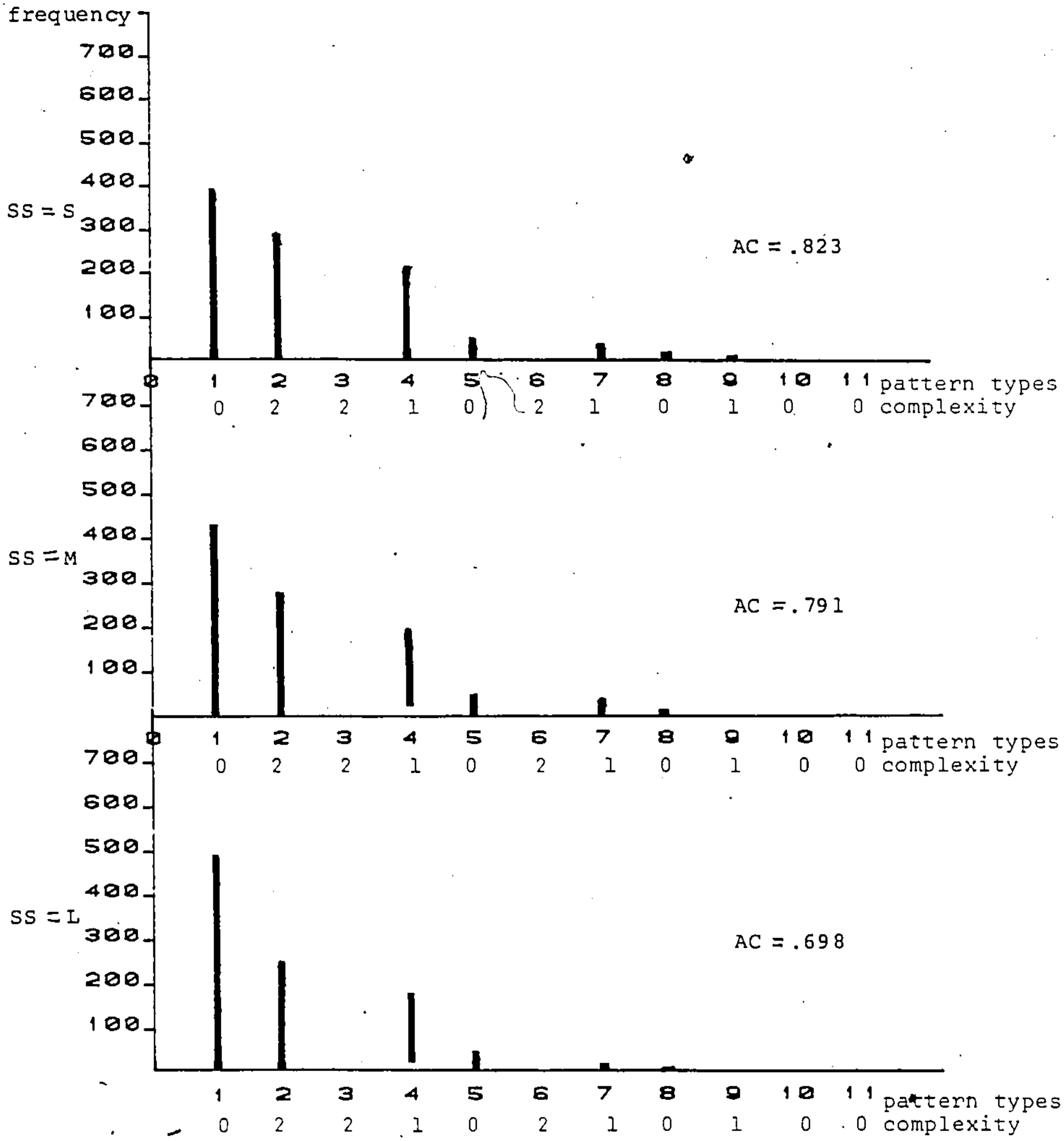


SS - sample size disparity, AC-average complexity, S-small, M-medium, L-large homogeneous variance condition (Fig. 20, 22) as sample size disparity increased, the frequency of pattern 1 increased, causing decreases in the frequencies of the other patterns. In the direct pairing condition, this trend was magnified. In the inverse pairing condition, it was reversed (Fig. 23). Frequency of pattern type 1 decreased as sample size disparity increased. This trend was consistently found across all the partitions for the Kramer method. Though changes in average complexity may indicate a shift in pattern frequencies, often it was unclear in which direction the shift was taking place, if pattern frequencies were not available.

Similar to its performance with other dependent variables, the Games and Howell technique did not exhibit changes in average complexity with changes in sample size disparity. A similar finding was evident upon examination of the pattern frequencies. Figures 24 and 25 display graphically the pattern frequencies under the varying sample size disparity levels for C = 0 and C = 1.0. The three sample size disparity conditions exhibit curves that were markedly similar in shape and size. Similar results were apparent in the inverse pairing condition.

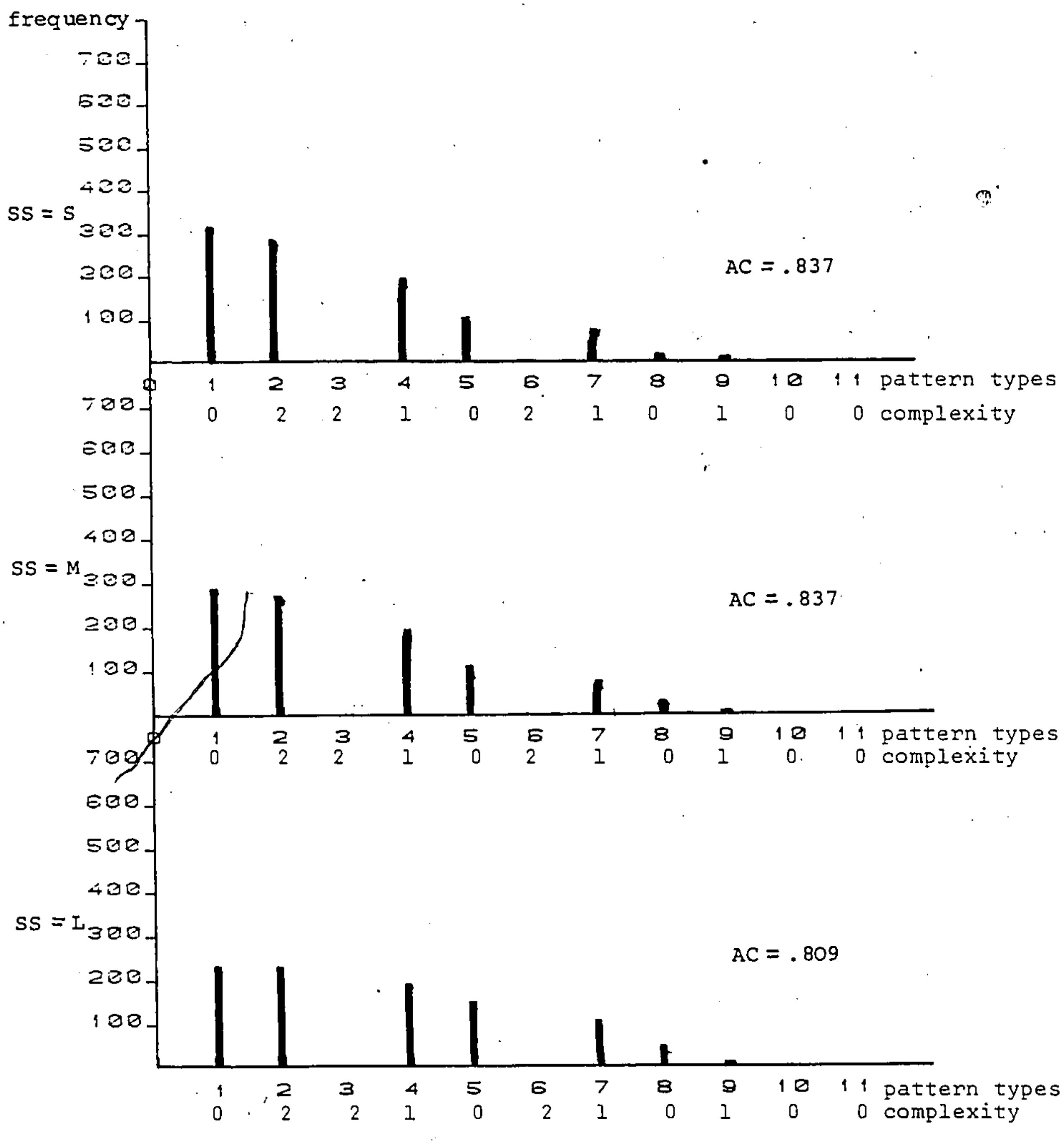
Variance heterogeneity increases did not produce a consistent trend in average complexity for the Kramer method in the direct pairing condition. Partitions I, II and V exhibited decreases in average complexity as variance heterogeneity increased. Partitions III and IV increased in average complexity. When explained by pattern type frequencies, the

Figure 22: Average complexity, Kramer method, partition IV, mean variability (medium), variance heterogeneity (C = 0), inverse pairing



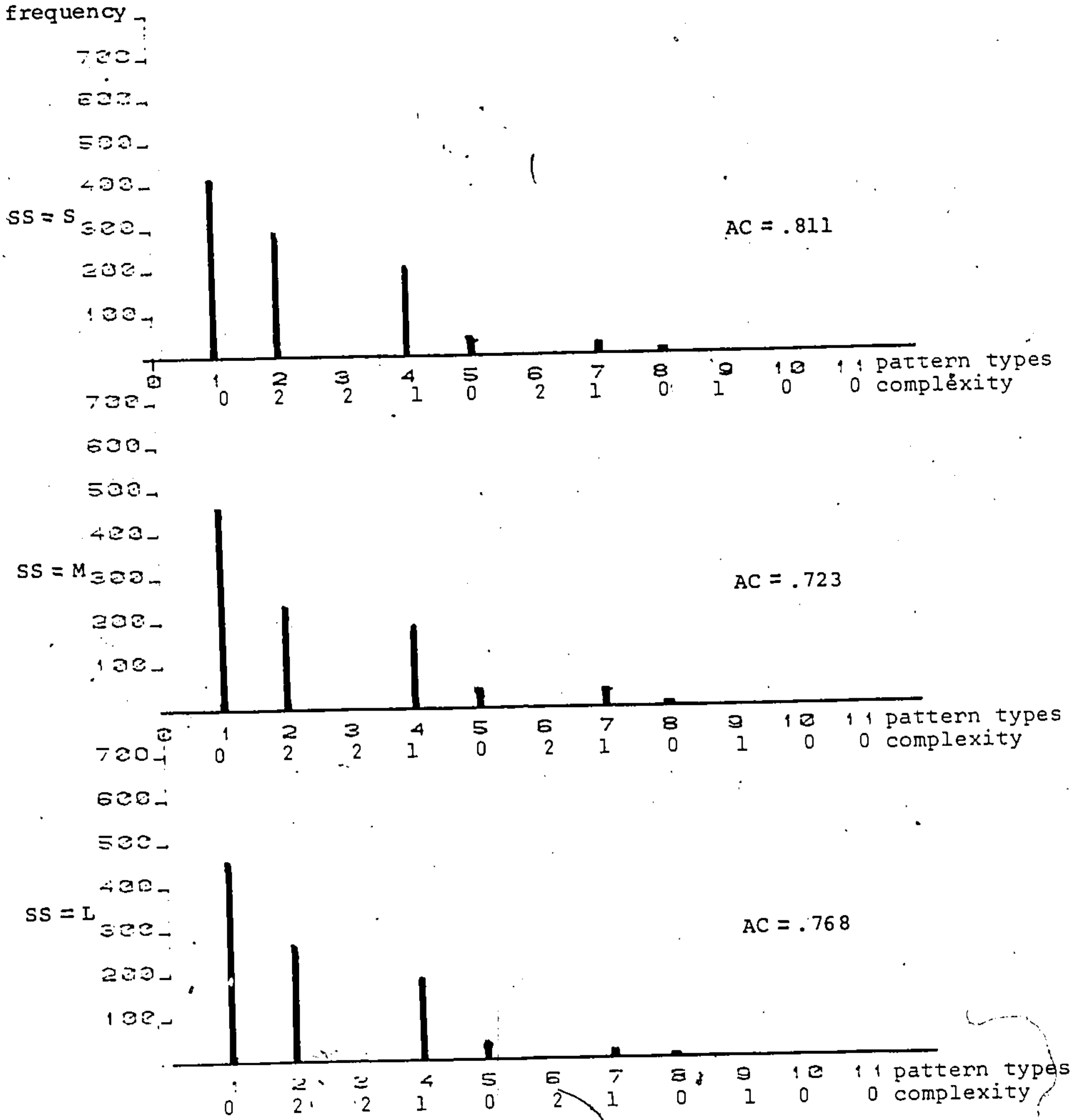
SS - sample size disparity, ,AC-average complexity,S-small,M-medium,L-large

Figure 23: Average complexity, Kramer method, partition IV, mean variability (medium), variance heterogeneity (C = 1.0), inverse pairing



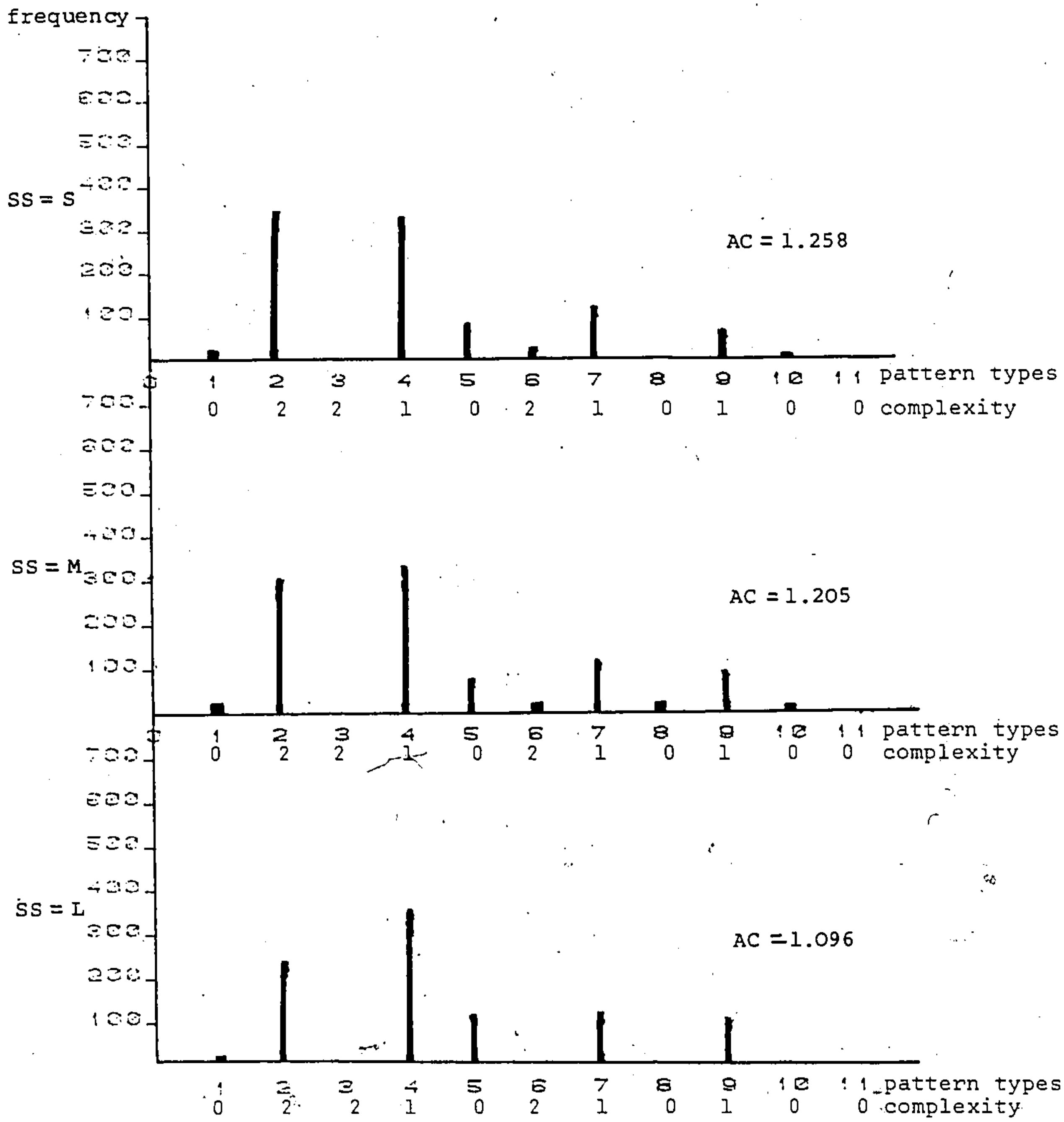
SS—sample size disparity,AC-average complexity,S-small,M-medium,L-large

Figure 24: Average complexity, Games and Howell method, partition IV, mean variability (medium), variance heterogeneity (C = 0), direct pairing



SS – sample size disparity ,AC-average complexity,S-small,M-medium,L-large

Figure 25: Average complexity, Games and Howell method, partition IV, mean variability (medium), variance hetergeneity (C = 1.0), direct pairing

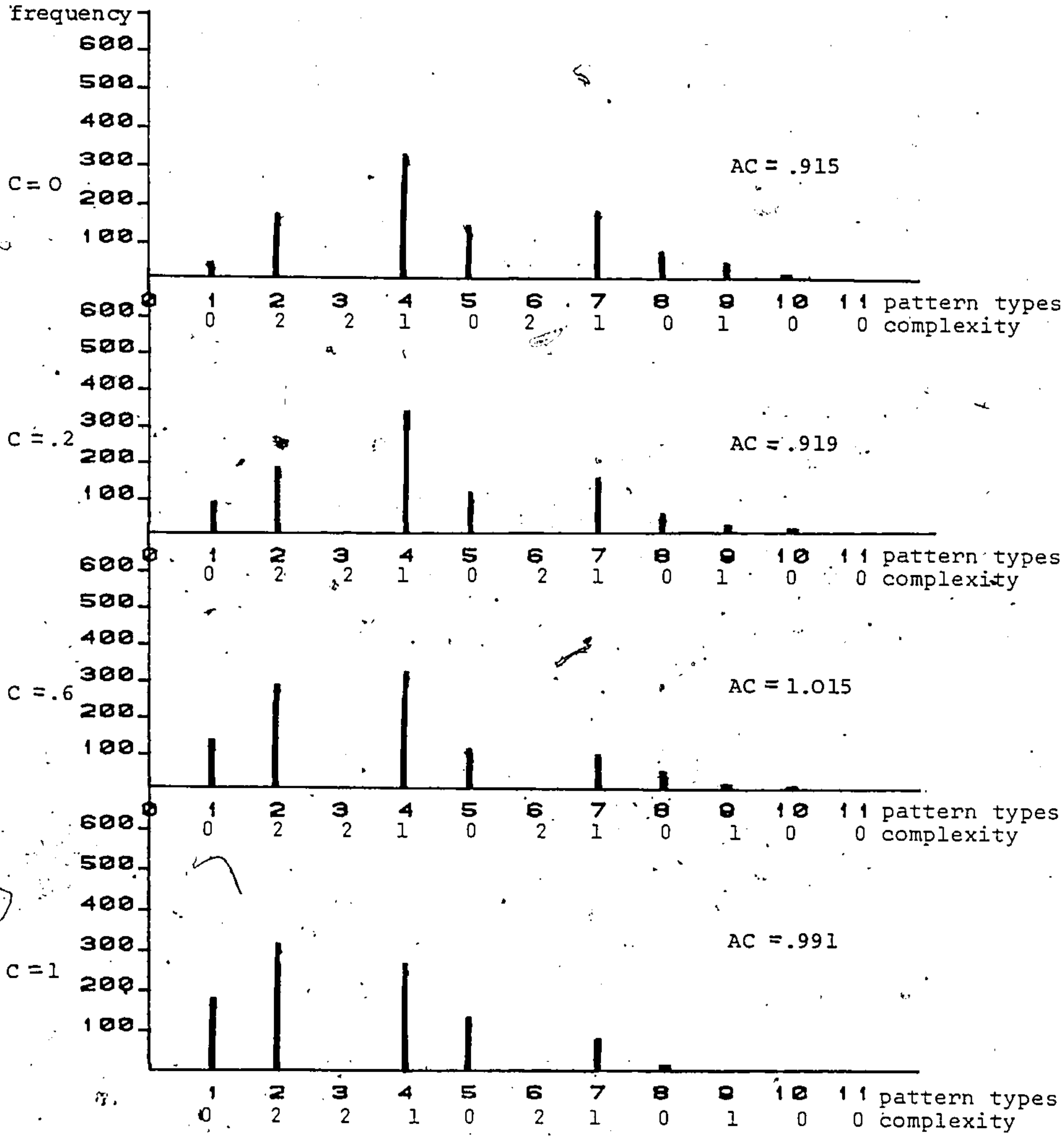


SS - sample size disparity ,AC-average complexity,S-small,M-medium,L-large trend became consistent with the other effects caused by the independent variables. Increased variance heterogeneity resulted in a decrease in the average number of rejections. This was reflected in a shift in the frequencies from the higher to lower order patterns (Fig. 26). When the result of this shift caused increased frequency of occurrence of patterns with large complexities (i.e., pattern 2 in partitions III and IV), average complexity tended to increase. When the major contributor to pattern frequencies was a type with a low complexity (i.e., pattern type 1 in partitions I, II and V), average complexity decreased.

In the inverse pairing condition, the shift was from lower order to higher order pattern types as variance heterogeneity increased (Fig. 27). Since most of the higher order patterns had lower complexities, a decrease in average complexity was observed.

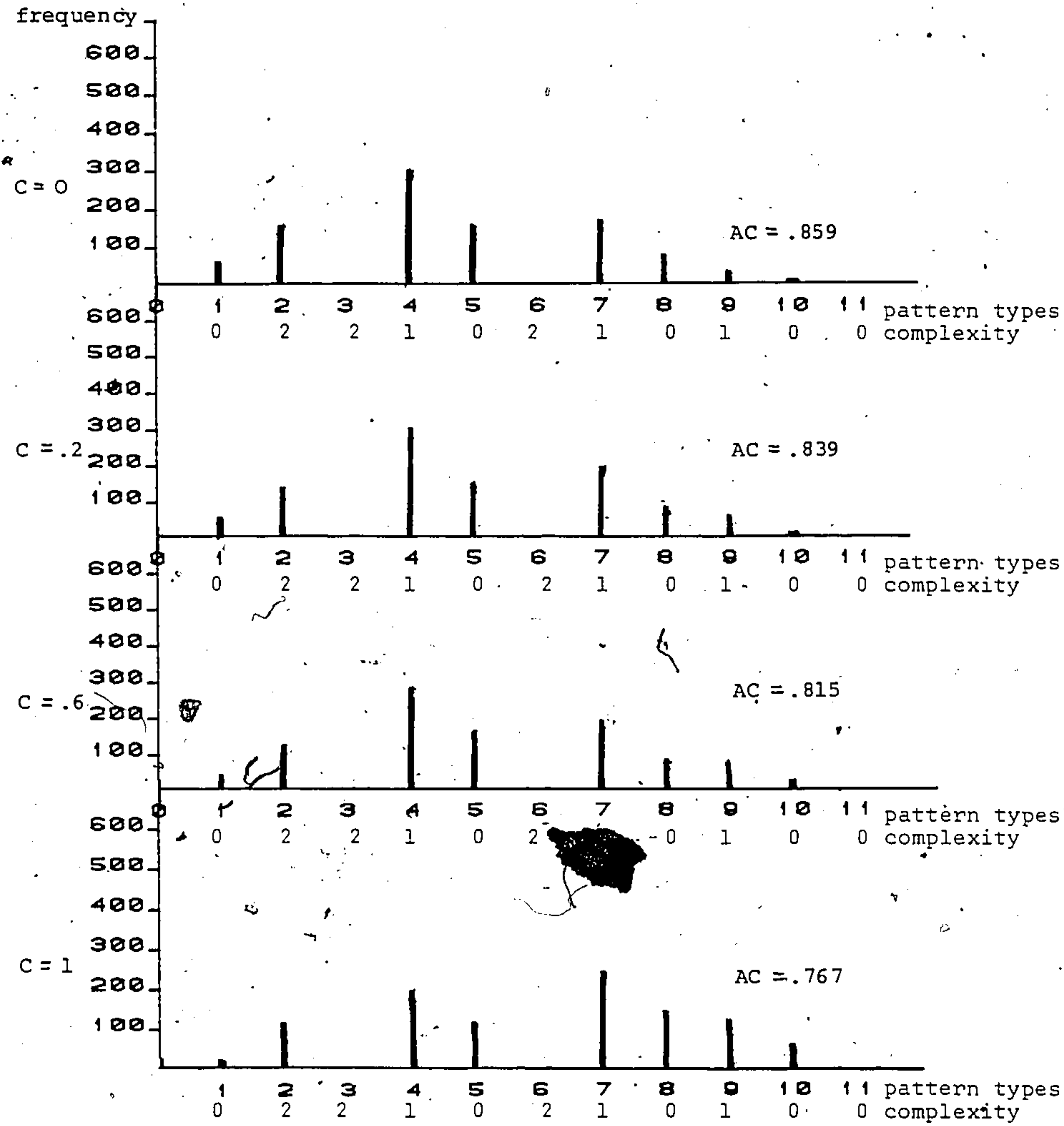
The Games and Howell technique experienced an effect in the opposite direction to that of the Kramer method, as indicated by the pattern type frequencies. In the direct pairing condition, a shift from lower order to higher order patterns occurred as variance heterogeneity increased (Fig. 28). Average complexity, in the case of partition IV, decreased but this was not necessarily the case for the other partitions. Partitions I, III and some conditions of partition IV had increased in average complexity as variance heterogeneity increased. In the inverse pairing condition, the shift was from higher order to lower order

Figure 26: Average complexity, Kramer method, partition IV, mean variability (large), sample size disparity (large), direct pairing



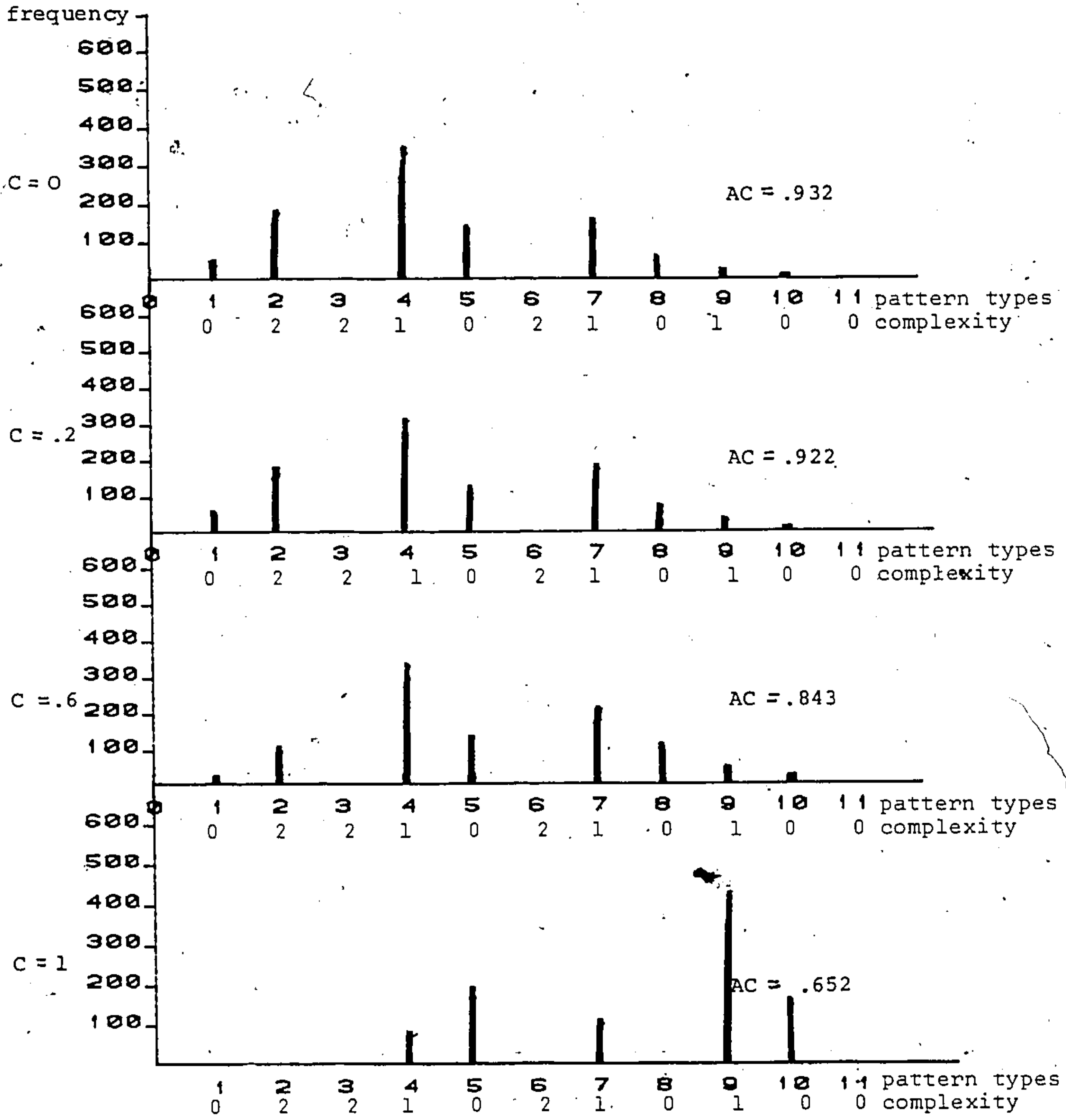
C - coefficient of variation, AC-average complexity,

Figure 27: Average complexity, Kramer method, partition IV, mean variability (large), sample size disparity (large), inverse pairing



C-coefficient of variation, AC-average complexity,

Figure 28: Average complexity, Games and Howell method, partition IV, mean variability (large), sample size disparity (large), direct pairing

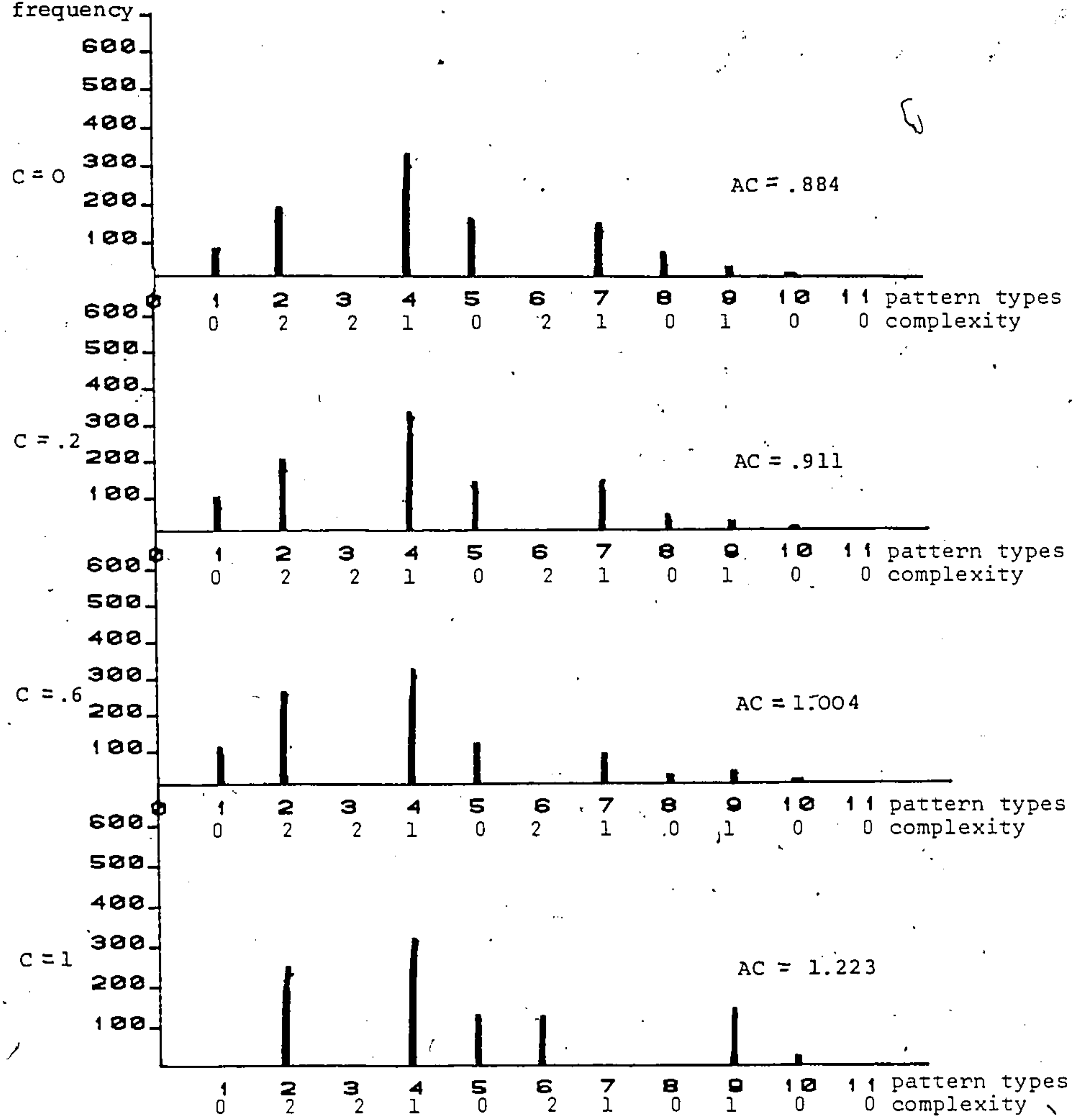


C-coefficient of variation ,AC-average complexity patterns as the divergence in variances increased. For partition IV, this resulted in an increase in average complexity (Fig. 29). Fig. 30 shows the frequency changes in pattern types for partition III, in which average complexity decreased as variance heterogeneity increased.

It is now apparent that when compared across the various conditions, average complexity is inconsistent. Without the information available from pattern frequencies, interpretation based on average complexity would be misleading. Yet, the concept of complexity, as perceived through pattern frequencies, held up remarkably well across all conditions. Changes in pattern frequency reflected changes in the independent variables and distinguished between the Kramer, and Games and Howell methods.

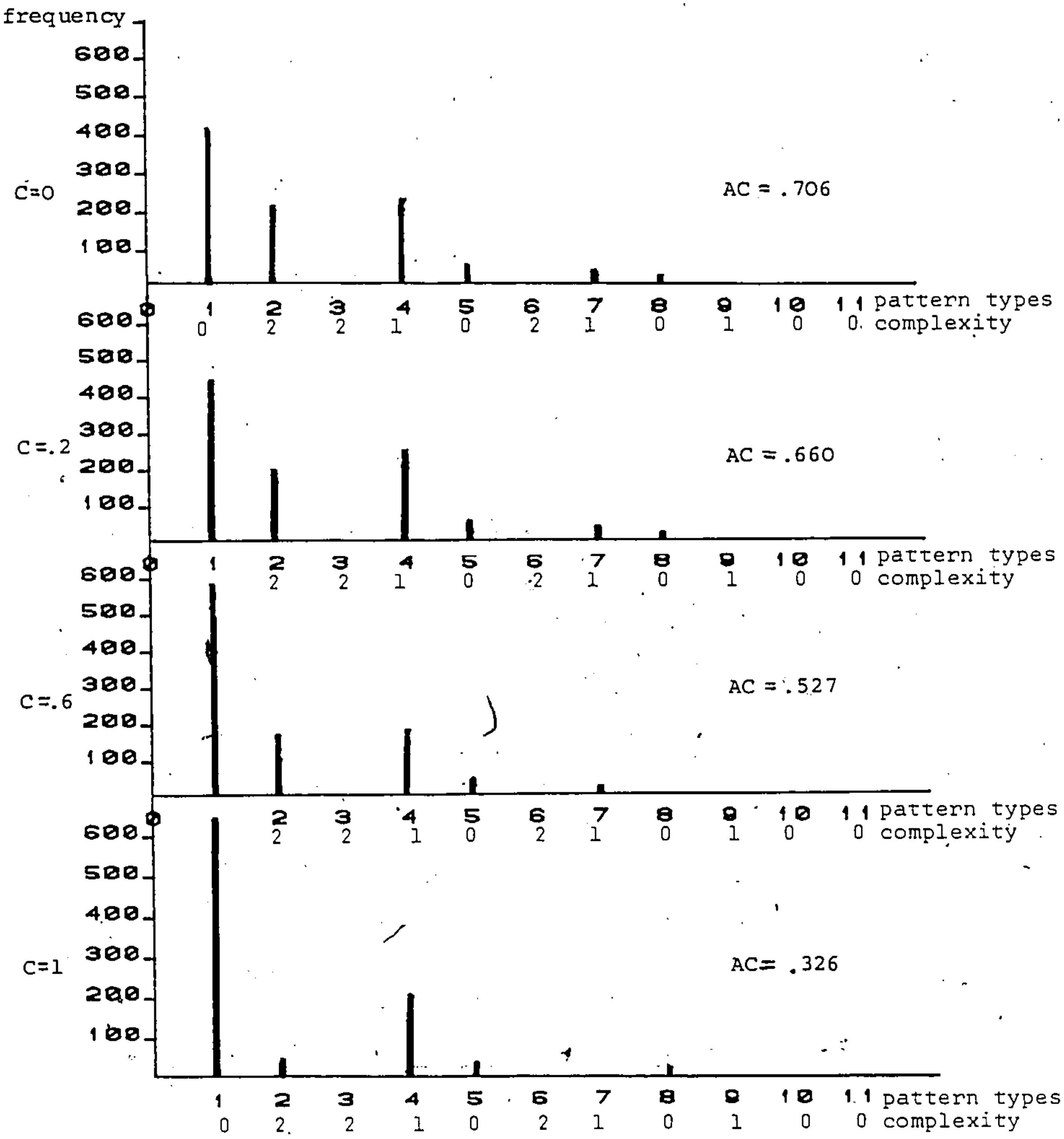
After observing the behaviour of complexity under a variety of conditions, the way in which the pattern type complexities were calculated can now be appreciated. Complexity decreases as the pattern types increase to a higher order. The implicit assumption would seem to be that in research, errors of omission (i.e., errors due to group differences found not to be significant) are more common than errors of inclusion. This assumption was verified in that lower order patterns always had greater frequencies than the higher order patterns even when the true simple pattern was of high order (i.e., Partition IV).

Figure 29: Average complexity, Games and Howell method, partition IV, mean variability (large), sample size disparity (large), inverse pairing



C – coefficient of variation ,AC-average complexity,

Figure 30: Average complexity, Games and Howell method, partition III, mean variability (large), sample size disparity (large), inverse pairing



C —coefficient of variation, AC-average complexity, Using a different method of categorization, the 64 possible patterns could be grouped as errors of omission or inclusion. Under this method, the true pattern would have a value of zero: those patterns with errors of omission would have negative values, while those patterns with errors of inclusion would have positive values. The size of the numerical value attributed to each pattern would be equal to the number of contrasts either omitted from or included above the true number of contrasts. This was done, and the results indicated grossly skewed distributions in favour of errors of omission. Errors of inclusion seldom occurred, and when they did, they evidenced very small frequencies. Shifts in frequencies of errors, similar to those occurring in the pattern types, were also apparent though not as clear.

In a similar manner, type II errors--rejecting the alternate when the alternate hypothesis is true--invariably occurred across all conditions. As such, type II errors as dependent variables, were unable to discriminate between the various conditions because they were so numerous.

Perhaps the most serious limitation of the present study can be seen in the frequencies of the various patterns. The frequency of occurrence of the complete pattern of true contrasts is very small. When taken individually, the identification of significant contrasts varied from extremely high to low. This may be related to the power of the tests used. It should be noted that the range of variability among means was relatively small in the present study. Shaffer (1981) employed a much larger range of constants, (.2 to 7.2), thus providing a larger frequency of occurrence of the true pattern. Keselman and Rogan (1978) employed a constant of .75 for their large mean variability condition. The largest constant employed in the present study was .357. As such, the generalizability of the present results are limited to situations involving a medium degree of mean variability.

A comparison of the number of correct rejections was made between the present study and that of Keselman and Rogan (1978). The average number of non-zero contrasts that were correctly identified was calculated as in Carmer and Swanson (1973). The trends were identical to those reported by Keselman and Rogan (1978) for their intermediate mean variability condition. However, slightly larger values were found in the present study since more discrepant means were used.

CHAPTER V

CONCLUSIONS

If complexity had replicated exactly the same trends as demonstrated by experiment-wise error rates, it would be a simple matter to draw conclusions about its usefulness as a criterion for the evaluation of multiple comparison procedures. It would also be redundant and of little use beyond novelty. On the other hand, if complexity contains totally different information from that of experiment-wise error rates, the opposite problem presents itself: finding a ground common-enough, for comparison. This appears to be the case of Shaffer's concept of complexity.

Consider complexity in its rawest form, the frequencies of the various pattern types. In this form, complexity is related directly to the average number of rejections, and reflects changes in sample size disparity, variance heterogeneity, direct and inverse pairing, mean variability and different partitions. As such, complexity provides information about patterns of results rather than simply the frequency of experiment-wise errors. It is also able to discriminate between the behaviour of the Kramer (1956) method and the Games and Howell (1976) method under the conditions cited above. Why some patterns acquire higher frequencies than others, or why patterns such as pattern type 3 seldom occur, become questions of some interest.

The latter was concerned with three subpatterns, each contaning two pair-wise contrasts [((12) (13)), ((13) (24)), ((14) (23))]. These pairs of contrasts appeared together under some conditions but always with other contrasts, thus changing the order of the pattern type. That the three pairs never appeared as a unique pattern of type 3 can be attributed to the partitions which were selected for study: if any of the three pairs occurred as significant estimated contrasts, then the nature of the partition dictated that other estimated contrasts also be significant.

As a tool rendering information to the researcher, the frequency of pattern types yielded information that was hitherto unavailable and not included within experiment-wise error rates. Frequency of pattern types provides a pictorial display which shows the change in patterns of results, as various conditions affecting the statistical test change. Unfortunately, this approach is unwieldy and difficult to tabulate --hence, the formation of average complexity. It is important, at this point, to distinguish between the usefulness of average complexity and complexity as a concept.

Average complexity was calculated in such a way as to measure the distance from a simple pattern. Simple patterns consequently had a complexity of zero, while other patterns' complexities varied depending on how far they were from a simple pattern. As such, complexity was defined as the number of non-significant contrasts which must become significant in order to change the pattern to a simple pattern. From the preceding discussion, it became apparent that the usefulness of average complexity is limited because of the way it is measured. The different complexities of the various pattern types cause inconsistent shifts in average complexity as the frequency of various patterns increase or decrease. The inconsistency disappears once pattern frequency is observed, yet without this information, the behaviour of average complexity becomes misleading. In order to clarify this situation, consider the relationship amongst average complexity, experiment-wise error rate and average number of rejections.

A decrease in average complexity indicates that the relative frequency of patterns with lower complexities (hopefully zero, i.e., simple patterns) has increased. The researcher would then wish to see low average complexities, something which would infer the possibility of more interpretable patterns. In terms of the average number of rejections, the greater the number of rejections, the greater the information provided to the researcher, and subsequently, the greater the chance of achieving a simple pattern. In conclusion, the educational researcher would prefer low experiment-wise error rates, high average number of rejections and low average complexity as characteristics of a good statistical method.

Table 21 lists values of the three dependent variables for both the Kramer and Games and Howell methods from a variety of conditions which were studied. The purpose of the table is to examine the three parameters together to see if they yield consistent interpretations, and to see if they discriminate between the two statistical methods under the various conditions.

TABLE 21

COMPARISON OF EXPERIMENT-WISE ERROR RATE, AVERAGE NUMBER OF REJECTIONS AND AVERAGE COMPLEXITY FOR THE KRAMER AND GAMES AND HOWELL METHODS

VARIABLES						KRAMER			GAMES & HOWELL		
Row	Partition	Pairing	Mean Variability	Sample Size Disparity	Variance Heterogeneity	ER	AR	AC	ER	AR	AC
1	I	D	S	S	1.0	.060	.029	.110	.029	.257	.411
2		D	S	M	0	.029	.071	.092	.033	.072	.099
3		D	M	L	1.0	.043	.047	.082	.040	1.055	1.680
4		D	L	S	1.0	.057	.242	.243	.021	1.732	1.732
5		D	L	L	0	.022	.287	.254	.016	.259	.253
6	II	I	S	S	0	.019	.073	.110	.021	.076	.104
7		I	S	S	1.0	.105	.275	.185	.027	.089	.084
8		I	L	L	0	.017	.435	.428	.023	.405	.406
9		I	L	L	1.0	.150	1.212	.401	.021	.318	.191
10	III	I	S	S	1.0	0	.288	.177	.004	.082	.074
11		I	L	S	1.0	0	1.793	.451	.005	1.053	.487
12		I	L	L	1.0	0	2.141	.403	.019	.710	.326
13	V	D		L	1.0	.026	.027	.051	.046	.067	.071
14		D		S	0	.050	.063	.087	.054	.066	.097
15		I		M	1.0	.160	.319	.180	.043	.028	.051
16		I		L	1.0	.222	.431	.243	.027	.045	.036

ER = EXPERIMENT-WISE ERROR RATES
AR = AVERAGE NUMBER OF REJECTIONS
AC = AVERAGE COMPLEXITY

The specific conditions were chosen on the basis of the experiment-wise error rates and as a consequence of this choice, cover the range from very high to very low error rates.

In some instances, average complexity was highly consistent with experiment-wise error rates and average number of rejections (rows 2, 5, 6, 7). In these cases, increased error rates coincided with increased average complexities. This was, however, not always true. Rows 1 and 9 show high error rates for the Kramer method as opposed to the Games and Howell method, yet the latter has a higher complexity.

Even though experiment-wise error rates may be controlled, average rejections may still increase, and this is the reason for inconsistency in the behaviour of average complexity. An increase in average rejections would indicate a shift towards higher order patterns, and thus complexity would fluctuate depending upon which patterns increased in frequency of occurrence. This can be seen in Rows 10, 11, 12, where very low error rates are evident dispite a variety of average complexities.

It would seem that average complexity, by itself, would be of little use to the researcher and in some instances may even be misleading. Due to the method in which it is calculated, the information contained in pattern frequencies and hence the frequency of various pair-wise contrasts is lost. The problems inherent in the interpretation of average complexity may be offset slightly if information regarding the number of rejections is available. Given both pieces of information and assuming that experiment-wise error rates are the same for two techniques, several options are available.

If the average rejections are the same for both techniques, then the technique with the lower average number of rejections is to be preferred since it indicates a smaller distance from a simple pattern. If the average complexities are the same, then the technique with the greater number of rejections would provide more information to the researcher. If one technique has both average number of rejections and average complexity greater than the other, then the researcher must choose either increase of information (i.e., more rejections), or an interpretable pattern, as having a higher priority. The situation becomes more complicated if type I error rates are also allowed to vary. The educational researcher must assume responsibility to align his priorities with the various criteria.

In conclusion, the concept of complexity provides a unique set of information which is not contained within experiment-wise error rates. As such, it opens a new dimension for the examination of the behaviour of multiple comparison procedures and allows for the development of a more complete picture of a statistical technique's performance. As a consequence, complexity should not be ignored and set aside. The concept is valid, and based upon pattern frequencies, demonstrates an ability to discriminate between statistical techniques under a variety of conditions. The difficulty arises when the information in the pattern frequencies is transformed into average complexity, for inconsistencies arise when comparing average complexity's interpretation with that of other criteria.

The present study was able to illustrate changes in the frequencies of significant estimated pair-wise contrasts and therefore the changes in frequencies of various patterns of significant differences under a variety of conditions. Knowing that the concept of complexity offers a new source of information to the researcher, future studies might wish to examine a more precise way of representing this information than the us of average complexity (Shaffer 1981). Other conditions such as the equal sample size and equal variance case should also be investigated. Since one of the the major limitations of the present study was the low power, future research should include intermean distances at least as great as Keselman and Rogan (1978) and perhaps as large as those used by Shaffer (1981). This would create the occurrence of larger frequencies of the true patterns; thus allowing the researcher to view a more complete range of pattern frequencies. It would also be interesting to examine complexity with respect to other statistical tests than the modifications of the Tukey techniques which were examined.

BIBLIOGRAPHY

Balaam, L.N., and Federer, W.T. (1965), "Ouery II: error rate bases," Technometrics, 7, 260-262.

Behrens, W.U. (1964), "The comparison of means of independent normal distributions with different variance," Biometrics, 20, 16-27.

Best, D.J., and Roberts, D.E. (1975), "The percentage points of the χ^2 distribution, algorithm AS 91," Applied Statistics, 24, 385-388.

Brown, R.A. (1974), "Robustness of the studentized range statistic," Biometrika, 61, 171-175.

Box, G.E.P. (1954), "Some theorems on quadratic forms applied in the study of analysis of variance problems, I Effect of inequality of variances in the one-way classification," Annals of Mathematical Statistics, 25, 290-302.

Box, G.E.P., and Muller, M. (1958), "A note on the generation of random normal deviates," Annals of Mathematical Statistics, 29, 610-611.

Carmer, S.G., and Swanson, M.R. (1973), "An evaluation of ten multiple comparison procedures by monte carlo methods," Journal of the American Statistical Association, 68, 66-74.

Cochran, W.G. (1964), "Approximate significance levels of the Behrens-Fisher test," Biometrics, 20, 191-195.

Cohen, J. (1969), Statistical Power Analysis for the Behavioral Sciences, New York: Academic Press.

Cox, D.R. (1965), "A remark on multiple comparison methods," Technometrics, 7, 223-224.

Dunn, O.J. (1958), "Estimation of the means of dependent variables," Annals of Mathematical Statistics, 29, 1095-1111.

Dunn, O.J. (1959), "Confidence intervals for the means of dependent normally distributed variables," Journal of the American Statistical Association, 54, 613-621.

Dunn, O.J. (1961), "Multiple comparisons among means," Journal of the American Statistical Association, 56, 52-64.

Dunn, O.J. (1974), "On multiple tests and confidence intervals," Communications in Statistics, 3, 101-103.

Dunn, O.J., and Massey, F.J. Jr. (1965), "Estimation of multiple contrasts using t-distributions," Journal of the American Statistical Association, 60, 573-583.

Dunnett, C.W. (1980a), "Pairwise multiple comparisons in the homogenous variance, unequal sample size case," Journal of the American Statistical Association, 75, 789-795.

Dunnett, C.W. (1980b), "Pairwise multiple comparisons in the unequal variance case," Journal of the American Statistical Association, 75, 796-800.

Einot, I., and Gabriel, K.R. (1975), "A study of the powers of several methods of multiple comparisons," Journal of the American Statistical Association, 70, 574-583.

Gabriel, K.R. (1978), "A simple method of multiple comparison," Journal of the American Statistical Association, 73, 724-729.

Games, P.A. (1971a), "Multiple comparisons of means," American Educational Research Journal, 8, 531-565.

Games, P.A. (1971b), "Inverse relation between the risks of type I and type II error and suggestions for the unequal n case in multiple comparisons," Psychological Bulletin, 75, 97-107.

Games, P.A. (1973), "Type IV errors revisited," Psychological Bulletin, 80, 304-307.

Games, P.A., and Howell, J.F. (1976), "Pairwise multiple comparison procedures with unequal n's and/or variances: a monte carlo study," Journal of Educational Statistics, 1, 113-125.

Genizi, A., and Hochberg, Y. (1978), "On improved extension of the T-method of multiple comparisons for unbalanced designs," Journal of the American Statistical Association, 73, 879-884.

Glass, G.V., and Stanley, J.C. (1970), Statistical Methods in Education and Psychology, Englewood Cliffs, New York: Prentice-Hall.

Greenwood, A.J. (1974), "A fast generator for gamma-distributed random variables," In Compstat 1974, eds, G. Bruckmann, F. Ferschl and L. Schmetterer, Vienna: Physica Verlag, 19-27.

Harter, L.H. (1960), "Tables of range and studentized range," Annals of Mathematical Statistics, 31, 1122-1147.

Hartley, H.O. (1955), "Some recent developments in analysis of variance," Communications on Pure and Applied Mathematics, 8, 47-72.

Hays, L.W. and Winkler, R.L. (1971), Statistics Probability, Inference and Decision, New York, Holt, Rinehart and Winston, Inc.

Hochberg, Y. (1974), "Some generalizations of the T-method in simultaneous inference," *Journal of Multivariate Analysis*, 4, 224-234.

Hochberg, Y. (1975), "An extension of the T-method to general unbalanced models of fixed effects," *Journal of the Royal Statistical Society, Series B*, 37, 426-433.

Hochberg, Y. (1976), "A modification of the T-method of multiple comparisons for a one-way layout with unequal variances," *Journal of the American Statistical Association*, 71, 200-203.

Howell, J.F., and Games, P.A. (1973), "The robustness of the analysis of variance and the Tukey WSD test under various patterns of heterogeneous variances," *Journal of Experimental Education*, 4, 33-37.

Howell, J.F., and Games, P.A. (1974), "The effects of variance heterogeneity on simultaneous multiple comparison procedures with equal sample size," *British Journal of Mathematical and Statistical Psychology*, 27, 72-81.

Hummel, T.J., and Sligo, J.R. (1971), "Empirical comparison of univariate and multivariate analysis of variance procedure," *Psychological Bulletin*, 76, 49-57.

IMSL Library Reference Manual, edition 9, vol. 2 (revised June 1982), Houston: IMSL Inc.

Keselman, H.J. (1976), "A power investigation of the Tukey multiple comparison statistic," *Educational and Psychological Measurement*, 36, 97-104.

Keselman, H.J. and Murray, R. (1974), "Tukey tests for pairwise contrasts following the analysis of variance: is there a type IV error," *Psychological Bulletin*, 81, 608-609.

Keselman, H.J., and Rogan, J.C. (1977), "The Tukey multiple comparison test: 1953-1976," *Psychological Bulletin*, 84, 1050-1056.

Keselman, H.J., and Rogan, J.C. (1978), "A comparison of the modified Tukey and Scheffé methods of multiple comparisons for pairwise contrasts," *Journal of the American Statistical Association*, 73, 47-52.

Keselman, H.J., and Toothaker, L.E. (1973), "Error rates for multiple comparison methods: some evidence concerning the misleading conclusions of Petrinovich and Hardyck," *Psychological Bulletin*, 80, 31-32.

Keselman, H.J., and Toothaker, L.E. (1974), "Comparison of Tukey's T-method and Scheffé's S-method for various numbers of all possible differences of averages contrasts under violation of assumptions," Educational and Psychological Measurement, 34, 511-519.

Kesselman, H.J., Games, P.A., and Rogan, J.C. (1979), "An addendum to a comparison of the modified Tukey and Scheffé methods of multiple comparisons for pairwise contrasts," Journal of the American Statistical Association, 74, 626-627.

Keselman, H.J., Murray, R., and Rogan, J. (1976), "Effect of very unequal group sizes on Tukey's multiple comparison test," Educational and Psychological Measurement, 36, 263-270.

Keslman, H.J., Toothaker, L.E., and Shooter, M. (1975), "An evaluation of two unequal n forms of the Tukey multiple comparison statistic," Journal of the American Statistical Association, 70, 584-587.

Kirk, R.E. (1968), Experimental Design: Procedures for the Behavioral Sciences, Belmont, California: Brooks/Cole Publishing Co.

Kirk, R.E. (1982), Experimental Design: Procedures for the Behavioral Sciences, second edition, Belmont, California: Brooks/Cole Publishing Co.

Kramer, C.Y. (1956), "Extension of multiple range tests to group means with unequal numbers of replications," Biometrics, 12, 307-310.

Kramer, C.Y. (1957), "Extension of multiple range tests to group correlated adjusted means," Biometrics, 13, 13-18.

Lee, A.F.S., and Gurland, J. (1975), "Size and power of tests for equality of means of two normal populations with unequal variances," Journal of the American Statistical Association, 70, 933-941.

Lehmann, E.L., and Shaffer, J.P. (1977), "On a fundamental theorem in multiple comparisons," Journal of the American Statistical Association, 72, 576-578.

Lehmann, E.L., and Shaffer, J.P. (1979), "Optimum significance levels for multistage comparison procedures," Annals of Statistics, 7, 27-45.

Levin, J.R., and Marascuilo, L.A. (1972), "Type IV errors and interactions," Psychological Bulletin, 78, 368-374.

Levin, J.R., and Marascuilo, L.A. (1973), "Type IV errors and Games," Psychological Bulletin, 80, 308-309.

Marsaglia, G., and Bray, T.A. (1964), "A convenient method for generating normal variables," SIAM Review, 6, 260-264.

Mehta, J.S., and Scrinivasan, R. (1970), "On the Behrens-Fisher problem," Biometrika, 57, 649-655.

Miller, R.G. (1966), Simultaneous Statistical Inference, New York: McGraw Hill.

Miller, R.G. (1977), "Developments in multiple comparisons 1966-1976," Journal of the American Statistical Association, 72, 779-788.

Morrison, D.F. (1976), Multivariate Statistical Methods, second edition, New York: McGraw-Hill Book Co.

O'Neill, R., and Wetherill, G.B. (1971), "The present state of multiple comparison methods (with discussion)," Journal of the Royal Statistical Society, Series B, 33(2), 218-250.

Petrinovich, L.F., and Hardyck, C. (1969), "Error rates for multiple comparison methods," Psychological Bulletin, 71, 43-54.

Ramsey, P.H. (1978), "Power differences between pairwise multiple comparisons," Journal of the American Statistical Association, 73, 479-487.

Ramsey, P.H. (1980), "Choosing the most powerful pairwise multiple comparison procedure in multivariate analysis of variance," Journal of Applied Psychology, 65, 317-326.

Ramseyer, G.C., and Tcheng, T.K. (1973), "The robustness of the studentized range statistic to violations of the normality and homogeneity of variance assumptions," American Educational Research Journal, 10, 235-240.

Ringland, J.T. (1983), "Robust multiple comparisons," Journal of the American Statistical Association, 78, 145-151.

Rogan, J.C., Keselman, H.J., and Breen, L.J. (1977), "Assumption violations and rates of the type I error for Tukey multiple comparison test: a review and empirical investigation via a coefficient of variance variation," Journal of Experimental Education, 46, 20-25.

Ryan, T.A. (1959), "Multiple comparisons in psychological research," Psychological Bulletin, 56, 26-47.

Satterthwaite, F.E. (1946), "An approximate distribution of estimates of variance components," Biometrics, 2, 110-114.

Scheffé, H. (1953), "A method for judging all contrasts in the analysis of variance," Biometrika, 40, 87-104.

Scheffé, H. (1959), The Analysis of Variance, New York: John Wiley.

Shaffer, J.P. (1981), "Complexity: an interpretability criterion for multiple comparisons," Journal of the American Statistical Association, 76, 395-401.

Sidak, Z. (1967), "Rectangular confidence regions for the means of multivariate normal distributions," Journal of the American Statistical Association, 62, 626-633.

Smith, R.A. (1971), "The effect of unequal group size on Tukey's HSD procedure," Psychometrika, 36, 31-34.

Spjotvoll, E., and Stoline, M.R. (1973), "An extension of the T-method of multiple comparisons to include cases with unequal sample sizes," Journal of the American Statistical Association, 68, 975-978.

Steel, R.G.D. (1961), "Querry 163: error rates in multiple comparisons," Biometrics, 17, 326-328.

Stoline, M.R. (1978), "Tables of the studentized augmented range and applications to problems of multiple comparisons," Journal of the American Statistical Association, 73, 656-660.

Stoline, M.R. (1981), "The status of multiple comparisons: simultaneous estimation of all pairwise comparisons in one-way ANOVA designs," The American Statistician, 35, 134-141.

Stoline, M.R., and Ury, H.K. (1979), "Tables of the studentized maximum modulus distribution and an application to multiple comparison among means," Technometrics, 21, 87-93.

Sukhatme, P.V. (1938), "On Fisher and Behrens' test of significance for the difference in means of two normal samples," Sankhya, 4, 39-48.

Swaminathan, H., and Gifford, J. (1978), "Procedure for the comparison of means: an alternative to analysis of variance," Paper presented at the annual meeting of the American Educational Research Association, Toronto.

Tamhane, A.C., Multiple comparisons in mode I one-way ANOVA with unequal variances," Communications in Statistics, A6, 15-32.

Tamhane, A.C. (1979), "A comparison of procedures for multiple comparisons of means with unequal variances," Journal of the American Statistical Association, 74, 471-480.

Timm, N.H. (1975), Multivariate Analysis With Applications in Education and Psychology, Belmont, California: Wadsworth Publishing Company, Inc.

Ury, H.K. (1976), "A comparison of four procedures for multiple comparisons among means (pairwise contrasts) for arbitrary sample sizes," *Technometrics*, 18, 89-97.

Ury, H.K., and Wiggins, A.D. (1975), "A comparison of three procedures for multiple comparisons among means," *British Journal of Mathematical and Statistical Psychology*, 28, 88-102.

Wang, Y.Y. (1971), "Probabilities of the type I errors of the tests for the Brehens-Fisher problem," *Journal of the American Statistical Association*, 66, 605-608.

Welch, B.L. (1938), "The significance of the difference between two means when the population variances are unequal," *Biometrika*, 29, 350-362.

Welch, B.L. (1947), "The generalization of Students' problem when several different population variances are involved," *Biometrika*, 34, 28-35.

Welsch, R.E. (1977), "Stepwise multiple comparison procedures," *Journal of the American Statistical Association*, 72, 566-575.

Winer, B.J. (1962), *Statistical Principles in Experimental Design*, New York: McGraw-Hill Book Co.

Winer, B.J. (1971), *Statistical Principles in Experimental Design*, second edition, New York: McGraw-Hill Book Co.

APPENDIX I

COMPUTER PROGRAM

```
//DADTA    JOB    DADTAHIG,'HURDJ',CLASS=N,
//      MSGLEVEL=(1,1)
/*SETUP       PLEASE MOUNT TAPE USC219  RING IN.
/*MESSAGE                    MAY RUN FOR 1  HOUR. THANKS
/*MESSAGE
//      EXEC IMSLS,TIME.GO=100
//FORT.SYSIN DD *
      INTEGER N(3,4),SIG(8,6),FREQ(8,11),PCNTRS(8,12,11),NCCNTR,
     &        PAT,SUMSIG,OCCURS(8,6),PAGE,HCNT,NOCOND
      REAL SAMP(48,4),COMP(11),REJ(11),VARHET(4,4,2),Z1(5),
     &     Z2(5),Y1(5),Y2(5),Y3(5),SCORE(48,4),ERROR1(8),ERROR2(8),
     &        HOLD(124),DELTA(3,4)
      COMPLEX*16  TITLE(8)
      DOUBLE PRECISION SEED
      COMMON XM1,B1,XM2,B2,XM3,B3
      DATA N/26,21,17,27,26,23,34,35,36,37,42,48/,
     &     VARHET/1.0,0.73,0.30,.01,1.0,.92,.80,.06,1.0,1.08,1.03,1.58,
     &     1.0,1.27,2.0,2.35,1.0,1.27,2.0,2.35,1.0,1.08,1.03,1.58,
     &     1.0,.92,.8,.06,1.0,.73,.30,.01/,
     &     Z1/24.0,30.0,40.0,60.0,120.0/,Z2/20.,24.,30.,40.,60./
      DATA Y1/3.961,3.845,3.791,3.737,3.685/,Y2/2.851,2.805,2.766,2.716,
     &        2.673/,Y3/3.958,3.961,3.845,3.791,3.737/,
     &     SZ1,SZ2,SY1,SY2,SY3,SSZ1,SSZ2,SZ1Y1,SZ1Y2,SZ2Y3/10*0./,
     &     PAGE,NOCOND/0,0/
      DATA   DELTA/.089333333,.223333333,.357333333,
     &              .089333333,.223333333,.357333333,
     &              .178666666,.446666666,.714666666,
     &              .268,.67,1.072/
      DATA   COMP/0.,2.,2.,1.,0.,2.,1.,0.,1.,0.,0./
      DATA   REJ/0.,1.,2.,2.,3.,3.,3.,4.,4.,5.,6./
      DATA   SEED/537424784D0/
      DATA   TITLE/'KRAMMER','SPJOTROLL,STOLIN','HOCHBERG(1974)',
     &       'GABRIEL','HOCHBERG(1976)','GAMES AND HOWELL',
     &       'T3 PROCEDURE','C PROCEDURE'/
C       CALCULATE CONSTANTS FOR STATISTICS 6,7,8(GAMES + HOWELL,T3,C)
      DO 1 K=1,5
        Z1(K)=1/Z1(K)
        Z2(K)=1/Z2(K)
        SZ1=SZ1+Z1(K)
        SZ2=SZ2+Z2(K)
        SY1=SY1+Y1(K)
        SY2=SY2+Y2(K)
        SY3=SY3+Y3(K)
        SZ1Y1=SZ1Y1+Z1(K)*Y1(K)
        SZ1Y2=SZ1Y2+Z1(K)*Y2(K)
        SZ2Y3=SZ2Y3+Z2(K)*Y3(K)
        SSZ1=SSZ1+Z1(K)**2
1       SSZ2=SSZ2+Z2(K)**2
      XM1 = (5.*SZ1Y1-SZ1*SY1)/(5.*SSZ1-SZ1**2)
      XM2 = (5.*SZ1Y2-SZ1*SY2)/(5.*SSZ1-SZ1**2)
      XM3 = (5.*SZ2Y3-SZ2*SY3)/(5.*SSZ2-SZ2**2)
      B1  = SY1/5.-XM1*SZ1/5.
      B2  = SY2/5.-XM2*SZ1/5.
      B3  = SY3/5.-XM3*SZ2/5.
C
      DO 90 I=1,4
        DO 90 J=1,4
          DO 90 K=1,2
90          VARHET(I,J,K) = SQRT(VARHET(I,J,K))
C
C     SET UP LOOPS FOR DIFFERENT CONDITIONS
      DO 900 NPAR=1,5
      DO 900 NPAIR=1,2
      DO 900 NMEAN=1,3
        IF ((NPAR .EQ. 5) .AND. (NMEAN .NE. 1)) GO TO 900
      DO 900 NDISP=1,3
      DO 900 NHET=1,4
C
C     INITIALIZE PATTERN FREQUENCY COUNTER.
      NOCOND = NOCOND + 1
      DO 2 K=1,8
        ERROR1(K) = 0
```

```
          ERROR2(K) = 0
          DO 3 J=1,6
3            OCCURS(K,J)=0
          DO 4 I=1,11
               DO 4 J=1,12
4              PCNTRS(K,J,I)=0
        DO 2 J=1,11
2            FREQ(K,J)=0
C
C       DO A SET OF 1000   SIMULATIONS
          DO 500 NSIM=1,1000
             CALL GGNPM(SEED,124,HOLD)
             HCNT = 1
             DO 120 NPOP=1,4
                ITEMP=N(NDISP,NPOP)
                DO 120 NN=1,ITEMP
C       FOR EACH SCORE IN ONE PARTICULAR PARTITION
                   SCORE(NN,NPOP)=HOLD(HCNT)*VARHET(NHET,NPOP,NPAIR)
                   HCNT = HCNT + 1
                   IF ((NPOP .EQ. 1) .OR. (NPAR .EQ. 5))  GO TO 120
C                  ADD APPROPRIATE DELTA TO SCORES
                   GO TO (131,132,133,134),NPAR
C       PARTITION 1
131                SCORE(NN,NPOP)=SCORE(NN,NPOP)+DELTA(NMEAN,NPAR)
                   GO TO 120
C       PARTITION 2
132                IF (NPOP .EQ. 2) GO TO 120
                   SCORE(NN,NPOP)= SCORE(NN,NPOP)+DELTA(NMEAN,NPAR)
                   GO TO 120
C       PARTITION 3
133                IF (NPOP .EQ. 2) GO TO 150
                   SCORE(NN,NPOP)=SCORE(NN,NPOP)+DELTA(NMEAN,NPAR)
                   GO TO 120
150                SCORE(NN,NPOP)=SCORE(NN,NPOP)+.5*DELTA(NMEAN,NPAR)
                   GO TO 120
C          PARTITION 4
134             SCORE(NN,NPOP)=SCORE(NN,NPOP)+DELTA(NMEAN,NPAR)*(NPOP-1)/3
120          CONTINUE
             CALL STATS(SCORE,N,NDISP,NPAR,SIG)
C            FIND PATTERN OF SIGNIFICANT CONTRASTS
             DO 500 NT=1,8
                LL=0
                DO 540 L=1,6
                   OCCURS(NT,L)=OCCURS(NT,L)+SIG(NT,L)
540                LL=LL+SIG(NT,L)
                SUMSIG=LL
                LL=LL+1
                NCONTR=0
                GO TO (549,551,552,553,554,555,556),LL
549             PAT=1
                GO TO 557
551             PAT=2
                DO 590 II=1,6
590                IF (SIG(NT,II) .EQ. 1) NCONTR = II
                GO TO 557
552             PAT=4
                IF   ((SIG(NT,1)+SIG(NT,6)) .EQ. 2) NCONTR=1
                IF   ((SIG(NT,2)+SIG(NT,5)) .EQ. 2) NCONTR=2
                IF   ((SIG(NT,3)+SIG(NT,4)) .EQ. 2) NCONTR=3
                IF   (NCONTR .NE. 0) PAT=3
                IF  (NCONTR .NE. 0) GO TO 557
                IF   ((SIG(NT,1)+SIG(NT,2)) .EQ. 2) NCONTR=1
                IF   ((SIG(NT,1)+SIG(NT,3)) .EQ. 2) NCONTR=2
                IF   ((SIG(NT,1)+SIG(NT,4)) .EQ. 2) NCONTR=3
                IF   ((SIG(NT,1)+SIG(NT,5)) .EQ. 2) NCONTR=4
                IF   ((SIG(NT,2)+SIG(NT,3)) .EQ. 2) NCONTR=5
                IF   ((SIG(NT,2)+SIG(NT,4)) .EQ. 2) NCONTR=6
                IF   ((SIG(NT,2)+SIG(NT,6)) .EQ. 2) NCONTR=7
                IF   ((SIG(NT,3)+SIG(NT,5)) .EQ. 2) NCONTR=8
                IF   ((SIG(NT,3)+SIG(NT,6)) .EQ. 2) NCONTR=9
                IF   ((SIG(NT,4)+SIG(NT,5)) .EQ. 2) NCONTR=10
                IF   ((SIG(NT,4)+SIG(NT,6)) .EQ. 2) NCONTR=11
                IF   ((SIG(NT.5)+SIG(NT.6)) .EQ. 2) NCONTR=12
```

```
            GO TO 557
553         PAT=7
            IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,3))  .EQ. 3)  NCONTR=1
            IF ((SIG(NT,1)+SIG(NT,4)+SIG(NT,5))  .EQ. 3)  NCONTR=2
            IF ((SIG(NT,2)+SIG(NT,4)+SIG(NT,6))  .EQ. 3)  NCONTR=3
            IF ((SIG(NT,3)+SIG(NT,5)+SIG(NT,6))  .EQ. 3)  NCONTR=4
            IF (NCONTR .NE. 0) PAT=5
            IF (NCONTR .NE. 0) GO TO 557
            IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,4))  .EQ. 3)  NCONTR=1
            IF ((SIG(NT,2)+SIG(NT,3)+SIG(NT,6))  .EQ. 3)  NCONTR=2
            IF ((SIG(NT,4)+SIG(NT,5)+SIG(NT,6))  .EQ. 3)  NCONTR=3
            IF ((SIG(NT,3)+SIG(NT,5)+SIG(NT,6))  .EQ. 3)  NCONTR=4
            IF (NCONTR .NE. 0) PAT=6
            IF (NCONTR .NE. 0) GO TO 557
            IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,5))  .EQ. 3)  NCONTR=1
            IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,6))  .EQ. 3)  NCONTR=2
            IF ((SIG(NT,1)+SIG(NT,3)+SIG(NT,4))  .EQ. 3)  NCONTR=3
            IF ((SIG(NT,1)+SIG(NT,3)+SIG(NT,6))  .EQ. 3)  NCONTR=4
            IF ((SIG(NT,1)+SIG(NT,4)+SIG(NT,6))  .EQ. 3)  NCONTR=5
            IF ((SIG(NT,1)+SIG(NT,5)+SIG(NT,6))  .EQ. 3)  NCONTR=6
            IF ((SIG(NT,2)+SIG(NT,3)+SIG(NT,4))  .EQ. 3)  NCONTR=7
            IF ((SIG(NT,2)+SIG(NT,3)+SIG(NT,5))  .EQ. 3)  NCONTR=8
            IF ((SIG(NT,2)+SIG(NT,4)+SIG(NT,5))  .EQ. 3)  NCONTR=9
            IF ((SIG(NT,2)+SIG(NT,5)+SIG(NT,6))  .EQ. 3)  NCONTR=10
            IF ((SIG(NT,3)+SIG(NT,4)+SIG(NT,5))  .EQ. 3)  NCONTR=11
            IF ((SIG(NT,3)+SIG(NT,4)+SIG(NT,6))  .EQ. 3)  NCONTR=12
            GO TO 557
554         PAT=9
            IF ((SIG(NT,2)+SIG(NT,3)+SIG(NT,4)+SIG(NT,5)).EQ.4)  NCONTR=1
            IF ((SIG(NT,1)+SIG(NT,3)+SIG(NT,4)+SIG(NT,6)).EQ.4)  NCONTR=2
            IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,5)+SIG(NT,6)).EQ.4)  NCONTR=3
            IF (NCONTR .NE. 0)  PAT=8
            IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,3)+SIG(NT,4)).EQ.4)  NCONTR=1
            IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,3)+SIG(NT,5)).EQ.4)  NCONTR=2
            IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,3)+SIG(NT,6)).EQ.4)  NCONTR=3
            IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,4)+SIG(NT,5)).EQ.4)  NCONTR=4
            IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,4)+SIG(NT,6)).EQ.4)  NCONTR=5
            IF ((SIG(NT,1)+SIG(NT,3)+SIG(NT,4)+SIG(NT,5)).EQ.4)  NCONTR=6
            IF ((SIG(NT,1)+SIG(NT,3)+SIG(NT,5)+SIG(NT,6)).EQ.4)  NCONTR=7
            IF ((SIG(NT,1)+SIG(NT,4)+SIG(NT,5)+SIG(NT,6)).EQ.4)  NCONTR=8
            IF ((SIG(NT,3)+SIG(NT,4)+SIG(NT,5)+SIG(NT,6)).EQ.4)  NCONTR=9
            IF ((SIG(NT,2)+SIG(NT,3)+SIG(NT,4)+SIG(NT,6)).EQ.4)  NCONTR=10
            IF ((SIG(NT,2)+SIG(NT,3)+SIG(NT,5)+SIG(NT,6)).EQ.4)  NCONTR=11
            IF ((SIG(NT,2)+SIG(NT,4)+SIG(NT,5)+SIG(NT,6)).EQ.4)  NCONTR=12
            GO TO 557
555         PAT=10
            IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,3)+SIG(NT,4)+SIG(NT,5)).EQ.5)
     &         NCONTR=1
            IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,3)+SIG(NT,4)+SIG(NT,6)).EQ.5)
     &         NCONTR=2
            IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,3)+SIG(NT,5)+SIG(NT,6)).EQ.5)
     &         NCONTR=3
            IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,4)+SIG(NT,5)+SIG(NT,6)).EQ.5)
     &         NCONTR=4
            IF ((SIG(NT,1)+SIG(NT,3)+SIG(NT,4)+SIG(NT,5)+SIG(NT,6)).EQ.5)
     &         NCONTR=5
            IF ((SIG(NT,2)+SIG(NT,3)+SIG(NT,4)+SIG(NT,5)+SIG(NT,6)).EQ.5)
     &         NCONTR=6
            GO TO 557
556         PAT=11
557         FREQ(NT,PAT)=FREQ(NT,PAT)+1
            IF (NCONTR .EQ. 0) NCONTR = 1
            PCNTRS(NT,NCONTR,PAT) = PCNTRS(NT,NCONTR,PAT) + 1
C           COMPUTE ERROR CONTRASTS (TABLE 4 PG 11)
            GO TO (561,562,563,564,565),NPAR
C       PARTITION 1
561         IF ((SIG(NT,4)+SIG(NT,5)+SIG(NT,6))  .GT. 0)
     &         ERROR1(NT)=ERROR1(NT)+1
            IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,3))  .LT. 3)
     &         ERROR2(NT)=ERROR2(NT)+1
            GO TO 506
C       PARTITION 2
```

```
562         IF ((SIG(NT,1)+SIG(NT,6)) .GT. 0) ERROR1(NT)=
     &        ERROR1(NT)+1
            IF ((SIG(NT,2)+SIG(NT,3)+SIG(NT,4)+SIG(NT,5)) .LT. 4)
     &        ERROR2(NT)=ERROR2(NT)+1
            GO TO 500
C     PARTITION 3
563         IF (SIG(NT,6) .EQ. 1) ERROR1(NT)=ERROR1(NT)+1
            IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,3)+SIG(NT,4)+SIG(NT,5))
     &          .LT. 5) ERRCR2(NT)=ERROR2(NT)+1
            GO TO 500
C     PARTITION 4
564         IF (SUMSIG .NE. 6) ERROR2(NT)=ERROR2(NT)+1
            GO TO 500
C     PARTITION 5
565         IF (SUMSIG .GE. 1) ERROR1(NT)=ERROR1(NT)+1
500       CONTINUE
C     END OF 1000 SIMULATIONS.  PRINT REPORT.
          PAGE = PAGE + 1
          WRITE(6,210) PAGE
          WRITE(6,211) NOCOND
          WRITE(6,212) NPAR
          WRITE(6,216) NPAIR
          WRITE(6,213) NMEAN
          WRITE(6,214) NDISP
          WRITE(6,215) NHET
          WRITE(6,206)
          WRITE(6,207) (K,(OCCURS(NT,K) ,NT=1,8) ,K=1,6)
720       WRITE(6,208)
          DO 800 NTEST=1,8
            ERROR1(NTEST)=ERROR1(NTEST)/1000
            ERROR2(NTEST)=ERROR2(NTEST)/1000
            ACOMP=0
            AREJ=0
            DO 610 K=1,11
              ACOMP=ACOMP+FREQ(NTEST,K)*COMP(K)
610           AREJ=AREJ+FREQ(NTEST,K)*REJ(K)
            ACOMP=ACOMP/1000
            AREJ=AREJ/1000
800         WRITE(6,209) TITLE(NTEST),ERROR1(NTEST),ERROR2(NTEST),
     &               ACOMP,AREJ
            WRITE(6,220)
            WRITE(6,221) (PCNTRS(I,1,1) ,I=1,8)
            WRITE(6,240) (FREQ(NT,1) ,NT=1,8)
            WRITE(6,222) (J, (PCNTRS(I,J,2) ,I=1,8) ,J=1,6)
            WRITE(6,240) (FREQ(NT,2) ,NT=1,8)
            WRITE(6,223) (J, (PCNTRS(I,J,3) ,I=1,8) ,J=1,3)
            WRITE(6,240) (FREQ(NT,3) ,NT=1,8)
            WRITE(6,224) (J, (PCNTRS(I,J,4) ,I=1,8) ,J=1,12)
            WRITE(6,240) (FREQ(NT,4) ,NT=1,8)
            PAGE = PAGE + 1
            WRITE(6,210) PAGE
            WRITE(6,220)
            WRITE(6,225) (J, (PCNTRS(I,J,5) ,I=1,8) ,J=1,4)
            WRITE(6,240) (FREQ(NT,5) ,NT=1,8)
            WRITE(6,226) (J, (PCNTRS(I,J,6) ,I=1,8) ,J=1,4)
            WRITE(6,240) (FREQ(NT,6) ,NT=1,8)
            WRITE(6,227) (J, (PCNTRS(I,J,7) ,I=1,8) ,J=1,12)
            WRITE(6,240) (FREQ(NT,7) ,NT=1,8)
            WRITE(6,228) (J, (PCNTRS(I,J,8) ,I=1,8) ,J=1,3)
            WRITE(6,240) (FREQ(NT,8) ,NT=1,8)
            WRITE(6,229) (J, (PCNTRS(I,J,9) ,I=1,8) ,J=1,12)
            WRITE(6,240) (FREQ(NT,9) ,NT=1,8)
            WRITE(6,230) (J, (PCNTRS(I,J,10) ,I=1,8) ,J=1,6)
            WRITE(6,240) (FREQ(NT,10) ,NT=1,8)
            WRITE(6,231) (PCNTRS(I,1,11) ,I=1,8)
            WRITE(6,240) (FREQ(NT,11) ,NT=1,8)
900   CONTINUE
C
206   FORMAT('0         CONTRAST FREQUENCY COUNT PER METHOD ',/,
     & ' CONTRAST NO.          1    2    3    4    5    6    7    8',/)
207   FORMAT(11X,I2,3X,8I5)
```

```
208    FORMAT(/,' STATISTICAL METHOD    TYPE 1 ERROR   TYPE 2 ERROR',
     &'     AVG. COMP.    AVG. NO. REJ.')
209    FORMAT(3X,2A8,1X,F12.3,1X,F12.3,3X,F12.3,3X,F12.3)
210    FORMAT('1',25X,'SIMULATION',25X,'PAGE',I5)
211    FORMAT('0','          SIMULATION NUMBER                          ',I4)
212    FORMAT('            PARTITION NUMBER                         ',I4)
213    FORMAT('            MEAN VARIABILITY (1=SML,2=MED,3=LGE)     ',I4)
214    FORMAT('            SAMPLE SIZE DISPARITY (1=SML,2=MED,3=LGE) ',I4)
215    FORMAT('            VARIANCE HET. (1=0,2=.2,3=.6,4=1.0)      ',I4)
216    FORMAT('            PAIRING (1=DIRECT,2=INVERSE)             ',I4)
220    FORMAT('0',27X,' PATTERN FREQUENCY COUNT ',/,
     &       18X,'    1     2     3     4     5     6     7     8',/)
221    FORMAT(' PATTERN  1     1',2X,8I5)
222    FORMAT(' PATTERN  2     ',I2,2X,8I5,  5(/,14X,I2,2X,8I5))
223    FORMAT(' PATTERN  3     ',I2,2X,8I5,   2(/14X,I2,2X,8I5))
224    FORMAT(' PATTERN  4     ',I2,2X,8I5,  11(/14X,I2,2X,8I5))
225    FORMAT(' PATTERN  5     ',I2,2X,8I5,   3(/14X,I2,2X,8I5))
226    FORMAT(' PATTERN  6     ',I2,2X,8I5,   3(/14X,I2,2X,8I5))
227    FORMAT(' PATTERN  7     ',I2,2X,8I5,  11(/14X,I2,2X,8I5))
228    FORMAT(' PATTERN  8     ',I2,2X,8I5,   2(/14X,I2,2X,8I5))
229    FORMAT(' PATTERN  9     ',I2,2X,8I5,  11(/14X,I2,2X,8I5))
230    FORMAT(' PATTERN 10     ',I2,2X,8I5,   5(/14X,I2,2X,8I5))
231    FORMAT(' PATTERN 11     1',2X,8I5)
240    FORMAT('                     TOTAL',2X,8I5,/)
       STOP
       END
       SUBROUTINE STATS(SCORE,N,NDISP,NPAR,SIG)
       COMMON XM1,B1,XM2,B2,XM3,B3
       DIMENSION SCORE(48,4),N(3,4),XBAR(4),SSQ(4),CRIT(8,6),S(4)
       INTEGER   SIG(8,6)
       DO 5 NPOP=1,4
         XBAR(NPOP)=0
         ITEMP=N(NDISP,NPOP)
         DO 6 NN=1,ITEMP
6          XBAR(NPOP)=XBAR(NPOP)+SCORE(NN,NPOP)
5        XBAR(NPOP)=XBAR(NPOP)/N(NDISP,NPOP)
       SS=0
       NTOT=0
       DO 7 NPOP=1,4
         SSQ(NPOP)=0
         ITEMP=N(NDISP,NPOP)
         DO 8 NN=1,ITEMP
8          SSQ(NPOP)=SSQ(NPOP)+(SCORE(NN,NPOP)-XBAR(NPOP))**2
         SS=SS+SSQ(NPOP)
         NTOT=NTOT+N(NDISP,NPOP)
         SSQ(NPOP)=SSQ(NPOP) / ( N(NDISP,NPOP)-1 )
7        S(NPOP)=(SSQ(NPOP))**.5
       SS=(SS/(NTOT-4))**.5
       NCONT=0
       DO 10 NX1=1,3
         NXX=NX1+1
         DO 10 NX2=NXX,4
           NCONT=NCONT+1
           CRIT(1,NCONT)=3.685*SS*((1./ N(NDISP,NX1)  + 1./ N(NDISP,NX2))
     &                  / 2.0)**.5
          CRIT(2,NCONT)=3.686*SS*(AMAX1( 1.0/ N(NDISP,NX1) ,
     &        1.0/ N(NDISP,NX2 )))**.5
          CRIT(3,NCONT)=2.673*SS*(1.0/ N(NDISP,NX1) + 1.0/N(NDISP,NX2))
     &                  **.5
          CRIT(4,NCONT)=2.673*SS*((2* N(NDISP,NX1))**(-0.5) +
     &        (2.0* N(NDISP,NX2))**(-0.5))
          CRIT(5,NCONT)=3.686*(2*(AMAX1((S(NX1)**2/N(NDISP,NX1)) ,
     &        (S(NX2)**2/N(NDISP,NX2)))))**.5
          V=(SSQ(NX1)/N(NDISP,NX1) + SSQ(NX2)/N(NDISP,NX2)) **2/
     &       (SSQ(NX1)**2 / (N(NDISP,NX1)**2 * (N(NDISP,NX1)-1)) +
     &        SSQ(NX2)**2 / (N(NDISP,NX2)**2 * (N(NDISP,NX2)-1)))
          Y1=XM1/V + B1
          CRIT(6,NCONT)=Y1*(SSQ(NX1) / (2.0*N(NDISP,NX1)) +
     &        SSQ(NX2) / (2.0*N(NDISP,NX2)))**.5
C        T3 PROCEDURE STATISTIC 7
          Y2 = XM2/V +B2
          CRIT(7,NCONT)=Y2*(SSQ(NX1)/N(NDISP,NX1)+SSQ(NX2)/N(NDISP,NX2)
```

```
     &                     )**.5
C     C PROCEDURE STATISTIC 8
           YI=XM3/(N(NDISP,NX1)-1) + B3
           YJ=XM3/(N(NDISP,NX2)-1) + B3
           A=(YI*SSQ(NX1)/N(NDISP,NX1)+YJ*SSQ(NX2)/N(NDISP,NX2)) /
     &       (SSQ(NX1)/N(NDISP,NX1) + SSQ(NX2)/N(NDISP,NX2))
           CRIT(8,NCONT)=A*(SSQ(NX1)/(2*N(NDISP,NX1)) +
     &             SSQ(NX2) / (2*N(NDISP,NX2)))**.5
C       COMPARE CONTRASTS WITH 8 STATS.
          DO 20 NT=1,8
            SIG(NT,NCONT)=0
            IF ( (ABS(XBAR(NX1)-XBAR(NX2)) ) .GE. CRIT(NT,NCONT))
     &        SIG(NT,NCONT)=1
20          CONTINUE
10     CONTINUE
       RETURN
       END
//GO.FT06F001 DD UNIT=TAPE,VOL=SER=US0219,DSN=REPORT,DISP=(NEW,PASS),
//            LABEL=(1,SL),DCB=(LRECL=133,BLKSIZE=3990,RECFM=FBA)
//       EXEC PGM=IEBGENER
//SYSPRINT DD SYSOUT=A
//SYSIN DD DUMMY
//SYSUT1 DD DSN=REPORT,UNIT=TAPE,VOL=SER=US0219,DISP=(OLD,PASS),
//          LABEL=(1,SL),DCB=(LRECL=133,BLKSIZE=3990,RECFM=FBA)
//SYSUT2 DD SYSOUT=A,DCB=(LRECL=133,BLKSIZE=3990,RECFM=FBA)
//       EXEC PGM=IEBGENER
//SYSPRINT DD SYSOUT=A
//SYSIN DD DUMMY
//SYSUT2 DD DSN=SOURCE,UNIT=TAPE,VOL=SER=US0219,DISP=(NEW,PASS),
//          LABEL=(2,SL),DCB=(LRECL=80,BLKSIZE=800,RECFM=FB)
//SYSUT1 DD *
**DADTA    JOB    DADTAHIG,'HURDJ',CLASS=N,
 NOTE TO RUN THIS: CHA ALL **  AND ALL @@  + REMOVE THIS LINE.
**     MSGLEVEL=(1,1)
@@SETUP  .    PLEASE MOUNT TAPE US0219  RING IN THNAKS.
@@ROUTE PRINT WYLBUR
**     EXEC IMSLS
**FORT.SYSIN DD *
       INTEGER N(3,4),SIG(8,6),FREQ(8,11),PCNTRS(8,12,11),NCCNTR,
     &          PAT,SUMSIG,OCCURS(8,6),PAGE,HCNT,NOCONC
       REAL SAMP(48,4),COMP(11),REJ(11),VARHET(4,4,2),Z1(5),
     &       Z2(5),Y1(5),Y2(5),Y3(5),SCORE(48,4),ERROR1(8),ERROR2(8),
     &          HOLD(124),DELTA(3,4)
       COMPLEX*16   TITLE(8)
       DOUBLE PRECISION SEED
       COMMON XM1,B1,XM2,B2,XM3,B3
       DATA N/26,21,17,27,26,23,34,35,36,37,42,48/,
     &       VARHET/1.0,0.73,0.30,.01,1.0,.92,.80,.06,1.0,1.08,1.03,1.58,
     &       1.0,1.27,2.0,2.35,1.0,1.27,2.0,2.35,1.0,1.08,1.03,1.58,
     &       1.0,.92,.8,.06,1.0,.73,.30,.01/,
     &       Z1/24.0,30.0,40.0,60.0,120.0/,Z2/20.,24.,30.,40.,60./
       DATA Y1/3.901,3.845,3.781,3.737,3.685/,Y2/2.851,2.805,2.760,2.716,
     &          2.673/,Y3/3.958,3.901,3.845,3.791,3.737/,
     &       SZ1,SZ2,SY1,SY2,SY3,SSZ1,SSZ2,SZ1Y1,SZ1Y2,SZ2Y3/10*0./,
     &       PAGE,NCCOND/0,0/
       DATA    DELTA/.089333333,.223333333,.357333333,
     &               .089333333,.223333333,.357333333,
     &               .178666666,.446666666,.714666666,
     &               .268,.67,1.072/
       DATA    COMP/0.,2.,2.,1.,0.,2.,1.,0.,1.,0.,0./
       DATA    REJ/0.,1.,2.,2.,3.,3.,3.,4.,4.,5.,6./
       DATA    SEED/537424784D0/
       DATA    TITLE/'KRAMMER','SPJOTROLL,STOLIN','HOCHBERG(1974)',
     &       'GABRIEL','HOCHBERG(1976)','GAMES AND HOWELL',
     &       'T3 PROCEDURE','C PROCEDURE'/
C
C                THIS PROGRAM IS A PRODUCT OF JIM HURD AND ANN HURD
C                         JULY 1983. WRITEN FOR DON DIXIE.
C       CALCULATE CONSTANTS FOR STATISTICS 6,7,8(GAMES + HOWELL,T3,C)
       DO 1 K=1,5
         Z1(K)=1/Z1(K)
         Z2(K)=1/Z2(K)
```

```
          SZ1=SZ1+Z1(K)
          SZ2=SZ2+Z2(K)
          SY1=SY1+Y1(K)
          SY2=SY2+Y2(K)
          SY3=SY3+Y3(K)
          SZ1Y1=SZ1Y1+Z1(K)*Y1(K)
          SZ1Y2=SZ1Y2+Z1(K)*Y2(K)
          SZ2Y3=SZ2Y3+Z2(K)*Y3(K)
          SSZ1=SSZ1+Z1(K)**2
1         SSZ2=SSZ2+Z2(K)**2
      XM1 = (5.*SZ1Y1-SZ1*SY1)/(5.*SSZ1-SZ1**2)
      XM2 = (5.*SZ1Y2-SZ1*SY2)/(5.*SSZ1-SZ1**2)
      XM3 = (5.*SZ2Y3-SZ2*SY3)/(5.*SSZ2-SZ2**2)
      B1  = SY1/5.-XM1*SZ1/5.
      B2  = SY2/5.-XM2*SZ1/5.
      B3  = SY3/5.-XM3*SZ2/5.
C
      DO 90 I=1,4
        DO 90 J=1,4
          DO 90 K=1,2
90          VARHET(I,J,K) = SQRT(VARHET(I,J,K))
C
C     SET UP LOOPS FOR DIFFERENT CONDITIONS
      DO 900 NPAR=1,5
      DO 900 NPAIR=1,2
      DO 900 NMEAN=1,3
        IF ((NPAR .EQ. 5) .AND. (NMEAN .NE. 1)) GO TC 900
      DO 900 NDISP=1,3
      DO 900 NHET=1,4
C
C     INITIALIZE PATTERN FREQUENCY COUNTER.
      NOCOND = NOCOND + 1
      DO 2 K=1,8
        ERROR1(K) = 0
        ERROR2(K) = 0
        DO 3 J=1,6
3         OCCURS(K,J)=0
        DO 4 I=1,11
            DO 4 J=1,12
4           PCNTRS(K,J,I)=0
      DO 2 J=1,11
2         FREQ(K,J)=0
C
C     DO A SET OF 1000   SIMULATIONS
        DO 500 NSIM=1,1000
          CALL GGNPM(SEED,124,HOLD)
          HCNT = 1
          DO 120 NPOP=1,4
            ITEMP=N(NDISP,NPOP)
            DO 120 NN=1,ITEMP
C     FOR EACH SCORE IN ONE PARTICULAR PARTITION
              SCORE(NN,NPOP)=HOLD(HCNT)*VARHET(NHET,NPOP,NPAIR)
              HCNT = HCNT + 1
              IF ((NPOP .EQ. 1) .OR. (NPAR .EQ. 5))  GO TO 120
C             ADD APPROPRIATE DELTA TO SCORES
              GO TO (131,132,133,134),NPAR
C     PARTITION 1
131           SCORE(NN,NPOP)=SCORE(NN,NPOP)+DELTA(NMEAN,NPAR)
              GO TO 120
C     PARTITION 2
132           IF (NPOP .EQ. 2) GO TO 120
              SCORE(NN,NPOP)= SCORE(NN,NPOP)+DELTA(NMEAN,NPAR)
              GO TO 120
C     PARTITION 3
133           IF (NPOP .EQ. 2) GO TO 150
              SCORE(NN,NPOP)=SCORE(NN,NPOP)+DELTA(NMEAN,NPAR)
              GO TO 120
150           SCORE(NN,NPOP)=SCORE(NN,NPOP)+.5*DELTA(NMEAN,NPAR)
              GO TO 120
C       PARTITION 4
134         SCORE(NN,NPOP)=SCORE(NN,NPOP)+DELTA(NMEAN,NPAR)*(NPOP-1)/3
120         CONTINUE
```

```
      CALL STATS(SCORE,N,NDISP,NPAR,SIG)
C     FIND PATTERN OF SIGNIFICANT CONTRASTS
      DO 500 NT=1[illegible]
        LL=0
        DO 540 L=1,6
          OCCURS(NT,L)=OCCURS(NT,L)+SIG(NT,L)
540       LL=LL+SIG(NT,L)
        SUMSIG=LL
        LL=LL+1
        NCONTR=0
        GO TO (549,551,552,553,554,555,556),LL
549     PAT=1
        GO TO 557
551     PAT=2
        DO 590 II=1,6
590       IF (SIG(NT,II) .EQ. 1) NCONTR = II
        GO TO 557
552     PAT=4
        IF   ((SIG(NT,1)+SIG(NT,6)) .EQ. 2) NCONTR=1
        IF   ((SIG(NT,2)+SIG(NT,5)) .EQ. 2) NCONTR=2
        IF   ((SIG(NT,3)+SIG(NT,4)) .EQ. 2) NCONTR=3
        IF   (NCONTR .NE. 0) PAT=3
        IF  (NCONTR .NE. 0) GO TO 557
        IF   ((SIG(NT,1)+SIG(NT,2)) .EQ. 2) NCONTR=1
        IF   ((SIG(NT,1)+SIG(NT,3)) .EQ. 2) NCONTR=2
        IF   ((SIG(NT,1)+SIG(NT,4)) .EQ. 2) NCONTR=3
        IF   ((SIG(NT,1)+SIG(NT,5)) .EQ. 2) NCONTR=4
        IF   ((SIG(NT,2)+SIG(NT,3)) .EQ. 2) NCONTR=5
        IF   ((SIG(NT,2)+SIG(NT,4)) .EQ. 2) NCONTR=6
        IF   ((SIG(NT,2)+SIG(NT,6)) .EQ. 2) NCONTR=7
        IF   ((SIG(NT,3)+SIG(NT,5)) .EQ. 2) NCONTR=8
        IF   ((SIG(NT,3)+SIG(NT,6)) .EQ. 2) NCONTR=9
        IF   ((SIG(NT,4)+SIG(NT,5)) .EQ. 2) NCONTR=10
        IF   ((SIG(NT,4)+SIG(NT,6)) .EQ. 2) NCONTR=11
        IF   ((SIG(NT,5)+SIG(NT,6)) .EQ. 2) NCONTR=12
        GO TO 557
553     PAT=7
        IF  ((SIG(NT,1)+SIG(NT,2)+SIG(NT,3)) .EQ. 3) NCONTR=1
        IF  ((SIG(NT,1)+SIG(NT,4)+SIG(NT,5)) .EQ. 3) NCONTR=2
        IF  ((SIG(NT,2)+SIG(NT,4)+SIG(NT,6)) .EQ. 3) NCONTR=3
        IF  ((SIG(NT,3)+SIG(NT,5)+SIG(NT,6)) .EQ. 3) NCONTR=4
        IF  (NCONTR .NE. 0) PAT=5
        IF  (NCONTR .NE. 0) GO TO 557
        IF  ((SIG(NT,1)+SIG(NT,2)+SIG(NT,4)) .EQ. 3) NCONTR=1
        IF  ((SIG(NT,2)+SIG(NT,3)+SIG(NT,6)) .EQ. 3) NCONTR=2
        IF  ((SIG(NT,4)+SIG(NT,5)+SIG(NT,6)) .EQ. 3) NCONTR=3
        IF  ((SIG(NT,3)+SIG(NT,5)+SIG(NT,6)) .EQ. 3) NCONTR=4
        IF  (NCONTR .NE. 0) PAT=6
        IF  (NCONTR .NE. 0) GO TO 557
        IF  ((SIG(NT,1)+SIG(NT,2)+SIG(NT,5)) .EQ. 3) NCONTR=1
        IF  ((SIG(NT,1)+SIG(NT,2)+SIG(NT,6)) .EQ. 3) NCONTR=2
        IF  ((SIG(NT,1)+SIG(NT,3)+SIG(NT,4)) .EQ. 3) NCONTR=3
        IF  ((SIG(NT,1)+SIG(NT,3)+SIG(NT,6)) .EQ. 3) NCONTR=4
        IF  ((SIG(NT,1)+SIG(NT,4)+SIG(NT,6)) .EQ. 3) NCONTR=5
        IF  ((SIG(NT,1)+SIG(NT,5)+SIG(NT,6)) .EQ. 3) NCONTR=6
        IF  ((SIG(NT,2)+SIG(NT,3)+SIG(NT,4)) .EQ. 3) NCONTR=7
        IF  ((SIG(NT,2)+SIG(NT,3)+SIG(NT,5)) .EQ. 3) NCONTR=8
        IF  ((SIG(NT,2)+SIG(NT,4)+SIG(NT,5)) .EQ. 3) NCONTR=9
        IF  ((SIG(NT,2)+SIG(NT,5)+SIG(NT,6)) .EQ. 3) NCONTR=10
        IF  ((SIG(NT,3)+SIG(NT,4)+SIG(NT,5)) .EQ. 3) NCONTR=11
        IF  ((SIG(NT,3)+SIG(NT,4)+SIG(NT,6)) .EQ. 3) NCONTR=12
        GO TO 557
554     PAT=9
        IF  ((SIG(NT,2)+SIG(NT,3)+SIG(NT,4)+SIG(NT,5)).EQ.4)  NCONTR=1
        IF  ((SIG(NT,1)+SIG(NT,3)+SIG(NT,4)+SIG(NT,6)).EQ.4)  NCONTR=2
        IF  ((SIG(NT,1)+SIG(NT,2)+SIG(NT,5)+SIG(NT,6)).EQ.4)  NCONTR=3
        IF  (NCONTR .NE. 0) PAT=8
        IF  ((SIG(NT,1)+SIG(NT,2)+SIG(NT,3)+SIG(NT,4)).EQ.4)  NCONTR=1
        IF  ((SIG(NT,1)+SIG(NT,2)+SIG(NT,3)+SIG(NT,5)).EQ.4)  NCONTR=2
        IF  ((SIG(NT,1)+SIG(NT,2)+SIG(NT,3)+SIG(NT,6)).EQ.4)  NCONTR=3
        IF  ((SIG(NT,1)+SIG(NT,2)+SIG(NT,4)+SIG(NT,5)).EQ.4)  NCONTR=4
        IF  ((SIG(NT,1)+SIG(NT,2)+SIG(NT,4)+SIG(NT,6)).EQ.4)  NCONTR=5
```

```
          IF ((SIG(NT,1)+SIG(NT,3)+SIG(NT,4)+SIG(NT,5)).EQ.4) NCONTR=6
          IF ((SIG(NT,1)+SIG(NT,3)+SIG(NT,5)+SIG(NT,6)).EQ.4) NCONTR=7
          IF ((SIG(NT,1)+SIG(NT,4)+SIG(NT,5)+SIG(NT,6)).EQ.4) NCONTR=8
          IF ((SIG(NT,3)+SIG(NT,4)+SIG(NT,5)+SIG(NT,6)).EQ.4) NCONTR=9
         IF ((SIG(NT,2)+SIG(NT,3)+SIG(NT,4)+SIG(NT,6)).EQ.4) NCONTR=10
         IF ((SIG(NT,2)+SIG(NT,3)+SIG(NT,5)+SIG(NT,6)).EQ.4) NCONTR=11
         IF ((SIG(NT,2)+SIG(NT,4)+SIG(NT,5)+SIG(NT,6)).EQ.4) NCONTR=12
          GO TO 557
555       PAT=10
         IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,3)+SIG(NT,4)+SIG(NT,5)).EQ.5)
     &      NCONTR=1
         IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,3)+SIG(NT,4)+SIG(NT,6)).EQ.5)
     &      NCONTR=2
         IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,3)+SIG(NT,5)+SIG(NT,6)).EQ.5)
     &      NCONTR=3
         IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,4)+SIG(NT,5)+SIG(NT,6)).EQ.5)
     &      NCONTR=4
         IF ((SIG(NT,1)+SIG(NT,3)+SIG(NT,4)+SIG(NT,5)+SIG(NT,6)).EQ.5)
     &      NCONTR=5
         IF ((SIG(NT,2)+SIG(NT,3)+SIG(NT,4)+SIG(NT,5)+SIG(NT,6)).EQ.5)
     &      NCONTR=6
          GO TO 557
556       PAT=11
557       FREQ(NT,PAT)=FREQ(NT,PAT)+1
          IF (NCONTR .EQ. 0) NCONTR = 1
          PCNTRS(NT,NCONTR,PAT) = PCNTRS(NT,NCONTR,PAT) + 1
C         COMPUTE ERROR CONTRASTS (TABLE 4 PG 11)
          GO TO (561,562,563,564,565),NPAR
C     PARTITION 1
561       IF ((SIG(NT,4)+SIG(NT,5)+SIG(NT,6)) .GT. 0)
     &       ERROR1(NT)=ERROR1(NT)+1
          IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,3)) .LT. 3)
     &       ERROR2(NT)=ERROR2(NT)+1
          GO TO 500
C     PARTITION 2
562       IF ((SIG(NT,1)+SIG(NT,6)) .GT. 0) ERROR1(NT)=
     &      ERROR1(NT)+1
          IF ((SIG(NT,2)+SIG(NT,3)+SIG(NT,4)+SIG(NT,5)) .LT. 4)
     &      ERROR2(NT)=ERROR2(NT)+1
          GO TO 500
C     PARTITION 3
563       IF (SIG(NT,6) .EQ. 1) ERROR1(NT)=ERROR1(NT)+1
          IF ((SIG(NT,1)+SIG(NT,2)+SIG(NT,3)+SIG(NT,4)+SIG(NT,5))
     &       .LT. 5) ERROR2(NT)=ERROR2(NT)+1
          GO TO 500
C     PARTITION 4
564       IF (SUMSIG .NE. 6) ERROR2(NT)=ERROR2(NT)+1
          GO TO 500
C     PARTITION 5
565       IF (SUMSIG .GE. 1) ERROR1(NT)=ERROR1(NT)+1
500     CONTINUE
C     END OF 1000 SIMULATIONS.  PRINT REPORT.
        PAGE = PAGE + 1
        WRITE(6,210) PAGE
        WRITE(6,211) NOCOND
        WRITE(6,212) NPAR
        WRITE(6,216) NPAIR
        WRITE(6,213) NMEAN
        WRITE(6,214) NDISP
        WRITE(6,215) NHET
        WRITE(6,206)
        WRITE(6,207) (K,(OCCURS(NT,K) ,NT=1,8) ,K=1,6)
720     WRITE(6,208)
        DO 600 NTEST=1,8
          ERROR1(NTEST)=ERROR1(NTEST)/1000
          ERROR2(NTEST)=ERROR2(NTEST)/1000
          ACOMP=0
          AREJ=0
          DO 610 K=1,11
            ACOMP=ACOMP+FREQ(NTEST,K)*COMP(K)
610         AREJ=AREJ+FREQ(NTEST,K)*REJ(K)
          ACOMP=ACOMP/1000
```

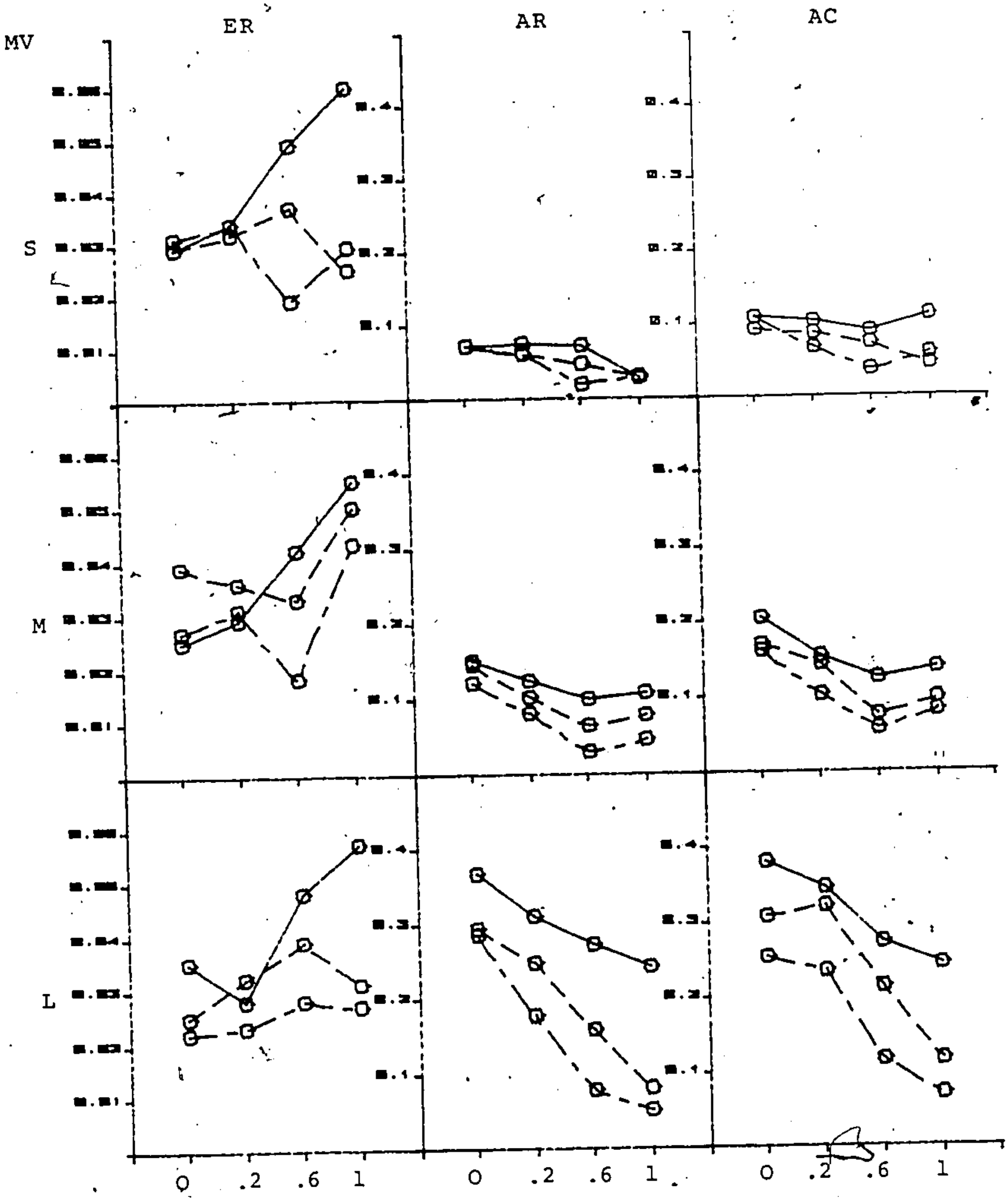
```
               AREJ=AREJ/1000
 600           WRITE(6,209) TITLE(NTEST),ERROR1(NTEST),ERROR2(NTEST),
     &              ACOMP,AREJ
               WRITE(6,220)
               WRITE(6,221)  (PCNTRS(I,1,1)  ,I=1,8)
               WRITE(6,240)  (FREQ(NT,1)  ,NT=1,8)
               WRITE(6,222)  (J,  (PCNTRS(I,J,2)  ,I=1,8)  ,J=1,6)
               WRITE(6,240)  (FREQ(NT,2)  ,NT=1,8)
               WRITE(6,223)  (J,  (PCNTRS(I,J,3)  ,I=1,8)  ,J=1,3)
               WRITE(6,240)  (FREQ(NT,3)  ,NT=1,8)
               WRITE(6,224)  (J,  (PCNTRS(I,J,4)  ,I=1,8)  ,J=1,12)
               WRITE(6,240)  (FREQ(NT,4)  ,NT=1,8)
               PAGE = PAGE + 1
               WRITE(6,210)  PAGE
               WRITE(6,220)
               WRITE(6,225)  (J,  (PCNTRS(I,J,5)  ,I=1,8)  ,J=1,4)
               WRITE(6,240)  (FREQ(NT,5)  ,NT=1,8)
               WRITE(6,226)  (J,  (PCNTRS(I,J,6)  ,I=1,8)  ,J=1,4)
               WRITE(6,240)  (FREQ(NT,6)  ,NT=1,8)
               WRITE(6,227)  (J,  (PCNTRS(I,J,7)  ,I=1,8)  ,J=1,12)
               WRITE(6,240)  (FREQ(NT,7)  ,NT=1,8)
               WRITE(6,228)  (J,  (PCNTRS(I,J,8)  ,I=1,8)  ,J=1,3)
               WRITE(6,240)  (FREQ(NT,8)  ,NT=1,8)
               WRITE(6,229)  (J,  (PCNTRS(I,J,9)  ,I=1,8)  ,J=1,12)
               WRITE(6,240)  (FREQ(NT,9)  ,NT=1,8)
               WRITE(6,230)  (J,  (PCNTRS(I,J,10)  ,I=1,8)  ,J=1,6)
               WRITE(6,240)  (FREQ(NT,10)  ,NT=1,8)
               WRITE(6,231)  (PCNTRS(I,1,11)  ,I=1,8)
               WRITE(6,240)  (FREQ(NT,11)  ,NT=1,8)

 900    CONTINUE
 C
 206    FORMAT('0            CONTRAST FREQUENCY COUNT PER METHOD ',/,
       & ' CONTRAST NO.          1    2    3    4    5    6    7    8',/)
 207    FORMAT(11X,I2,3X,8I5)
 208    FORMAT(/,' STATISTICAL METHOD   TYPE 1 ERROR   TYPE 2 ERROR',
       &'    AVG. COMP.    AVG. NO. REJ.')
 209    FORMAT(3X,2A8,1X,F12.3,1X,F12.3,3X,F12.3,3X,F12.3)
 210    FORMAT('1',25X,'SIMULATION',25X,'PAGE',I5)
 211    FORMAT('0','            SIMULATION NUMBER                             ',I4)
 212    FORMAT('           PARTITION NUMBER -                            ',I4)
 213    FORMAT('           MEAN VARIABILITY (1=SML,2=MED,3=LGE)          ',I4)
 214    FORMAT('           SAMPLE SIZE DISPARITY (1=SML,2=MED,3=LGE)     ',I4)
 215    FORMAT('           VARIANCE HET. (1=0,2=.2,3=.6,4=1.0)           ',I4)
 216    FORMAT('           PAIRING (1=DIRECT,2=INVERSE)                  ',I4)
 220    FORMAT('0',27X,' PATTERN FREQUENCY COUNT ',/,
       &          18X,'    1    2    3    4    5    6    7    8',/)
 221    FORMAT(' PATTERN  1      1',2X,8I5)
 222    FORMAT(' PATTERN  2      ',I2,2X,8I5, 5(/,14X,I2,2X,8I5))
 223    FORMAT(' PATTERN  3      ',I2,2X,8I5,   2(/14X,I2,2X,8I5))
 224    FORMAT(' PATTERN  4      ',I2,2X,8I5,  11(/14X,I2,2X,8I5))
 225    FORMAT(' PATTERN  5      ',I2,2X,8I5,   3(/14X,I2,2X,8I5))
 226    FORMAT(' PATTERN  6      ',I2,2X,8I5,   3(/14X,I2,2X,8I5))
 227    FORMAT(' PATTERN  7      ',I2,2X,8I5,  11(/14X,I2,2X,8I5))
 228    FORMAT(' PATTERN  8      ',I2,2X,8I5,   2(/14X,I2,2X,8I5))
 229    FORMAT(' PATTERN  9      ',I2,2X,8I5,  11(/14X,I2,2X,8I5))
 230    FORMAT(' PATTERN 10      ',I2,2X,8I5,   5(/14X,I2,2X,8I5))
 231    FORMAT(' PATTERN 11      1',2X,8I5)
 240    FORMAT('                 TOTAL',2X,8I5,/)
        STOP
        END
        SUBROUTINE STATS(SCORE,N,NDISP,NPAR,SIG)
        COMMON XM1,B1,XM2,B2,XM3,B3
        DIMENSION SCORE(48,4),N(3,4),XBAR(4),SSO(4),CRIT(8,8),S(4)
        INTEGER   SIG(8,8)
        DO 5 NPOP=1,4
          XBAR(NPOP)=0
          ITEMP=N(NDISP,NPOP)
          DO 6 NN=1,ITEMP
 6          XBAR(NPOP)=XBAR(NPOP)+SCORE(NN,NPOP)
 5        XBAR(NPOP)=XBAR(NPOP)/N(NDISP,NPOP)
        SS=0
```

```
        NTOT=0
        DO 7 NPOP=1,4
          SSQ(NPOP)=0
          ITEMP=N(NDISP,NPOP)
          DO 8 NN=1,ITEMP
8           SSQ(NPOP)=SSQ(NPOP)+(SCORE(NN,NPOP)-XBAR(NPOP))**2
          SS=SS+SSQ(NPOP)
          NTOT=NTOT+N(NDISP,NPOP)
          SSQ(NPOP)=SSQ(NPOP) / ( N(NDISP,NPOP)-1. )
7         S(NPOP)=(SSQ(NPCP))**.5
        SS=(SS/(NTOT-4))**.5
        NCONT=0
        DO 10 NX1=1,3
          NXX=NX1+1
          DO 10 NX2=NXX,4
            NCONT=NCONT+1
            CRIT(1,NCONT)=3.685*SS*((1./ N(NDISP,NX1) + 1./ N(NDISP,NX2))
     &                      / 2.0)**.5
           CRIT(2,NCONT)=3.686*SS*(AMAX1( 1.0/ N(NDISP,NX1) ,
     &         1.0/ N(NDISP,NX2 )))**.5
           CRIT(3,NCONT)=2.673*SS*(1.0/ N(NDISP,NX1) + 1.0/N(NDISP,NX2))
     &                      **.5
            CRIT(4,NCONT)=2.673*SS*((2* N(NDISP,NX1))**(-0.5) +
     &          (2.0* N(NDISP,NX2))**(-0.5))
           CRIT(5,NCONT)=3.686*(2*(AMAX1((S(NX1)**2/N(NDISP,NX1)) ,
     &         (S(NX2)**2/N(NDISP,NX2)))))**.5
           V=(SSQ(NX1)/N(NDISP,NX1) + SSQ(NX2)/N(NDISP,NX2)) **2/
     &        (SSQ(NX1)**2 / (N(NDISP,NX1)**2 * (N(NDISP,NX1)-1)) +
     &         SSQ(NX2)**2 / (N(NDISP,NX2)**2 * (N(NDISP,NX2)-1)))
           Y1=XM1/V + B1
           CRIT(6,NCONT)=Y1*(SSQ(NX1) / (2.0*N(NDISP,NX1)) +
     &         SSQ(NX2) / (2.0*N(NDISP,NX2)))**.5
C       T3 PROCEDURE STATISTIC 7
           Y2 = XM2/V +B2
           CRIT(7,NCONT)=Y2*(SSQ(NX1)/N(NDISP,NX1)+SSQ(NX2)/N(NDISP,NX2)
     &                      )**.5
C       C PROCEDURE STATISTIC 8
              YI=XM3/(N(NDISP,NX1)-1) + B3
              YJ=XM3/(N(NDISP,NX2)-1) + B3
              A=(YI*SSQ(NX1)/N(NDISP,NX1)+YJ*SSQ(NX2)/N(NDISP,NX2)) /
     &          (SSQ(NX1)/N(NDISP,NX1) + SSQ(NX2)/N(NDISP,NX2))
              CRIT(8,NCONT)=A*(SSQ(NX1)/(2*N(NDISP,NX1)) +
     &                 SSQ(NX2) / (2*N(NDISP,NX2)))**.5
C         COMPARE CONTRASTS WITH 8 STATS.
             DO 20 NT=1,8
               SIG(NT,NCONT)=0
               IF ( (ABS(XBAR(NX1)-XBAR(NX2)) ) .GE. CRIT(NT,NCONT))
     &            SIG(NT,NCONT)=1
20             CONTINUE
10        CONTINUE
          RETURN
          END
**FT06F001 DD UNIT=TAPE,VOL=SER=US0219,DSN=REPORT,DISP=(NEW,PASS),
**            LABEL=(1,SL),DCB=(LRECL=133,BLKSIZE=3990,RECFM=FBA,)
**        EXEC PGM=IEBGENER
**SYSPRINT DD SYSOUT=A
**SYSIN DD DUMMY
**SYSUT1 DD DSN=REPORT,UNIT=TAPE,VOL=SER=US0219,DISP=(OLD,PASS),
**           LABEL=(1,SL),DCB=(LRECL=133,BLKSIZE=3990,RECFM=FBA)
**SYSUT2 DD SYSOUT=A
**        EXEC PGM=IEBGENER
**SYSPRINT DD SYSOUT=A
**SYSIN DD DUMMY
**SYSUT1 DD DSN=SOURCE,UNIT=TAPE,VOL=SER=US0219,DISP=(OLD,KEEP),
**           LABEL=(1,SL),DCB=(LRECL=80,BLKSIZE=800,RECFM=FB)
**SYSUT2 DD *
@@
/*
//        EXEC TLDUMP,VOL=US0219,CNT=99
/*
```

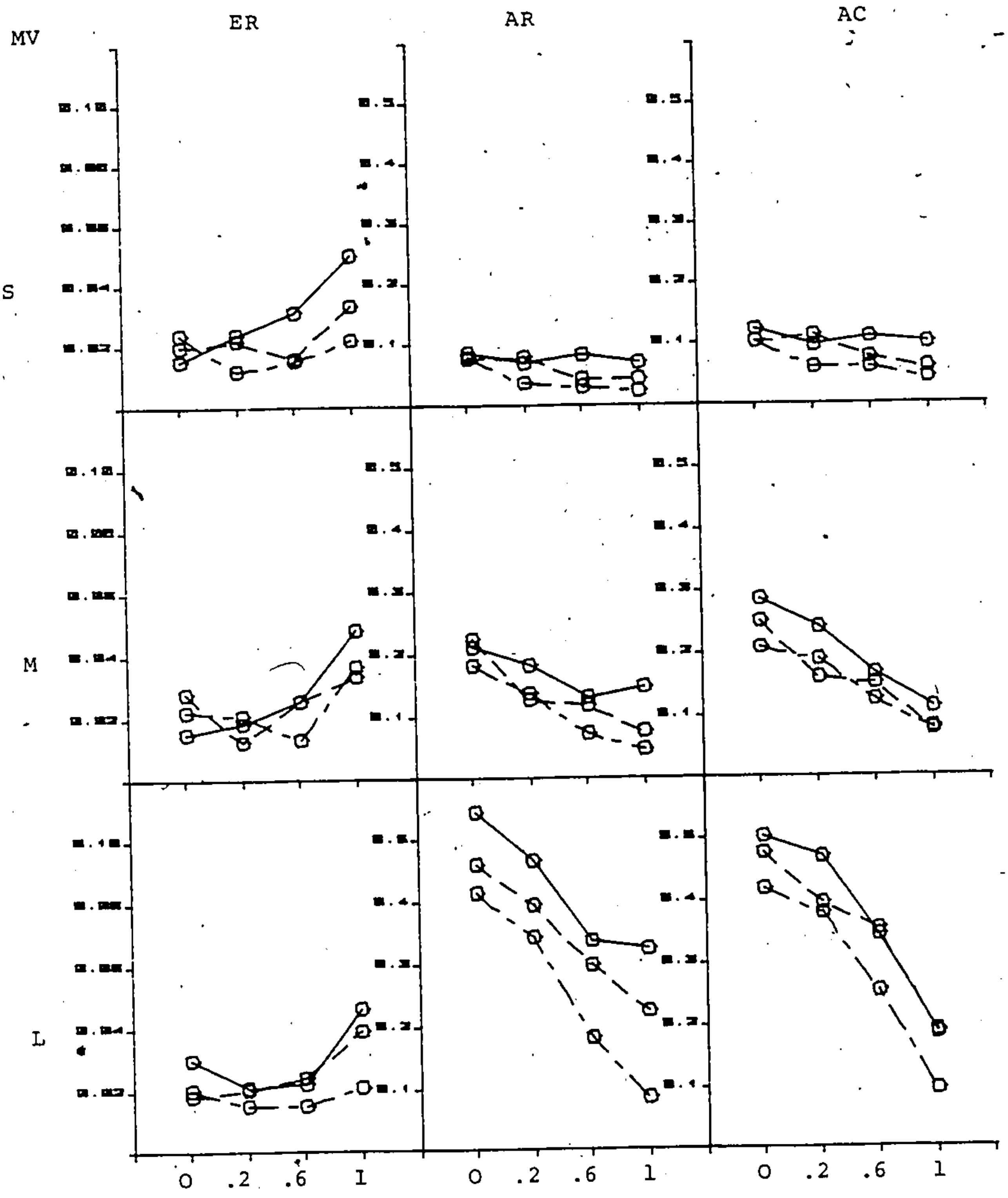
APPENDIX II

SUMMARY OF ERROR RATES, AVERAGE NUMBER OF REJECTIONS AND AVERAGE COMPLEXITY FOR THE KRAMER AND GAMES AND HOWELL METHODS UNDER EACH PARTITION AND IN THE INVERSE AND DIRECT PAIRING CONDITIONS

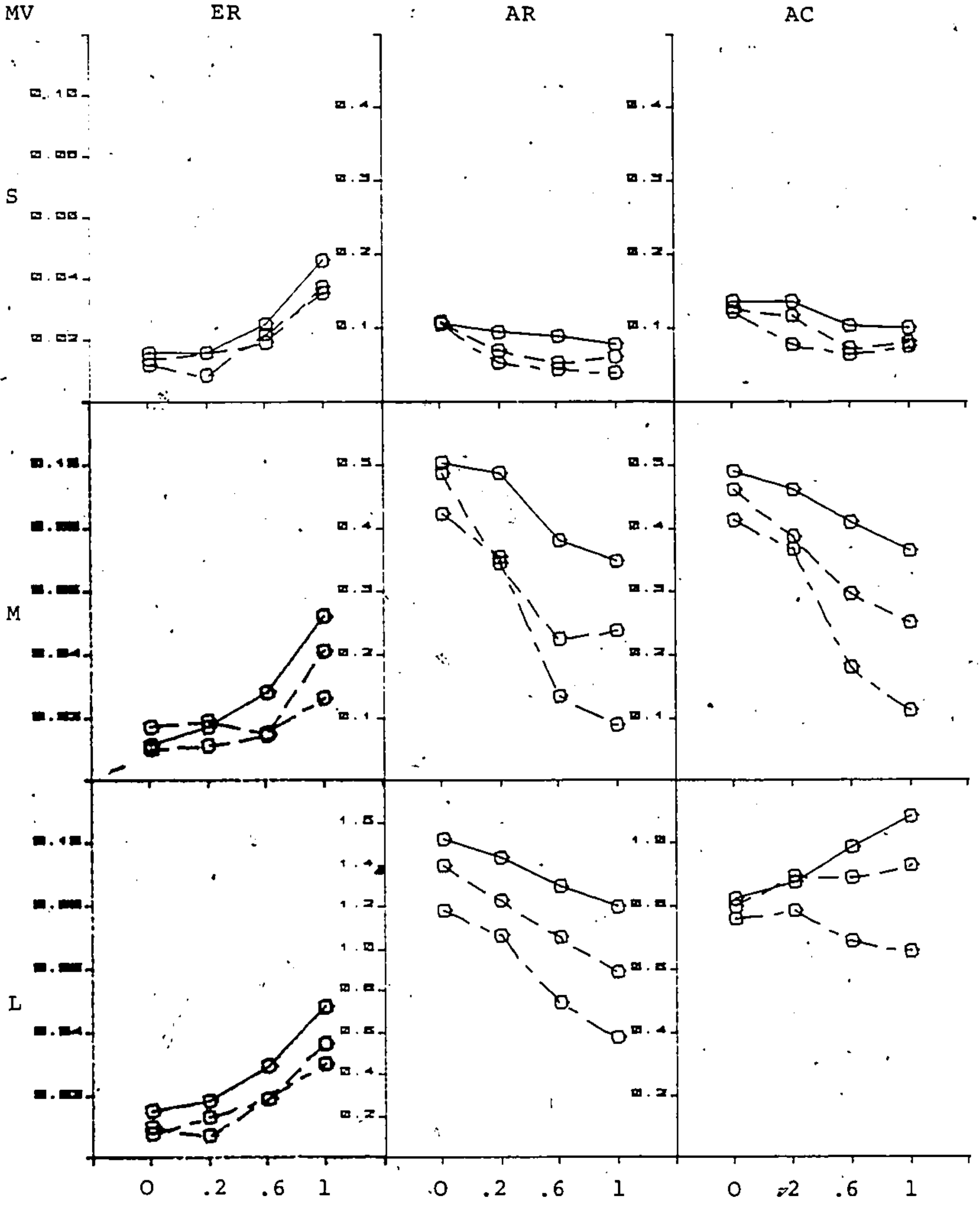
Partition I, Direct, Kramer Method



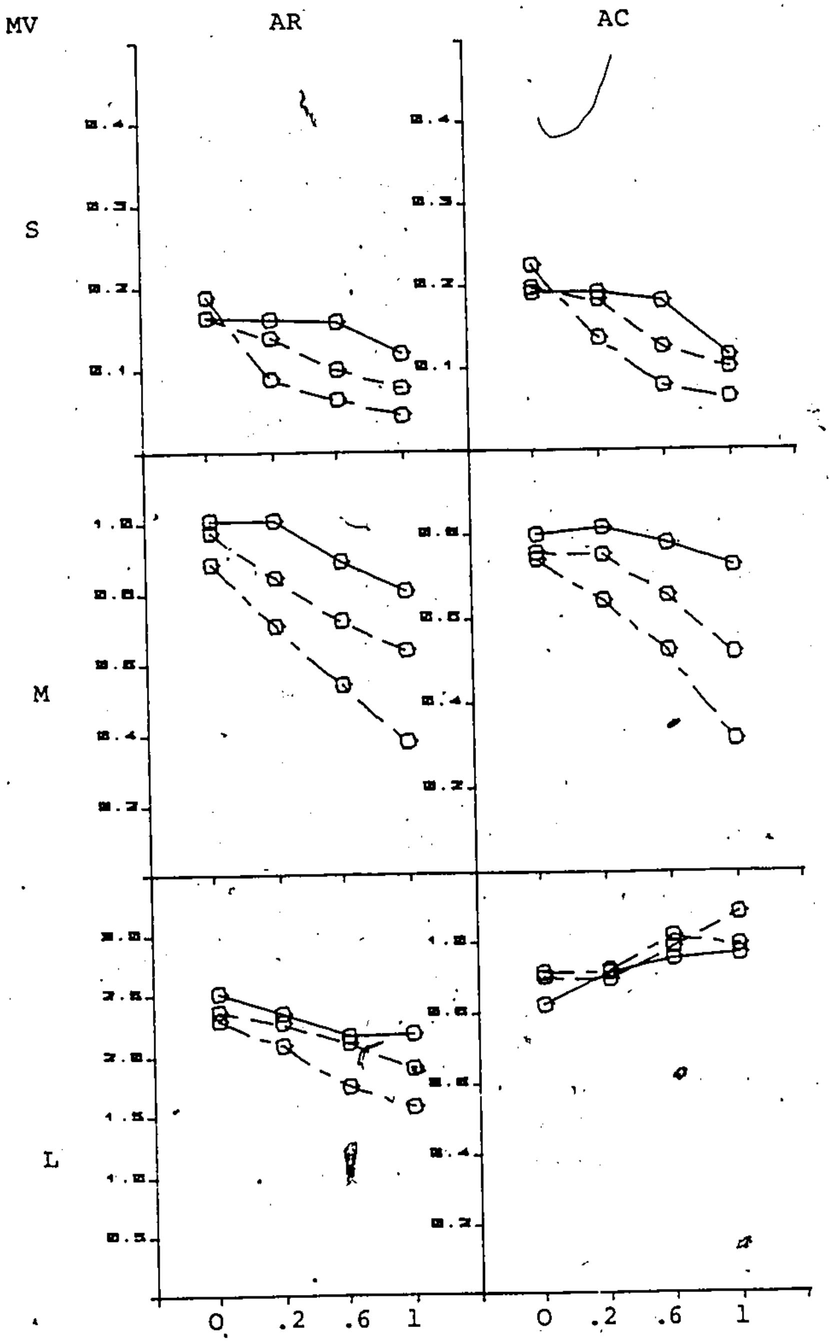
Partition II, Direct, Kramer Method



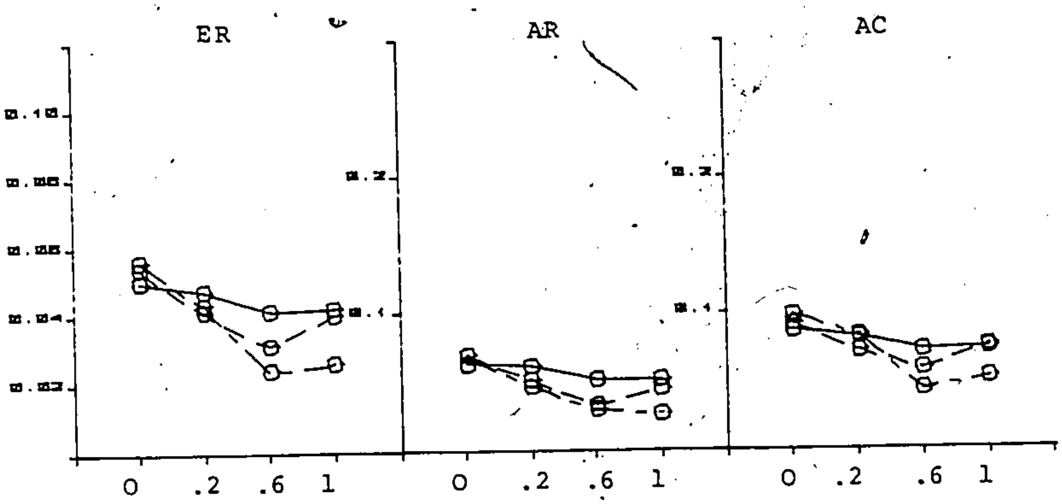
Partition III, Direct, Kramer Method



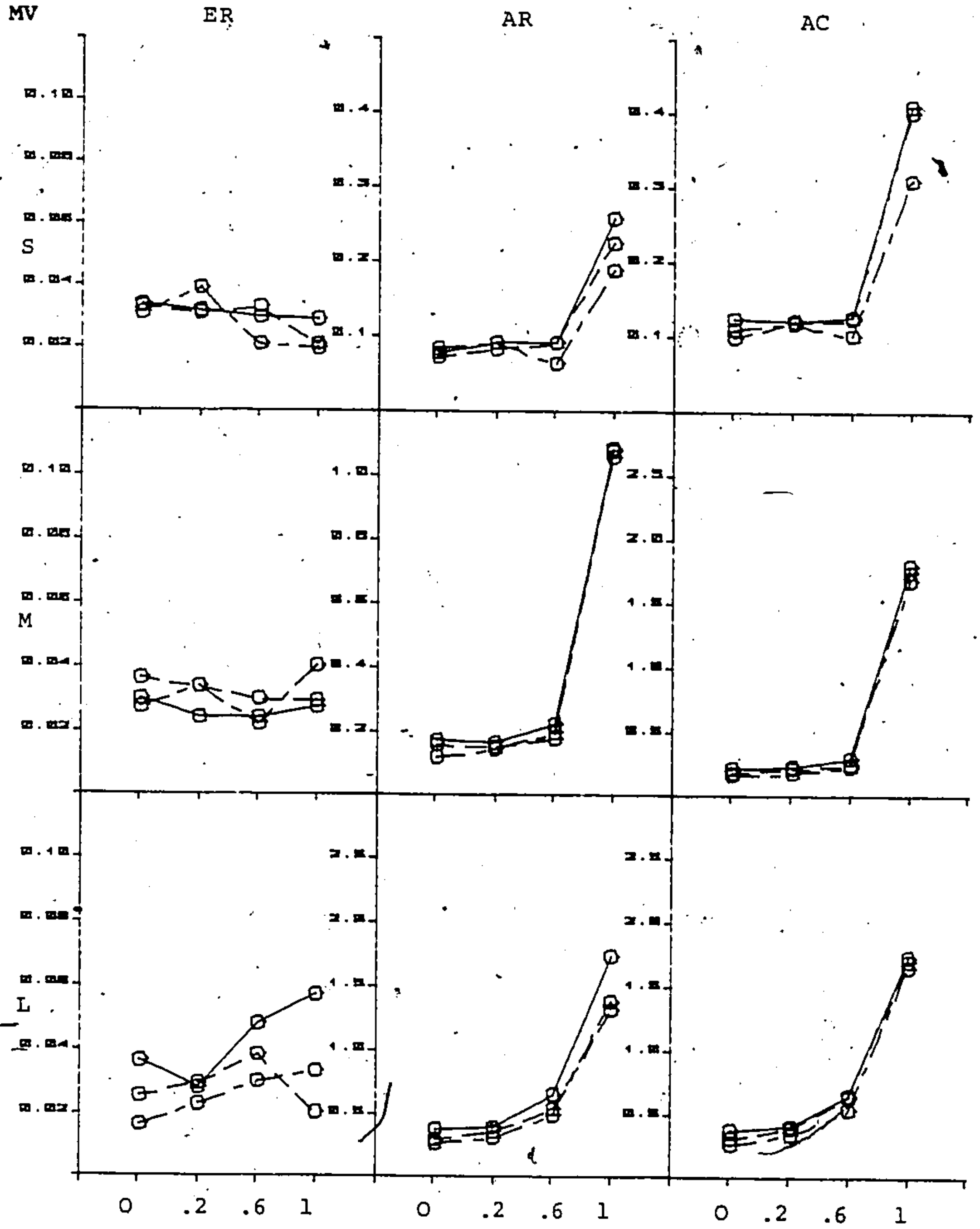
Partition IV, Direct, Kramer Method



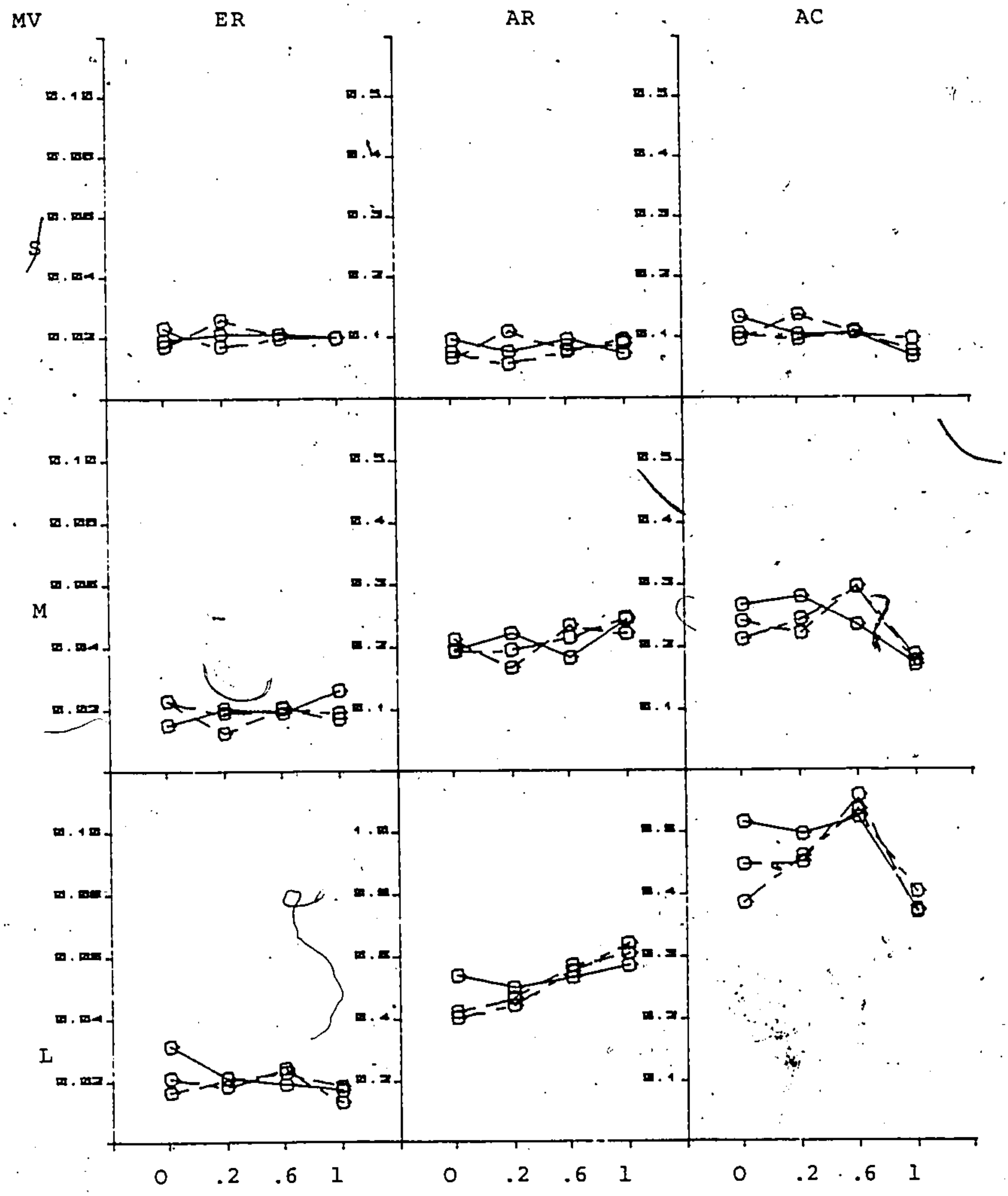
Partition V, Direct, Kramer Method



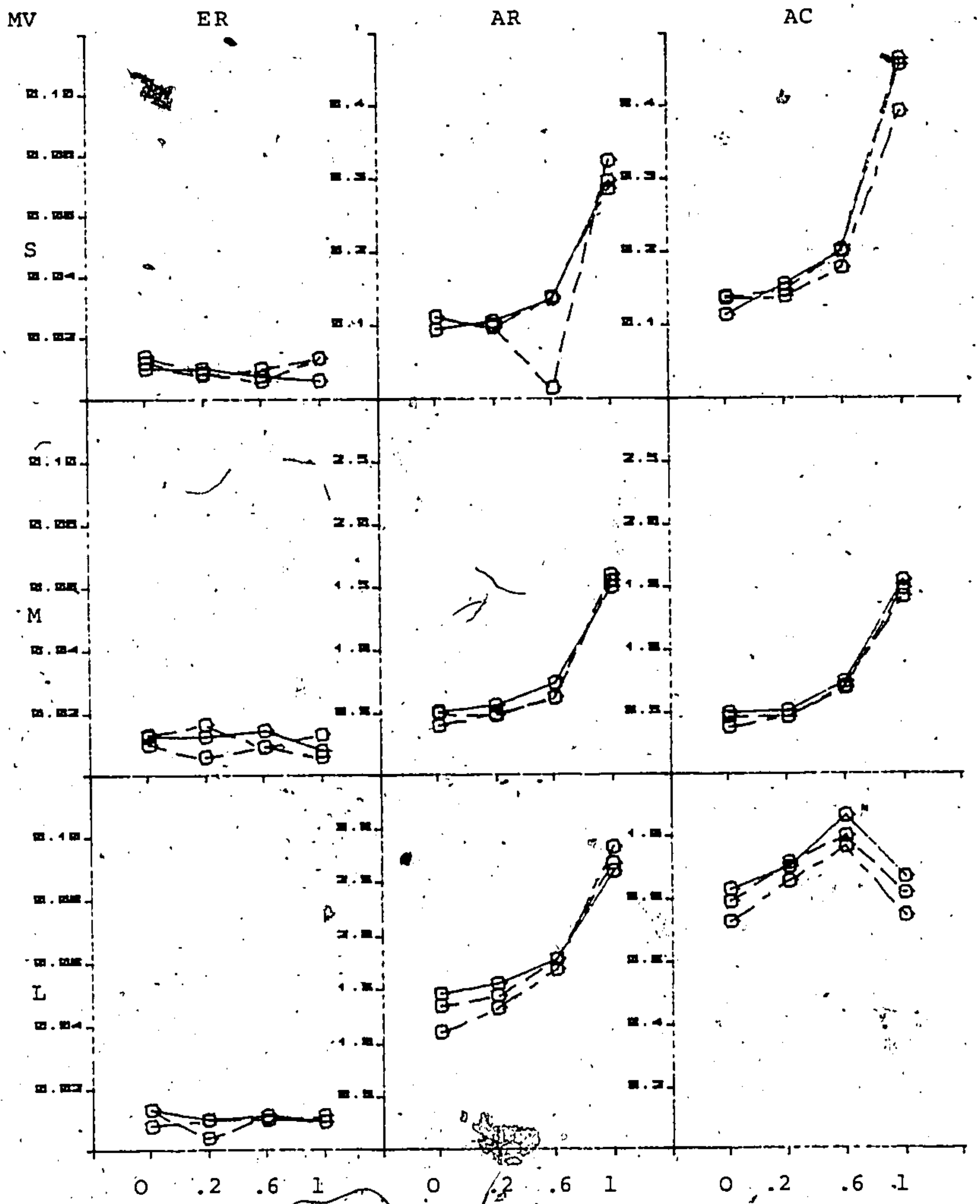
Partition I, Direct, Games and Howell Method



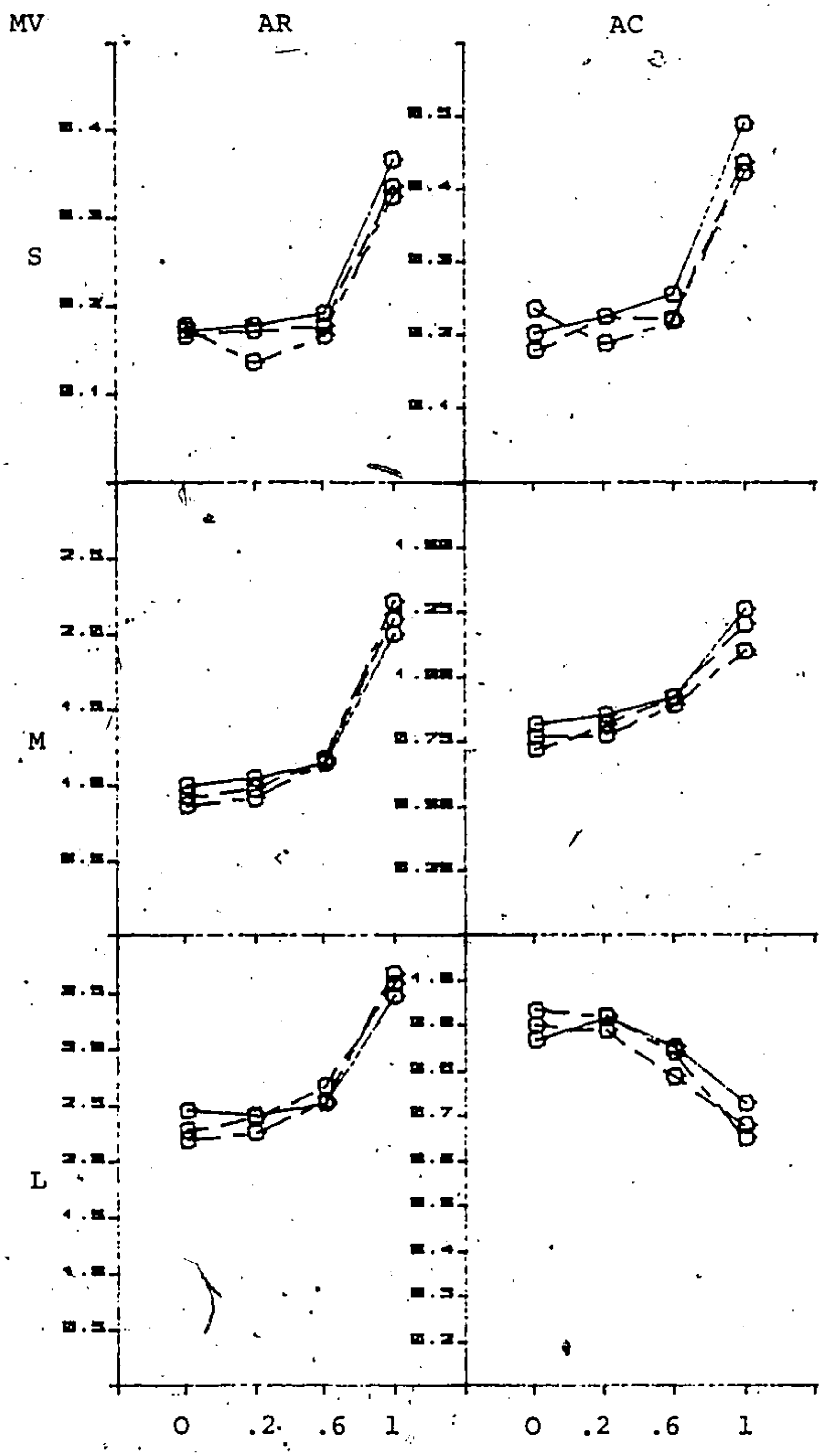
Partition II, Direct, Games and Howell Method



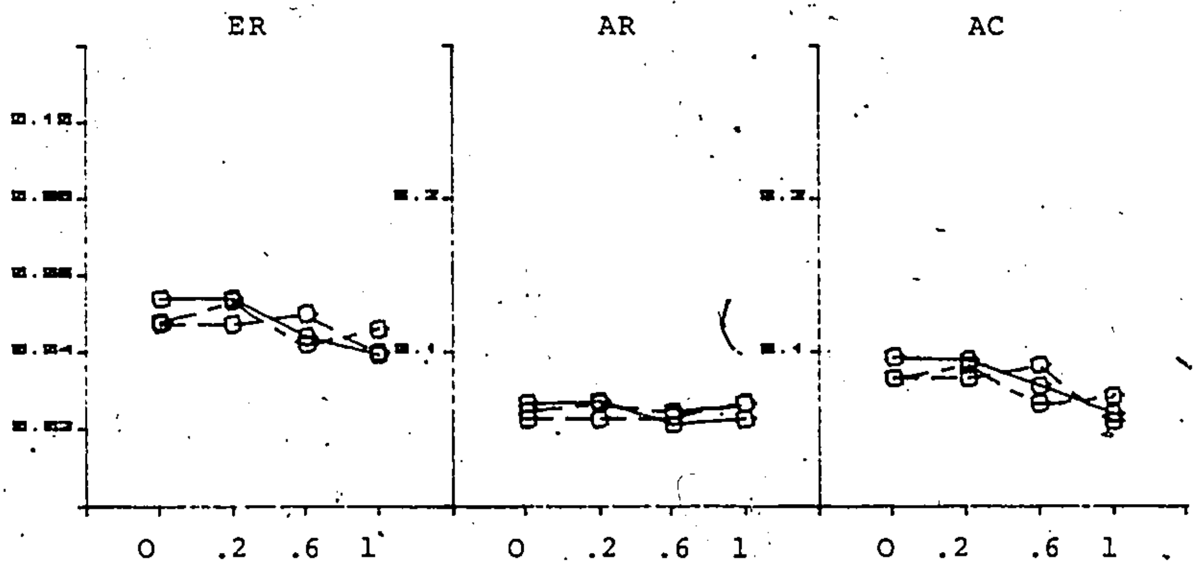
Partition III, Direct, Games and Howell Method



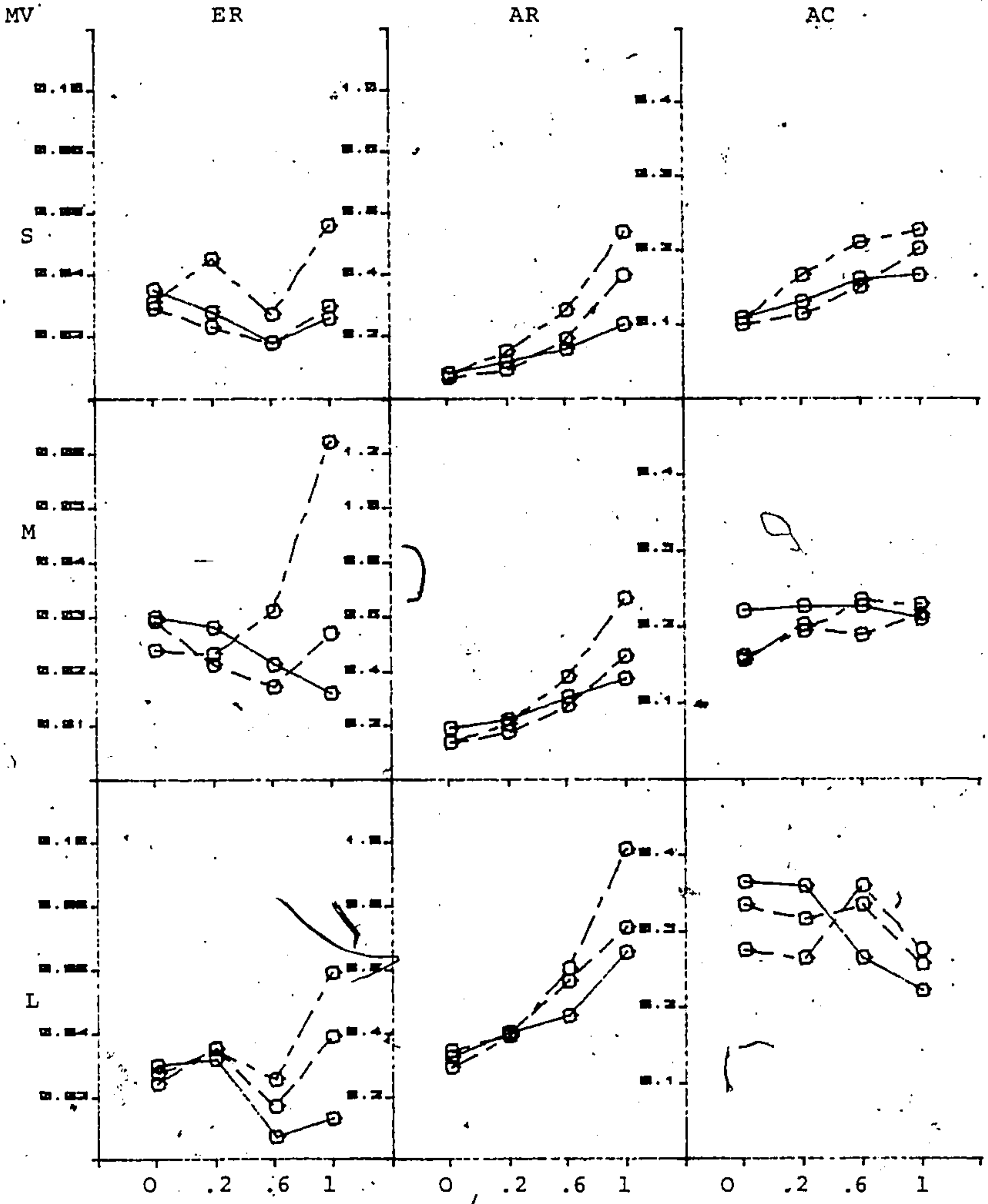
Partition IV, Direct, Games and Howell Method



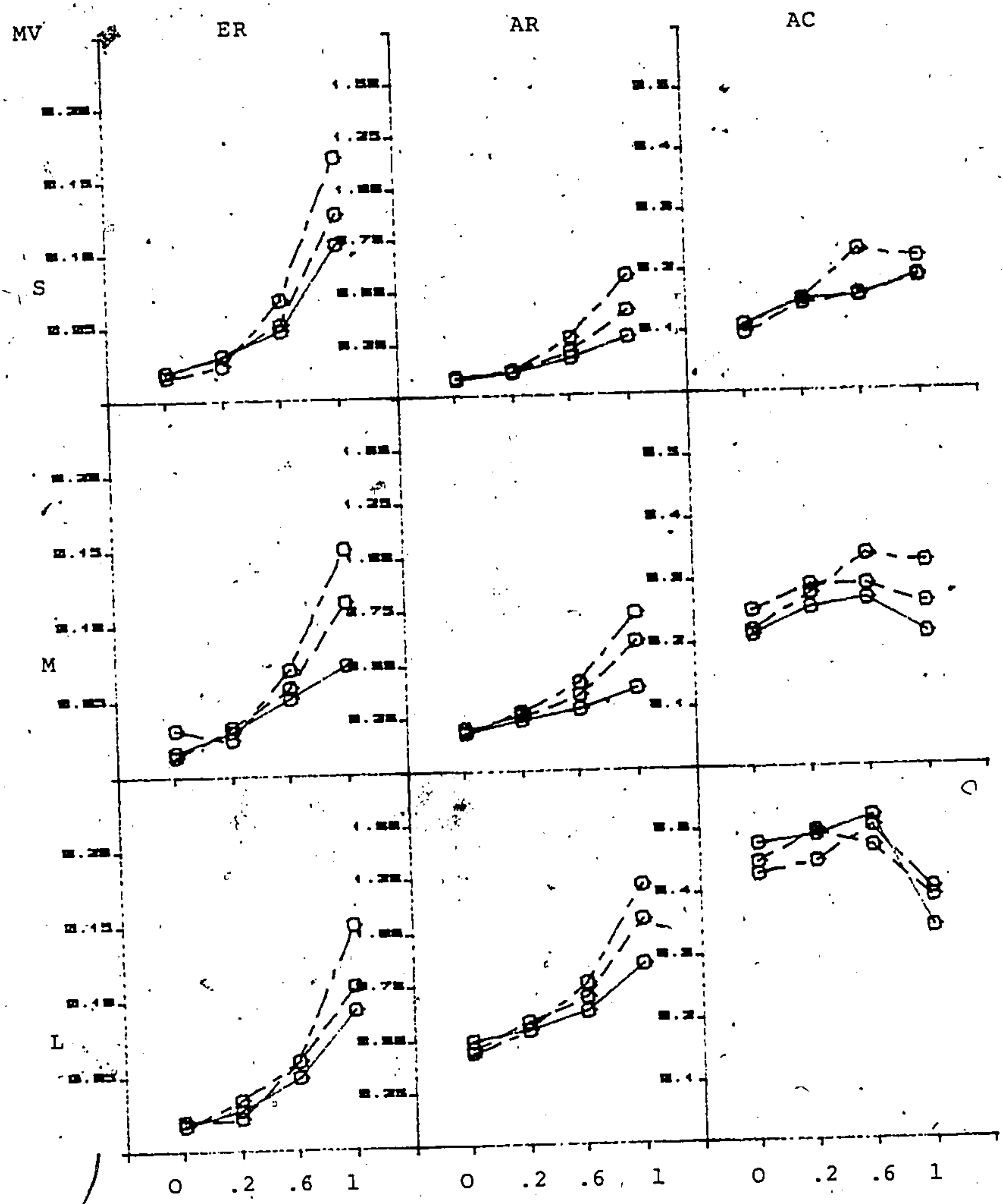
Partition V, Direct, Games and Howell Method



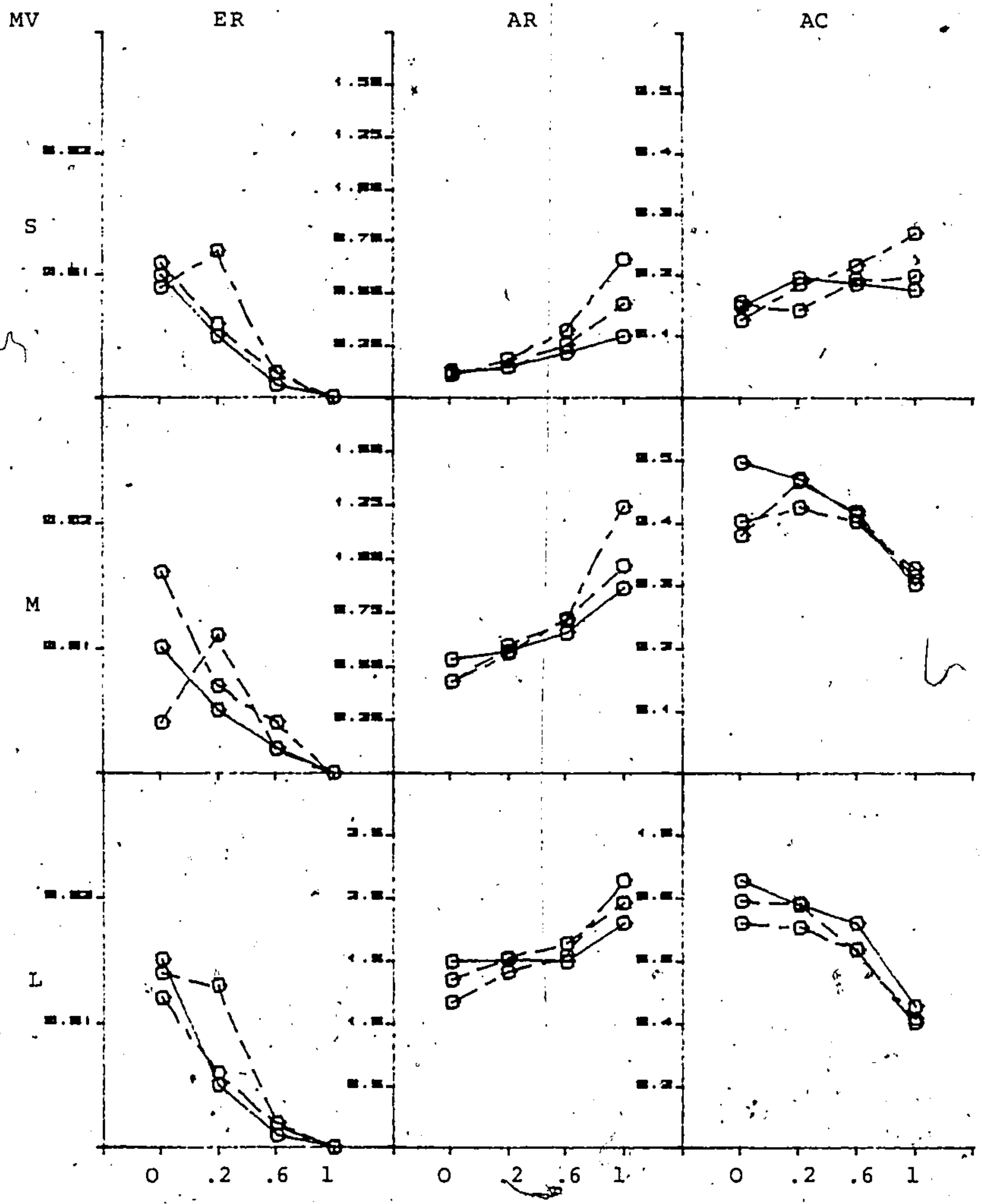
Partition I, Inverse, Kramer Method



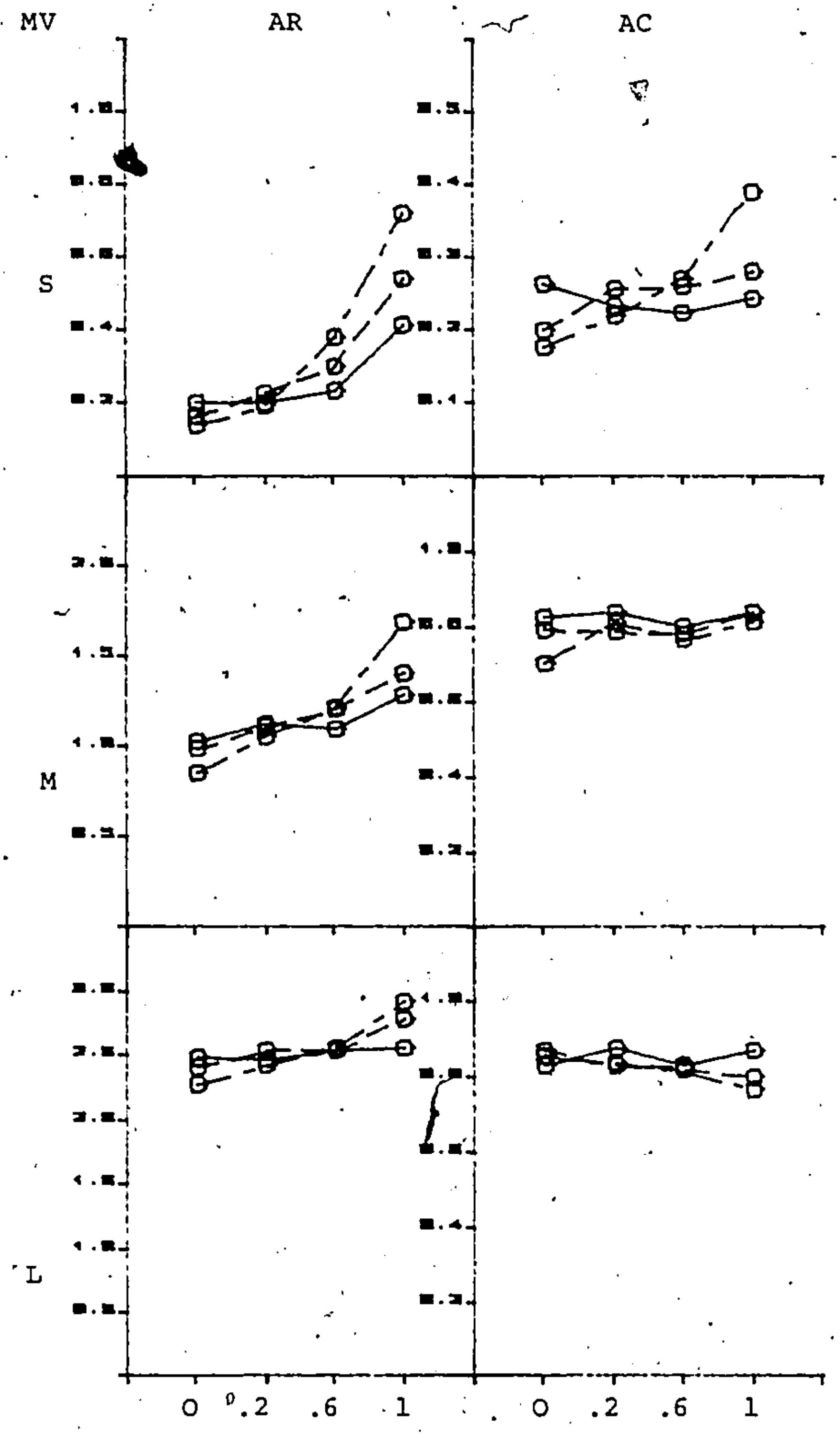
Partition II, Inverse, Kramer Method



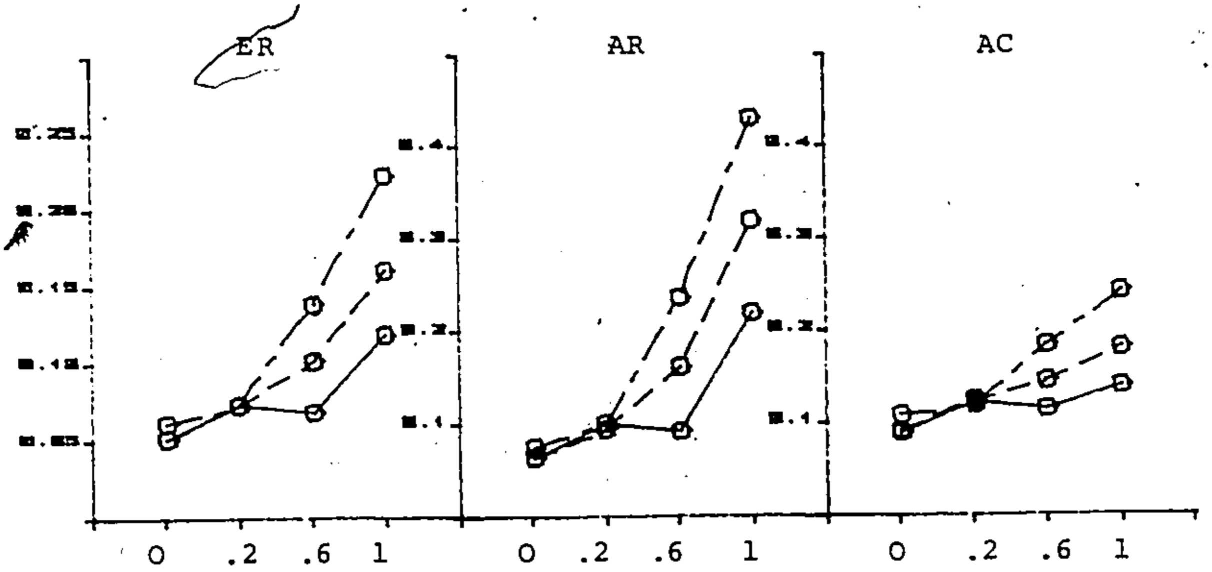
Partition III, Inverse, Kramer Method



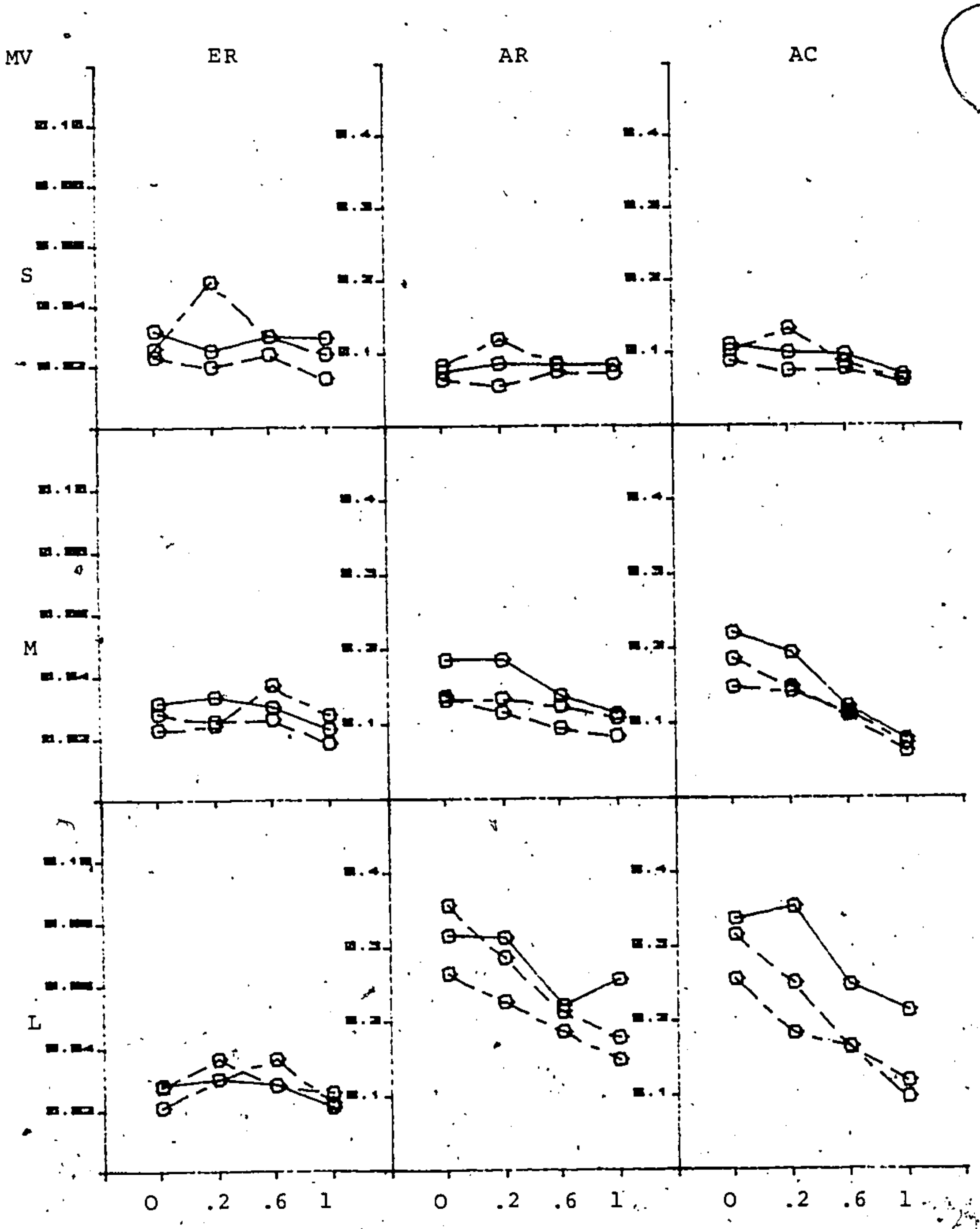
Partition IV, Inverse, Kramer Method



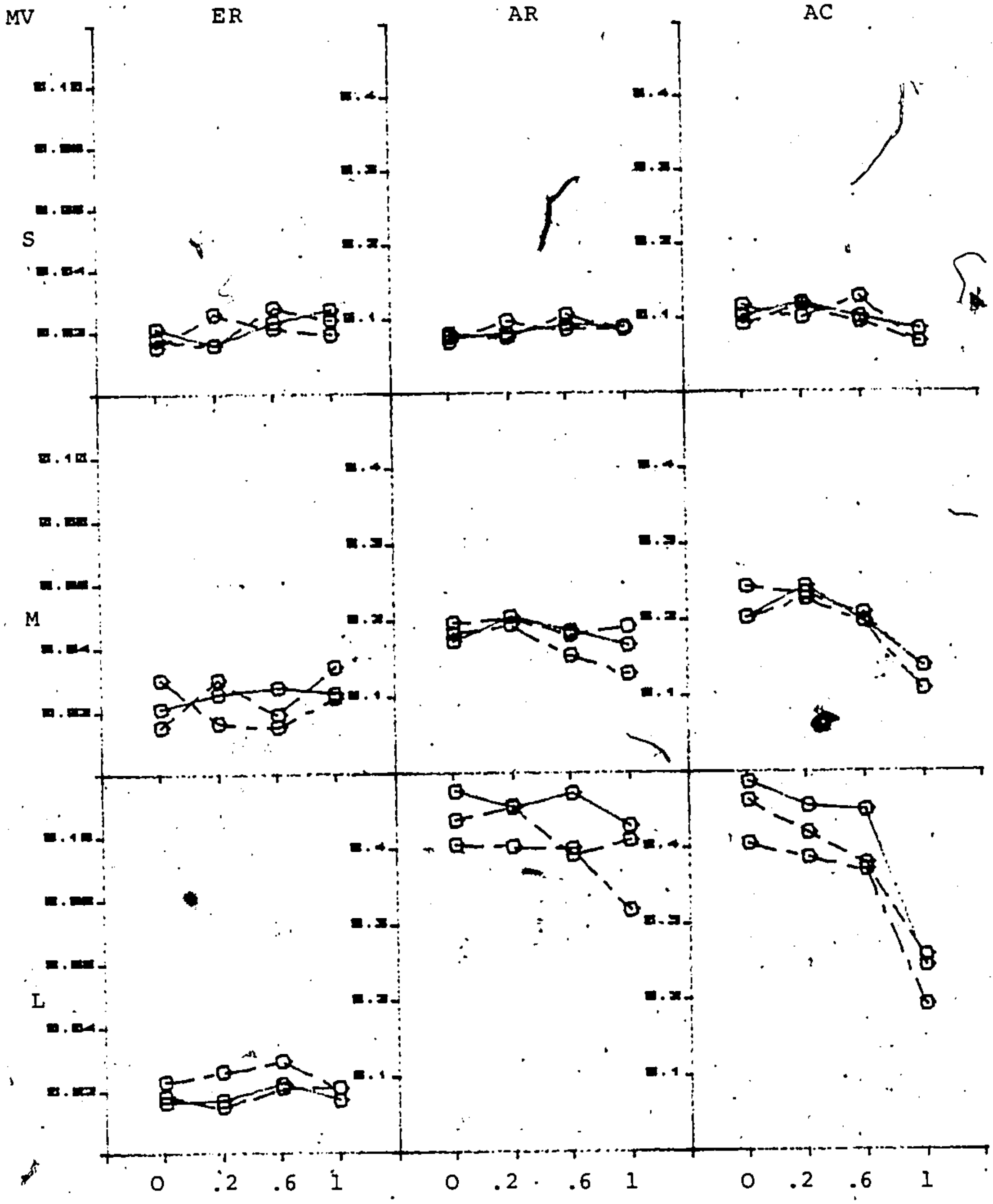
Partition V, Inverse, Kramer Method



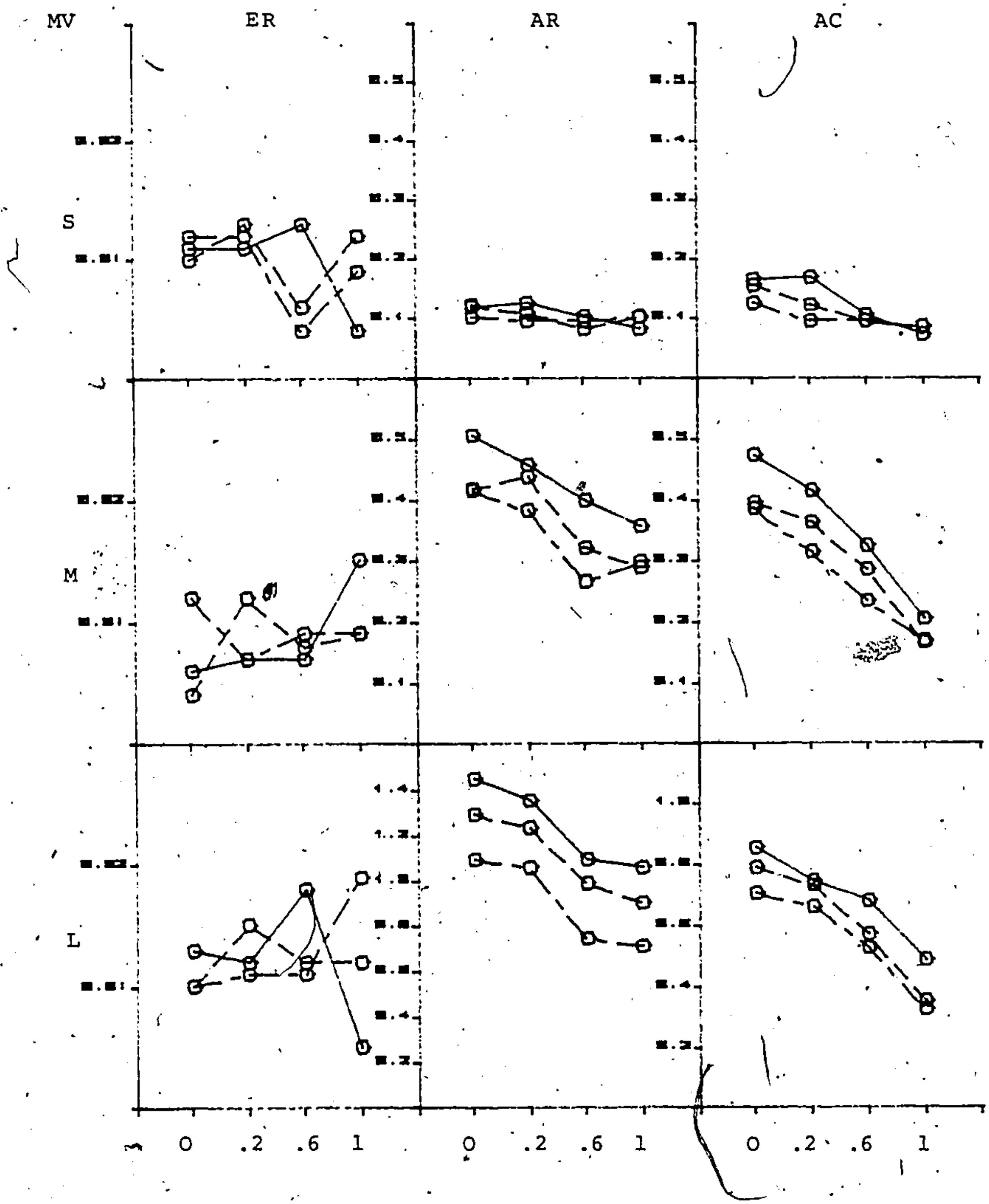
Partition I, Inverse, Games and Howell Method



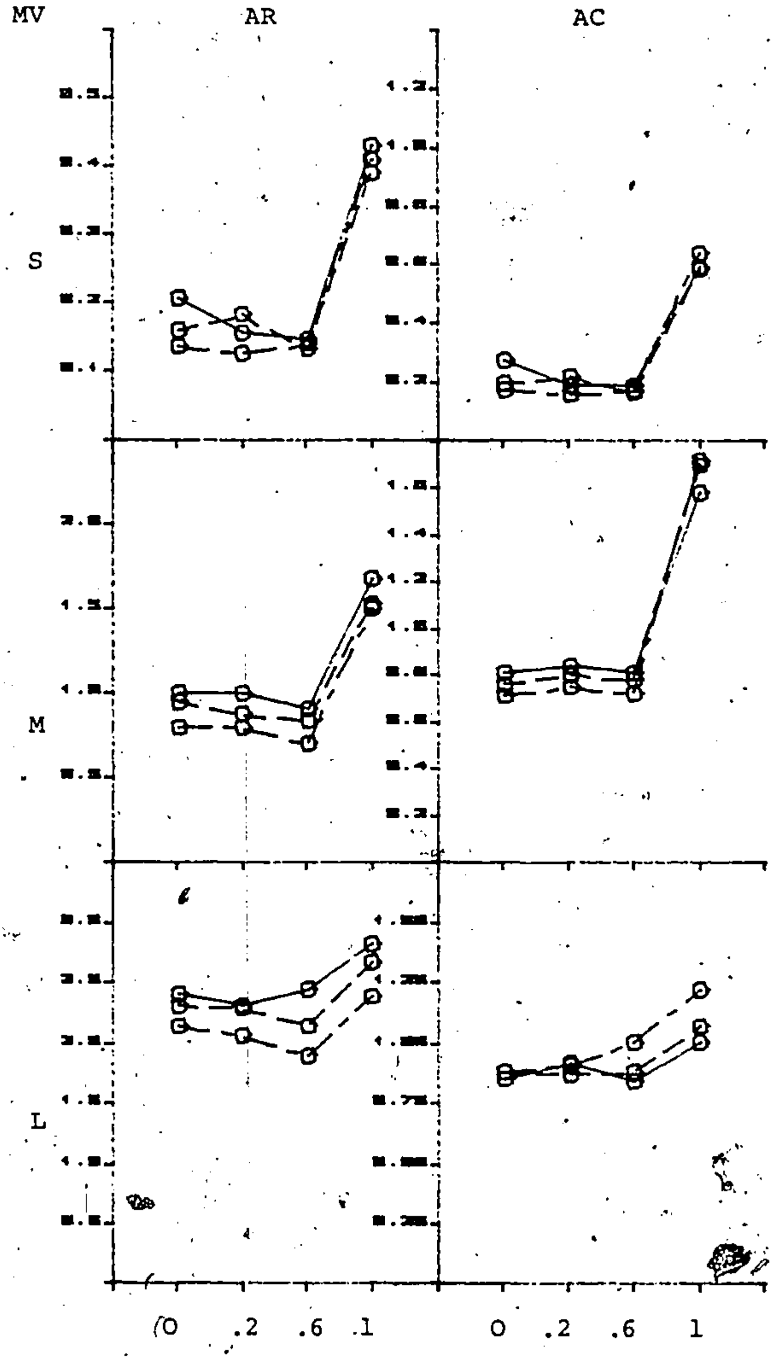
Partition II, Inverse, Games and Howell Method



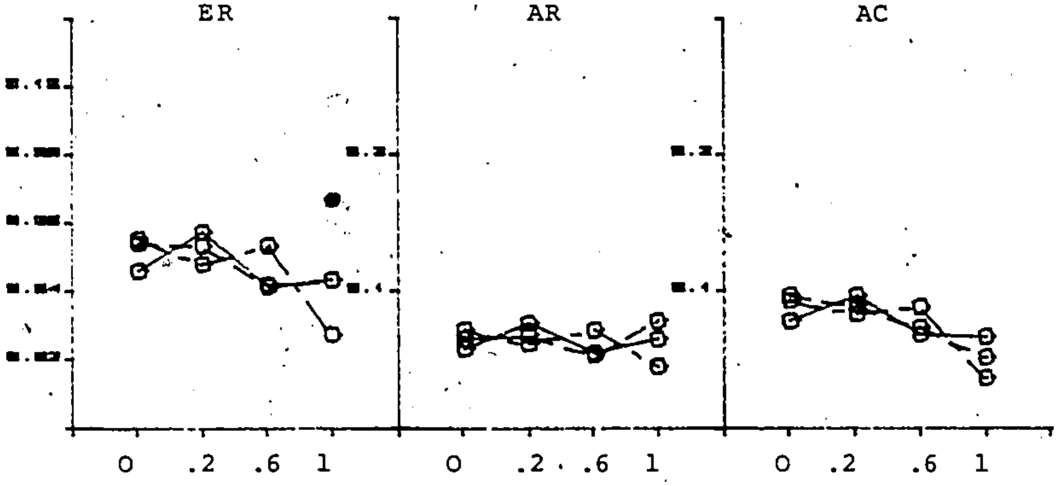
Partition III, Inverse, Games and Howell Method



Partition IV, Inverse, Games and Howell Method



Partition V, Inverse, Games and Howell Method



APPENDIX III

FREQUENCY OF SIGNIFICANT CONTRASTS FOR KRAMER AND GAMES AND HOWELL METHODS

PARTITION I, MEAN VARIABILITY (SMALL)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	10	12	8	12	1	17	0	189
	(13)	17	19	15	22	1	19	4	19
	(14)	12	12	18	24	13	24	8	15
	(23)	8	11	6	7	4	9	6	12
	(24)	14	12	10	9	24	12	7	8
	(34)	10	13	19	17	29	12	54	14
Medium Sample Size Disparity	(12)	15	12	12	21	1	28	0	178
	(13)	13	13	6	14	1	21	1	11
	(14)	12	10	6	14	4	18	2	13
	(23)	16	18	7	10	6	13	1	9
	(24)	6	9	12	13	13	12	2	9
	(34)	9	10	17	11	22	9	24	5
Large Sample Size Disparity	(12)	9	12	5	17	0	16	0	136
	(13)	15	17	9	14	0	16	0	16
	(14)	15	17	5	15	0	11	0	13
	(23)	7	12	8	13	0	9	1	11
	(24)	18	14	14	19	7	8	0	5
	(34)	9	10	19	15	12	5	29	9

PARTITION I, MEAN VARIABILITY (MEDIUM)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	39	52	24	46	8	78	0	946
	(13)	41	39	29	40	9	72	15	87
	(14)	46	47	38	47	36	45	33	48
	(23)	10	13	11	10	1	7	3	10
	(24)	10	11	6	9	20	13	7	8
	(34)	6	8	16	9	26	7	50	12
Medium Sample Size Disparity	(12)	31	33	15	40	8	44	0	924
	(13)	41	43	25	40	5	57	5	65
	(14)	29	34	24	35	12	38	18	53
	(23)	19	16	6	8	2	15	0	7
	(24)	10	11	12	12	13	10	8	17
	(34)	15	16	21	16	25	14	47	12
Large Sample Size Disparity	(12)	24	18	10	36	2	54	0	889
	(13)	25	30	15	33	1	68	1	54
	(14)	45	35	19	38	10	50	2	66
	(23)	11	11	11	16	1	12	1	13
	(24)	12	14	11	11	6	8	0	14
	(25)	8	8	15	11	11	4	43	19

PARTITION I, MEAN VARIABILITY (LARGE)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	102	102	81	124	28	185	0	1000
	(13)	107	105	89	116	41	191	52	163
	(14)	121	120	106	119	88	112	75	101
	(23)	14	14	6	8	3	13	1	6
	(24)	10	9	11	10	26	14	9	12
	(34)	13	18	17	10	30	12	55	9
Medium Sample Size Disparity	(12)	87	81	50	100	22	167	0	1000
	(13)	83	87	82	115	34	191	14	167
	(14)	97	89	82	107	58	132	30	124
	(23)	11	9	4	4	4	13	0	5
	(24)	8	9	15	15	21	20	6	10
	(34)	9	11	16	12	21	11	30	9
Large Sample Size Disparity	(12)	69	69	37	74	8	134	0	1000
	(13)	95	83	51	94	10	173	9	175
	(14)	96	87	66	109	28	140	14	163
	(23)	12	7	9	12	2	9	0	21
	(24)	7	5	7	8	9	11	0	7
	(34)	8	8	8	4	18	12	27	10

PARTITION II, MEAN VARIABILITY (SMALL)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	9	13	3	7	0	10	0	13
	(13)	21	23	7	13	3	20	1	15
	(14)	19	15	11	16	16	17	6	12
	(23)	18	21	10	12	8	20	4	14
	(24)	14	18	18	13	24	17	8	11
	(34)	6	6	21	14	31	11	50	7
Medium Sample Size Disparity	(12)	13	13	10	16	0	13	0	12
	(13)	14	10	8	17	0	17	2	21
	(14)	17	17	8	17	4	12	1	15
	(23)	11	11	20	22	7	16	2	18
	(24)	14	12	21	27	15	13	4	16
	(34)	7	4	12	10	16	8	33	8
Large Sample Size Disparity	(12)	8	8	2	11	0	11	0	12
	(13)	10	12	3	14	0	15	0	23
	(14)	15	13	7	11	1	12	0	19
	(23)	12	11	5	6	3	17	0	17
	(24)	20	18	9	9	12	12	0	18
	(34)	16	15	10	6	15	9	22	8

PARTITION II, MEAN VARIABILITY (MEDIUM)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	6	5	6	11	0	10	0	14
	(13)	36	36	34	59	4	53	15	62
	(14)	44	42	39	49	28	35	31	49
	(23)	55	46	43	48	21	41	14	59
	(24)	60	57	47	42	52	33	39	48
	(34)	9	10	12	10	25	9	48	12
Medium Sample Size Disparity	(12)	15	11	2	5	1	12	0	12
	(13)	52	51	22	37	5	76	3	62
	(14)	40	43	28	39	16	38	15	42
	(23)	51	47	24	37	35	66	4	58
	(24)	53	48	39	40	37	31	19	41
	(34)	13	12	10	7	23	9	33	5
Large Sample Size Disparity	(12)	16	17	4	7	1	14	0	9
	(13)	41	43	16	39	2	56	1	58
	(14)	43	39	22	40	5	47	1	58
	(23)	35	42	30	47	6	38	5	64
	(24)	39	44	47	49	45	54	3	49
	(34)	6	6	17	12	12	6	36	10

PARTITION II, MEAN VARIABILITY (LARGE)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	14	17	4	10	0	11	0	7
	(13)	119	112	100	135	39	200	55	168
	(14)	135	133	109	117	74	96	78	112
	(23)	119	124	115	121	89	141	58	158
	(24)	142	141	123	107	114	78	87	117
	(34)	16	14	17	11	22	6	46	10
Medium Sample Size Disparity	(12)	7	8	3	9	1	13	0	11
	(13)	105	96	74	114	27	188	30	179
	(14)	102	86	87	108	54	122	38	122
	(23)	112	107	96	107	80	138	53	168
	(24)	122	119	116	118	112	98	64	117
	(34)	11	8	17	12	23	10	39	7
Large Sample Size Disparity	(12)	8	6	5	10	0	13	0	11
	(13)	84	81	58	106	6	173	3	176
	(14)	99	93	66	99	27	131	11	141
	(23)	96	88	92	103	46	121	11	169
	(24)	114	115	113	116	86	103	38	140
	(34)	13	16	10	8	15	11	21	2

PARTITION III, MEAN VARIABILITY (SMALL)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	13	13	6	13	2	22	0	189
	(13)	16	16	17	25	4	45	5	47
	(14)	17	20	28	30	27	31	19	29
	(23)	21	19	13	14	6	12	1	15
	(24)	22	16	14	14	26	18	7	11
	(34)	16	10	16	10	25	7	46	6
Medium Sample Size Disparity	(12)	14	15	8	18	0	20	0	180
	(13)	23	27	16	32	1	45	5	52
	(14)	27	25	9	18	10	31	10	38
	(23)	10	11	10	12	6	14	1	26
	(24)	19	20	11	11	18	17	8	16
	(34)	14	12	16	8	19	10	37	13
Large Sample Size Disparity	(12)	15	15	0	4	0	15	0	155
	(13)	31	27	8	28	0	38	1	40
	(14)	24	26	12	25	7	47	1	47
	(23)	11	15	13	14	2	12	0	13
	(24)	12	13	10	16	13	20	1	19
	(34)	12	14	9	9	22	6	35	13

PARTITION III, MEAN VARIABILITY (MEDIUM)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	31	38	16	37	5	71	0	941
	(13)	190	179	173	211	110	364	107	276
	(14)	201	207	186	199	169	196	143	177
	(23)	38	35	46	48	29	53	16	54
	(24)	32	31	47	41	37	27	30	43
	(34)	11	12	17	12	28	14	52	8
Medium Sample Size Disparity	(12)	34	35	21	39	2	56	0	930
	(13)	178	180	117	190	56	297	51	278
	(14)	180	179	132	168	94	185	99	255
	(23)	47	50	35	37	16	40	15	60
	(24)	36	36	29	28	41	34	32	60
	(34)	10	10	11	6	14	9	41	13
Large Sample Size Disparity	(12)	31	28	18	45	2	56	0	873
	(13)	140	140	112	171	31	289	23	305
	(14)	171	149	134	184	52	193	32	239
	(23)	26	28	29	31	10	34	4	63
	(24)	38	39	42	43	23	30	4	50
	(34)	17	13	19	16	15	9	26	6

PARTITION III, MEAN VARIABILITY (LARGE)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	119	119	76	116	42	184	0	1000
	(13)	580	560	557	624	479	799	465	734
	(14)	584	561	542	570	535	572	532	583
	(23)	111	99	119	111	76	119	58	145
	(24)	111	112	123	114	132	95	94	113
	(34)	15	13	18	10	29	11	48	9
Medium Sample Size Disparity	(12)	96	89	51	95	16	179	0	1000
	(13)	505	483	464	557	384	767	337	743
	(14)	540	531	504	563	448	595	394	635
	(23)	119	118	86	98	68	123	43	152
	(24)	122	118	111	112	119	102	75	137
	(34)	10	13	7	4	19	11	36	9
Large Sample Size Disparity	(12)	62	65	37	81	11	135	0	1000
	(13)	434	411	381	492	250	729	215	771
	(14)	446	409	452	532	330	602	253	712
	(23)	113	110	79	110	52	123	31	169
	(24)	119	102	99	97	81	89	45	157
	(34)	8	8	13	10	19	10	30	11

PARTITION IV, MEAN VARIABILITY (SMALL)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	15	17	7	18	3	24	0	193
	(13)	33	33	31	40	7	44	10	52
	(14)	48	52	50	65	52	67	39	61
	(23)	19	18	15	16	6	13	4	19
	(24)	34	37	39	26	54	34	21	28
	(34)	16	16	20	13	35	10	45	14
Medium Sample Size Disparity	(12)	13	16	6	15	0	17	0	165
	(13)	25	24	16	28	4	36	0	38
	(14)	61	61	52	66	25	67	19	73
	(23)	13	15	11	14	8	16	0	15
	(24)	36	36	33	36	30	26	14	33
	(34)	16	15	22	14	32	15	45	12
Large Sample Size Disparity	(12)	18	14	1	7	0	9	0	147
	(13)	31	25	10	23	1	36	1	38
	(14)	61	55	20	37	16	71	6	84
	(23)	18	18	11	18	2	13	1	15
	(24)	37	42	27	33	22	21	3	32
	(34)	23	24	21	17	24	16	34	10

PARTITION IV, MEAN VARIABILITY (MEDIUM)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	41	45	28	52	8	69	0	941
	(13)	190	187	162	213	119	349	102	287
	(14)	496	483	519	534	455	497	445	498
	(23)	47	42	35	41	28	59	12	60
	(24)	188	184	202	165	191	127	153	177
	(34)	45	43	61	35	67	34	91	24
Medium Sample Size Disparity	(12)	42	39	14	34	2	57	0	925
	(13)	176	173	120	174	64	324	51	287
	(14)	464	440	423	485	380	531	341	559
	(23)	42	40	27	36	31	52	13	65
	(24)	199	190	190	184	174	154	130	216
	(34)	51	48	69	56	70	42	101	42
Large Sample Size Disparity	(12)	22	29	6	29	1	51	0	886
	(13)	129	118	89	148	29	279	20	310
	(14)	410	385	343	449	271	560	192	666
	(23)	41	39	29	38	15	48	5	60
	(24)	212	233	183	192	162	174	82	251
	(34)	70	62	57	48	59	43	76	37

PARTITION IV, MEAN VARIABILITY (LARGE)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	112	102	63	108	27	179	0	1000
	(13)	587	562	548	602	488	791	465	718
	(14)	950	942	937	944	911	918	910	931
	(23)	132	130	96	108	79	115	54	148
	(24)	593	571	548	520	520	433	549	593
	(34)	134	126	163	112	139	77	207	68
Medium Sample Size Disparity	(12)	85	79	61	109	15	182	0	1000
	(13)	479	447	457	545	402	797	359	754
	(14)	916	902	928	948	912	952	879	966
	(23)	101	100	85	88	66	127	37	172
	(24)	616	584	583	568	534	495	461	609
	(34)	175	169	161	131	165	106	160	69
Large Sample Size Disparity	(12)	65	60	46	92	5	130	0	1000
	(13)	429	392	414	492	269	750	200	753
	(14)	900	865	859	897	836	951	789	978
	(23)	113	113	93	107	66	146	21	176
	(24)	607	587	525	547	442	459	398	665
	(34)	176	170	141	110	125	95	162	84

PARTITION V

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	9	7	6	13	0	9	0	8
	(13)	15	15	4	9	0	7	0	11
	(14)	10	12	11	16	7	8	6	12
	(23)	7	6	4	6	12	17	0	7
	(24)	6	9	14	10	7	3	7	8
	(34)	16	17	22	14	25	9	39	11
Medium Sample Size Disparity	(12)	12	8	3	5	0	10	0	4
	(13)	11	9	7	10	0	10	0	17
	(14)	11	9	7	11	0	6	1	14
	(23)	13	11	8	11	2	10	0	13
	(24)	10	8	7	7	10	9	4	13
	(34)	13	12	19	13	22	14	40	5
Large Sample Size Disparity	(12)	11	11	6	13	0	8	0	15
	(13)	13	8	4	13	0	8	0	6
	(14)	9	10	4	12	2	13	0	15
	(23)	7	4	11	13	4	8	0	8
	(24)	15	14	8	9	7	11	1	13
	(34)	15	15	13	7	17	14	26	10

PARTITION I, MEAN VARIABILITY (SMALL)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	11	9	30	17	52	8	97	13
	(13)	15	12	29	16	49	17	51	14
	(14)	15	19	25	20	39	20	53	13
	(23)	14	12	16	12	12	12	19	18
	(24)	12	10	15	14	5	16	23	15
	(34)	13	13	3	6	3	10	0	11
Medium Sample Size Disparity	(12)	12	10	26	8	64	17	136	17
	(13)	9	12	21	10	52	14	104	16
	(14)	12	13	19	15	52	13	114	16
	(23)	9	9	8	7	15	12	21	9
	(24)	14	11	9	6	8	17	28	8
	(34)	9	9	9	10	0	2	0	6
Large Sample Size Disparity	(12)	16	18	37	17	74	16	175	17
	(13)	13	16	31	22	93	17	129	18
	(14)	19	19	31	24	82	17	140	13
	(23)	12	10	19	18	20	17	45	10
	(24)	12	9	22	21	13	13	53	11
	(34)	11	10	11	17	2	6	0	13

PARTITION I, MEAN VARIABILITY (MEDIUM)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	43	44	61	42	94	38	146	16
	(13)	52	53	67	48	105	31	94	30
	(14)	60	57	62	60	82	31	100	32
	(23)	11	10	12	13	19	20	15	14
	(24)	11	12	16	17	3	10	13	13
	(34)	10	11	2	7	0	7	0	8
Medium Sample Size Disparity	(12)	30	34	45	19	88	16	156	18
	(13)	37	39	48	32	86	24	117	23
	(14)	33	35	58	40	81	27	134	19
	(23)	9	8	8	9	12	9	25	9
	(24)	6	6	9	8	5	9	24	9
	(34)	18	16	5	9	2	9	0	7
Large Sample Size Disparity	(12)	35	28	59	34	119	28	196	20
	(13)	37	37	58	31	108	24	176	26
	(14)	41	44	57	40	110	30	189	24
	(23)	5	5	11	10	19	14	48	14
	(24)	10	9	12	9	14	9	59	14
	(34)	10	10	6	11	5	20	0	10

PARTITION I, MEAN VARIABILITY (LARGE)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	88	82	115	78	134	63	216	50
	(13)	92	99	117	81	165	50	199	86
	(14)	111	107	135	120	144	77	214	93
	(23)	10	9	19	13	5	6	13	8
	(24)	9	9	16	13	3	9	10	7
	(34)	13	11	2	9	0	16	0	12
Medium Sample Size Disparity	(12)	95	93	108	72	177	54	237	46
	(13)	99	110	127	89	183	59	209	49
	(14)	119	120	120	85	184	69	219	51
	(23)	8	10	18	13	10	9	32	9
	(24)	14	12	11	12	7	9	33	11
	(34)	9	11	9	16	2	13	0	13
Large Sample Size Disparity	(12)	77	71	112	54	178	44	278	30
	(13)	87	82	107	67	196	54	290	44
	(14)	103	89	120	65	201	49	305	43
	(23)	11	8	24	20	14	11	50	12
	(24)	10	9	18	16	13	17	54	12
	(34)	6	7	5	6	2	11	0	8

PARTITION II, MEAN VARIABILITY (SMALL)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	4	6	25	10	43	12	105	13
	(13)	10	14	22	11	61	19	56	11
	(14)	16	15	20	20	38	12	64	11
	(23)	15	15	20	19	19	20	23	20
	(24)	13	11	13	15	7	17	27	20
	(34)	15	15	5	6	4	11	0	14
Medium Sample Size Disparity	(12)	9	7	20	11	50	7	127	7
	(13)	11	10	23	13	67	19	97	17
	(14)	18	17	24	17	55	14	103	19
	(23)	14	16	24	18	13	11	35	16
	(24)	13	12	20	22	13	20	41	15
	(34)	10	8	10	15	2	14	0	12
Large Sample Size Disparity	(12)	9	9	17	7	64	14	165	10
	(13)	21	19	26	9	76	18	121	17
	(14)	13	10	24	13	73	15	132	15
	(23)	14	13	20	18	26	15	65	15
	(24)	17	21	20	18	29	30	79	14
	(34)	8	9	7	9	4	14	0	14

PARTITION II, MEAN VARIABILITY (MEDIUM)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	8	12	24	13	51	13	73	8
	(13)	41	39	54	35	88	33	92	27
	(14)	38	42	46	37	79	32	101	28
	(23)	37	33	58	50	41	47	53	46
	(24)	44	39	53	58	21	48	57	42
	(34)	8	9	6	12	1	14	0	17
Medium Sample Size Disparity	(12)	7	10	26	13	58	13	115	18
	(13)	44	48	68	45	104	24	143	34
	(14)	58	54	54	38	92	27	158	28
	(23)	38	35	55	40	57	49	87	47
	(24)	49	44	44	51	43	63	94	48
	(34)	7	6	9	18	0	6	0	16
Large Sample Size Disparity	(12)	13	12	20	9	67	4	150	5
	(13)	34	32	58	35	117	20	173	24
	(14)	40	40	57	36	106	28	179	22
	(23)	33	36	69	53	63	42	103	30
	(24)	45	46	63	54	49	48	124	31
	(34)	18	18	4	7	3	11	0	19

PARTITION II, MEAN VARIABILITY (LARGE)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	11	10	24	10	47	9	93	13
	(13)	110	105	112	82	169	68	213	76
	(14)	116	120	123	102	172	89	215	74
	(23)	121	117	128	109	130	120	159	131
	(24)	122	116	145	144	106	175	164	133
	(34)	8	7	3	7	1	13	0	4
Medium Sample Size Disparity	(12)	11	9	18	9	59	14	110	10
	(13)	105	104	133	84	179	50	242	62
	(14)	102	95	133	88	194	63	249	61
	(23)	119	112	142	122	123	100	210	135
	(24)	111	110	146	147	131	158	237	133
	(34)	9	9	3	6	1	7	0	11
Large Sample Size Disparity	(12)	11	14	26	12	57	18	150	13
	(13)	83	77	122	70	187	53	273	57
	(14)	100	86	121	86	185	46	292	53
	(23)	111	103	125	103	163	123	236	95
	(24)	124	116	137	117	161	149	261	92
	(34)	6	9	8	14	2	11	0	8

PARTITION III, MEAN VARIABILITY (SMALL)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	16	16	26	11	52	13	99	16
	(13)	32	28	37	26	79	23	68	18
	(14)	27	29	42	38	60	22	79	20
	(23)	17	21	22	21	14	16	22	13
	(24)	18	16	18	20	5	15	20	11
	(34)	10	11	5	11	1	13	0	4
Medium Sample Size Disparity	(12)	15	16	23	14	59	10	125	11
	(13)	28	27	44	24	85	19	117	23
	(14)	35	36	35	22	81	20	136	22
	(23)	16	16	15	14	13	12	33	18
	(24)	18	19	21	23	8	15	35	16
	(34)	11	10	6	13	2	6	0	12
Large Sample Size Disparity	(12)	12	14	30	9	81	16	167	12
	(13)	23	20	49	22	104	21	158	24
	(14)	31	33	57	25	100	27	184	24
	(23)	15	12	19	13	21	13	69	15
	(24)	16	12	17	14	10	14	78	18
	(34)	9	12	12	12	2	4	0	9

PARTITION III, MEAN VARIABILITY (MEDIUM)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	36	31	47	30	83	25	138	19
	(13)	208	193	214	156	244	115	294	114
	(14)	204	199	212	179	238	128	310	123
	(23)	38	39	45	41	56	58	59	44
	(24)	35	35	43	44	30	65	63	43
	(34)	10	8	5	7	2	7	0	15
Medium Sample Size Disparity	(12)	37	41	57	31	100	23	156	18
	(13)	157	152	199	135	253	89	313	92
	(14)	159	150	214	157	267	105	332	92
	(23)	30	32	55	48	48	39	81	38
	(24)	40	41	61	56	36	58	87	39
	(34)	4	4	11	12	2	8	0	9
Large Sample Size Disparity	(12)	19	22	67	35	108	21	207	20
	(13)	159	152	186	123	252	84	361	77
	(14)	147	137	203	140	246	93	381	76
	(23)	43	41	47	34	62	34	135	56
	(24)	46	50	50	43	50	45	157	60
	(34)	16	12	7	7	4	9	0	9

PARTITION III, MEAN VARIABILITY (LARGE)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	97	90	124	88	161	58	230	44
	(13)	555	532	546	476	554	346	612	374
	(14)	578	566	568	525	572	394	621	376
	(23)	120	116	134	119	100	97	156	121
	(24)	120	128	125	128	97	174	174	133
	(34)	15	13	5	12	1	18	0	5
Medium Sample Size Disparity	(12)	98	104	109	73	195	56	258	56
	(13)	492	472	542	414	570	316	628	279
	(14)	509	488	582	471	615	346	655	283
	(23)	109	98	139	113	133	109	196	130
	(24)	121	114	147	146	124	143	221	137
	(34)	14	10	13	15	2	12	0	12
Large Sample Size Disparity	(12)	76	63	113	61	183	46	296	40
	(13)	421	392	474	345	518	215	644	222
	(14)	444	409	516	392	558	242	660	219
	(23)	109	102	151	115	137	90	255	102
	(24)	113	110	147	130	145	138	286	108
	(34)	12	10	6	11	2	11	0	19

PARTITION IV, MEAN VARIABILITY (SMALL)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	18	21	26	11	45	4	114	23
	(13)	43	44	42	29	61	16	96	36
	(14)	67	66	82	64	96	40	135	54
	(23)	22	24	10	9	12	15	20	19
	(24)	28	28	29	23	17	48	51	39
	(34)	25	25	12	22	2	22	0	260
Medium Sample Size Disparity	(12)	12	16	32	19	57	8	143	9
	(13)	31	28	48	25	81	20	110	26
	(14)	54	49	77	52	115	34	180	38
	(23)	17	13	22	23	13	9	37	19
	(24)	40	42	39	44	27	38	70	32
	(34)	13	12	11	19	8	23	0	266
Large Sample Size Disparity	(12)	11	14	33	17	71	10	157	12
	(13)	26	26	35	20	102	18	159	26
	(14)	47	46	78	48	132	32	225	33
	(23)	17	16	14	8	26	17	67	14
	(24)	25	23	22	19	44	43	115	30
	(34)	16	13	11	14	6	19	0	295

PARTITION IV, MEAN VARIABILITY (MEDIUM)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	38	37	72	38	83	31	139	22
	(13)	189	185	224	171	246	114	270	104
	(14)	506	492	531	467	526	338	563	313
	(23)	45	48	56	50	45	43	55	36
	(24)	192	176	195	207	166	276	262	211
	(34)	61	57	42	62	28	100	0	990
Medium Sample Size Disparity	(12)	35	33	63	36	97	24	143	19
	(13)	167	162	211	139	251	94	307	88
	(14)	464	443	503	392	524	285	577	230
	(23)	43	46	53	43	58	42	68	30
	(24)	216	205	241	222	242	294	315	164
	(34)	55	57	33	49	37	95	0	991
Large Sample Size Disparity	(12)	40	31	63	27	102	14	197	25
	(13)	131	128	204	128	259	77	353	79
	(14)	408	365	476	344	500	240	600	187
	(23)	31	25	59	44	64	34	129	42
	(24)	188	191	239	203	244	233	411	171
	(34)	58	61	42	56	41	102	0	991

PARTITION IV, MEAN VARIABILITY (LARGE)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	103	96	139	87	166	64	233	45
	(13)	570	556	565	482	590	375	611	350
	(14)	438	924	934	902	942	859	941	773
	(23)	129	122	113	101	116	119	133	104
	(24)	592	574	587	584	633	739	645	552
	(34)	151	137	119	156	94	283	0	1000
Medium Sample Size Disparity	(12)	89	90	127	77	150	42	241	32
	(13)	516	484	528	417	531	254	639	271
	(14)	920	902	935	867	913	717	931	659
	(23)	120	118	129	112	140	121	216	134
	(24)	617	591	635	613	675	720	739	569
	(34)	134	131	175	211	120	287	16	1000
Large Sample Size Disparity	(12)	83	75	97	54	167	48	267	31
	(13)	442	405	477	324	499	214	638	215
	(14)	876	826	893	799	875	558	906	527
	(23)	104	101	128	107	138	104	250	102
	(24)	595	562	635	573	688	663	770	500
	(34)	172	174	187	205	195	311	76	1000

PARTITION V

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Contrast	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	(12)	12	13	29	16	32	5	95	13
	(13)	10	10	22	14	28	4	45	11
	(14)	9	6	15	12	14	4	48	10
	(23)	11	11	15	11	13	14	16	9
	(24)	13	11	15	15	5	11	15	9
	(34)	8	8	5	8	1	17	0	13
Medium Sample Size Disparity	(12)	6	7	22	10	44	8	116	17
	(13)	12	13	23	13	53	7	72	17
	(14)	12	10	15	10	41	8	81	14
	(23)	6	7	17	16	14	6	26	12
	(24)	13	13	13	13	8	13	24	8
	(34)	15	15	5	7	1	12	0	10
Large Sample Size Disparity	(12)	8	6	37	14	66	10	139	5
	(13)	9	14	21	13	74	14	87	8
	(14)	14	15	19	9	70	14	109	7
	(23)	14	12	9	6	19	11	41	8
	(24)	15	10	7	7	4	11	55	8
	(34)	14	15	7	12	3	11	0	9

APPENDIX IV

FREQUENCY OF OCCURRENCE OF PATTERN TYPES

PARTITION I, MEAN VARIABILITY (SMALL)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	940	932	940	929	946	926	937	779
	2	49	57	45	52	37	56	52	192
	3	0	0	0	0	0	0	0	1
	4	11	11	14	18	16	17	6	21
	5	0	0	1	1	1	1	5	3
	6								0
	7								4
	8								
Medium Sample Size Disparity	1	946	943	952	932	960	928	975	791
	2	39	43	39	55	34	54	22	194
	3	0	0	0	0	0	0	0	0
	4	13	13	7	12	5	17	1	14
	5	1	1	1	0	1	1	2	1
	6	0	0	0	0	0	0	0	0
	7	1	0	0	0	0	0	0	0
	8			1	1	0	0	0	0
Large Sample Size Disparity	1	940	936	958	929	981	945	971	833
	2	47	50	25	52	19	47	28	148
	3	0	0	0	0	0	0	0	0
	4	13	10	16	16	0	6	1	15
	5	0	4	1	3	0	0	0	3
	6	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	2	0	1
	8	0	0	0	0	0	0	0	0

PARTITION I, MEAN VARIABILITY (MEDIUM)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	881	871	908	872	925	829	918	49
	2	89	90	67	104	51	126	58	839
	3	0	0	0	0	0	0	0	4
	4	27	37	18	16	23	39	22	92
	5	3	2	5	6	1	4	2	3
	6	0	0	0	0	0	0	0	3
	7	0	0	2	1	0	2	0	18
	8	0	0	0	1	0	0	0	1
	9								1
Medium Sample Size Disparity	1	895	890	918	882	953	862	942	65
	2	71	74	63	90	33	105	42	817
	3	0	0	0	0	0	0	0	3
	4	29	29	17	23	10	27	12	96
	5	4	5	1	3	3	5	4	3
	6	0	0	0	0	0	0	0	2
	7	0	2	1	2	1	0	0	8
	8	1	0	0	0	0	1	0	0
	9	0	0	0	0	0	0	0	6
Large Sample Size Disparity	1	906	910	939	891	971	857	957	99
	2	70	70	44	77	27	96	39	775
	3	0	0	0	0	0	0	0	5
	4	21	15	14	28	2	41	4	99
	5	2	3	2	3	0	5	0	3
	6	0	0	0	0	0	0	0	3
	7	1	1	1	1	0	1	0	10
	8	0	1	0	0	0	0	0	1
	9	0	0	0	0	0	0	0	5

PARTITION I, MEAN VARIABILITY (LARGE)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	753	754	782	739	837	608	856	0
	2	161	157	143	164	116	273	102	745
	3	0	0	0	0	0	0	0	0
	4	54	58	58	70	41	104	36	225
	5	25	25	13	21	5	13	4	15
	6	0	0	0	0	0	0	0	3
	7	5	4	2	4	1	1	2	6
	8	1	1	2	2	0	1	0	1
	9	1	1	0	0	0	0	0	5
Medium Sample Size Disparity	1	799	805	810	752	876	615	935	0
	2	124	123	139	158	95	266	53	723
	3	0	0	0	0	0	0	0	0
	4	60	53	43	75	22	90	9	242
	5	17	18	8	15	5	21	3	22
	6	0	0	0	0	0	0	0	2
	7	0	1	0	0	2	7	0	8
	8	0	0	0	0	0	1	0	0
	9	0	0	0	0	0	0	0	3
Large Sample Size Disparity	1	821	830	863	790	936	664	960	0
	2	99	106	104	138	55	215	30	678
	3	0	0	0	0	0	0	0	0
	4	53	40	25	53	7	99	10	273
	5	24	23	6	15	2	21	0	27
	6	0	0	0	0	0	0	0	8
	7	2	0	2	4	0	1	0	9
	8	0	0	0	0	0	0	0	2
	9	1	1	0	0	0	0	0	3

PARTITION II, MEAN VARIABILITY (SMALL)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	930	924	945	941	936	933	944	954
	2	54	57	43	45	51	46	46	24
	3	0	0	0	0	0	0	0	0
	4	15	18	9	12	8	14	7	19
	5	0	0	3	2	5	6	3	2
	6	0	0	0	0	0	0	0	0
	7	1	1	0	0	0	1	0	0
	8								0
	9								1
Medium Sample Size Disparity	1	940	945	936	918	961	937	966	944
	2	45	45	51	59	36	48	29	26
	3	0	0	0	0	0	0	0	0
	4	14	8	11	19	3	14	2	26
	5	1	2	2	4	0	1	3	4
	6	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	0	0
	9	0	0	0	0	0	0	0	0
Large Sample Size Disparity	1	939	939	968	949	970	939	978	936
	2	45	47	28	45	29	46	22	35
	3	0	0	0	0	0	0	0	0
	4	12	12	4	6	1	15	0	25
	5	3	1	0	0	0	0	0	4
	6	0	0	0	0	0	0	0	0
	7	1	1	0	0	0	0	0	0
	8	0	0	0	0	0	0	0	0
	9	0	0	0	0	0	0	0	0

PARTITION II, MEAN VARIABILITY (MEDIUM)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	835	846	860	835	901	862	914	863
	2	124	117	102	119	74	100	43	46
	3	0	0	0	0	0	0	0	0
	4	37	32	35	39	19	34	25	78
	5	3	4	1	3	5	3	18	10
	6	0	0	0	0	0	0	0	0
	7	1	1	2	3	0	0	0	0
	8				1	1	1	0	3
Medium Sample Size Disparity	1	843	850	905	872	911	825	949	869
	2	97	95	67	94	65	127	33	49
	3	0	0	0	0	0	0	0	0
	4	53	49	26	31	20	40	13	76
	5	4	5	2	2	4	7	5	3
	6	0	0	0	0	0	0	0	0
	7	3	0	0	1	0	0	0	2
	8	0	1	0	0	0	1	0	1
Large Sample Size Disparity	1	871	867	893	856	935	831	960	857
	2	84	83	86	105	59	131	35	54
	3	0	0	0	0	0	0	0	0
	4	39	43	16	31	6	30	4	77
	5	5	5	2	2	0	6	1	7
	6	0	0	0	0	0	0	0	0
	7	1	1	0	3	0	2	0	1
	8	0	1	3	3	0	0	0	4

PARTITION II, MEAN VARIABILITY (LARGE)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	660	656	694	677	775	654	833	691
	2	185	193	185	190	130	191	50	80
	3	0	0	0	0	0	0	0	0
	4	115	114	91	99	79	129	80	204
	5	13	12	9	7	9	9	30	9
	6	0	0	0	0	0	0	0	1
	7	17	16	10	16	5	10	4	6
	8	10	9	11	10	2	7	3	9
	9				1				
Medium Sample Size Disparity	1	695	716	743	703	787	632	864	685
	2	191	178	145	168	147	206	69	80
	3	0	0	0	0	0	0	0	0
	4	84	80	90	93	51	133	50	202
	5	8	8	3	7	6	7	12	5
	6	0	0	0	0	0	0	0	1
	7	12	10	15	23	6	12	1	6
	8	10	8	4	6	2	9	4	21
	9	0	0	0	0	1	1	0	0
Large Sample Size Disparity	1	729	743	762	705	857	647	941	663
	2	159	138	150	176	111	200	38	83
	3	0	0	0	0	0	0	0	0
	4	86	102	72	96	27	118	17	222
	5	7	4	6	5	1	8	4	2
	6	0	0	0	0	0	0	0	0
	7	14	7	8	13	4	16	0	14
	8	4	5	2	5	0	11	0	15
	9	1	1	0	0	0	0	0	1

PARTITION III, MEAN VARIABILITY (SMALL)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	920	931	924	914	936	889	941	751
	2	58	46	63	70	42	91	44	207
	3	0	0	0	0	0	0	0	1
	4	19	21	8	12	18	17	11	36
	5	5	1	3	1	3	2	4	2
	6	0	0	0	0	0	0	0	1
	7	2	1	2	3	1	0	0	1
	8	0	0	0	0	0	1	0	1
Medium Sample Size Disparity	1	922	917	938	918	958	887	953	740
	2	52	59	55	67	32	92	37	211
	3	0	0	0	0	0	0	0	0
	4	23	21	6	13	8	18	6	36
	5	3	3	1	1	2	3	4	7
	6	0	0	0	0	0	0	0	2
	7	0	0	0	1	0	0	0	2
	8	0	0	0	0	0	0	0	1
	9								0
	10								1
Large Sample Size Disparity	1	924	918	958	923	964	895	962	776
	2	53	60	36	62	30	78	38	168
	3	0	0	0	0	0	0	0	1
	4	17	16	3	12	4	21	0	49
	5	6	5	0	0	2	6	0	1
	6	0	0	0	0	0	0	0	1
	7	0	1	2	2	0	0	0	3
	8	0	0	1	1	0	0	0	1

PARTITION III, MEAN VARIABILITY (MEDIUM)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	672	672	693	660	742	522	764	38
	2	183	182	179	183	168	281	146	576
	3	0	0	0	0	0	0	0	0
	4	118	121	86	116	66	154	69	264
	5	18	18	17	17	12	27	18	35
	6	0	0	0	0	0	0	0	32
	7	6	4	17	14	6	9	2	28
	8	3	3	8	10	6	6	1	0
	9						1		25
	10								2
Medium Sample Size Disparity	1	690	690	758	692	828	572	838	37
	2	179	169	154	176	126	268	100	519
	3	0	0	0	0	0	0	0	0
	4	91	104	74	105	42	131	49	286
	5	23	21	9	22	2	20	11	49
	6	0	0	0	0	0	0	0	39
	7	13	14	4	4	1	5	1	41
	8	4	2	1	1	1	3	1	0
	9						1		27
	10								2
Large Sample Size Disparity	1	724	744	762	687	897	566	934	67
	2	158	142	143	169	75	280	46	496
	3	0	0	0	0	0	0	0	0
	4	90	88	76	112	27	134	17	302
	5	21	17	13	23	0	12	3	44
	6	0	0	0	0	0	0	0	33
	7	6	8	4	8	0	5	0	27
	8	1	1	2	1	1	2	0	2
	9						1		29
	10								

PARTITION III, MEAN VARIABILITY (LARGE)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	255	269	261	217	270	96	264	0
	2	213	229	246	248	324	303	382	96
	3	0	0	0	0	0	0	0	0
	4	321	305	323	342	271	360	253	387
	5	102	98	75	98	46	117	27	272
	6	0	1	0	0	0	0	0	59
	7	77	64	62	61	67	90	68	39
	8	30	31	33	34	22	25	5	0
	9	2	3	0	0	0	9	1	131
	10								16
Medium Sample Size Disparity	1	305	316	318	247	373	116	412	0
	2	217	213	276	264	297	271	368	94
	3	0	0	0	0	0	0	0	0
	4	295	304	291	325	247	374	148	357
	5	84	76	42	77	23	125	25	289
	6	0	0	0	0	46	0	0	43
	7	63	61	53	64	14	75	42	28
	8	31	27	19	23	0	32	5	0
	9	5	3	1	0	0	5	0	157
	10						2		32
Large Sample Size Disparity	1	382	414	404	302	534	155	598	0
	2	218	215	245	250	245	243	267	71
	3	0	0	0	0	0	0	0	0
	4	261	244	258	296	175	393	100	303
	5	54	55	33	70	16	93	15	329
	6	0	1	0	0	0	1	0	33
	7	60	50	39	58	20	84	18	29
	8	22	18	20	20	9	27	2	0
	9	3	3	1	4	1	3	0	205
	10						1		30

PARTITION IV, MEAN VARIABILITY (SMALL)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	882	876	883	866	888	851	923	718
	2	79	83	81	97	74	110	48	213
	3	0	0	0	0	0	0	0	0
	4	31	33	27	31	31	35	16	56
	5	6	5	6	4	6	2	12	6
	6	0	0	0	0	0	0	0	1
	7	2	3	3	1	1	2	1	3
	8	0	0	0	1	0	0	0	1
	9								2
Medium Sample Size Disparity	1	881	885	893	868	925	867	941	745
	2	85	76	78	95	55	95	45	185
	3	0	0	0	0	0	0	0	0
	4	24	27	25	34	16	32	9	59
	5	7	10	2	1	4	4	5	4
	6	0	0	0	0	0	0	0	0
	7	2	1	2	1	0	2	0	7
	8	1	1	0	1	0	0	0	0
Large Sample Size Disparity	1	863	862	925	892	952	872	964	752
	2	94	103	60	82	35	94	30	181
	3	0	0	0	0	0	0	0	0
	4	36	30	15	25	9	30	3	56
	5	6	4	0	1	4	3	3	6
	6	0	0	0	0	0	0	0	1
	7	0	1	0	0	0	1	0	4
	8	1	0	0	0	0	0	0	0

PARTITION IV, MEAN VARIABILITY (MEDIUM)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	414	417	410	388	460	338	498	24
	2	274	283	282	294	288	306	276	345
	3	0	0	0	0	0	0	0	0
	4	223	214	213	225	183	254	152	332
	5	44	42	43	42	44	41	55	83
	6	0	0	0	0	0	0	0	28
	7	26	29	39	34	18	46	18	115
	8	17	13	12	14	7	14	1	0
	9	1	2	0	3	0	1	0	65
	10	1	0	1	0	0	0	0	8
Medium Sample Size Disparity	1	443	467	479	421	549	328	618	27
	2	256	240	268	286	247	306	202	309
	3	0	0	0	0	0	0	0	0
	4	197	199	190	206	146	260	109	334
	5	47	40	32	44	38	45	60	80
	6	1	0	0	1	0	2	0	21
	7	44	44	25	33	12	43	8	115
	8	12	10	4	8	8	13	3	5
	9	0	0	2	0	0	3	0	96
	10	0	0	0	1	0	0	0	13
Large Sample Size Disparity	1	474	467	559	452	649	336	771	27
	2	258	274	236	277	201	289	126	240
	3	0	0	0	0	0	0	0	1
	4	193	196	149	195	115	274	62	359
	5	32	29	28	37	28	47	38	114
	6	0	0	0	0	0	0	0	12
	7	34	23	23	30	6	39	1	125
	8	11	10	5	8	1	13	2	7
	9	0	1	0	1	0	2	0	106
	10	0	0	0	0	0	0	0	9

PARTITION IV, MEAN VARIABILITY (LARGE)

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
	1	27	32	33	28	61	26	54	0
	2	118	128	155	150	195	111	217	18
	3	0	0	0	0	0	0	0	0
Small Sample Size Disparity	4	333	362	369	363	350	331	278	126
	5	157	135	129	124	120	153	170	202
	6	0	1	0	0	0	0	0	13
	7	221	202	187	200	189	239	222	124
	8	94	90	73	78	59	78	31	0
	9	36	43	48	51	24	58	28	414
	10	14	7	6	6	2	4	0	103
	1	38	51	49	35	62	16	68	0
	2	167	174	157	138	226	86	311	5
	3	0	0	0	0	0	0	0	0
Medium Sample Size Disparity	4	322	336	359	351	348	315	308	107
	5	146	153	154	149	131	161	135	205
	6	0	2	0	2	0	0	0	6
	7	196	176	190	215	155	244	148	115
	8	72	64	57	61	47	95	22	0
	9	48	34	30	43	29	67	8	437
	10	11	10	4	6	2	16	0	125
	1	46	59	97	61	133	22	179	0
	2	178	192	192	188	294	116	319	6
	3	0	0	0	0	0	0	0	0
Large Sample Size Disparity	4	338	353	344	319	321	345	270	84
	5	140	136	116	125	104	134	136	199
	6	1	1	1	1	0	0	0	5
	7	181	163	161	190	98	219	82	114
	8	72	61	53	74	41	107	12	0
	9	38	30	28	35	8	47	1	432
	10	6	5	8	7	1	10	1	160

PARTITION V

DIRECT PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	950	946	953	946	959	956	958	961
	2	39	43	35	43	31	35	34	22
	3	0	0	0	0	0	0	0	0
	4	9	10	10	8	10	9	6	16
	5	2	0	1	2	0	0	2	1
	6	0	0	0	0	0	0	0	0
	7	0	1	1	1	0	0	0	0
	8	0	0	0	0	0	0	0	0
Medium Sample Size Disparity	1	946	953	959	953	969	950	960	960
	2	40	39	32	38	28	42	36	18
	3	0	0	0	0	0	0	0	0
	4	12	6	8	8	3	7	3	19
	5	2	2	1	1	0	1	1	2
	6	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	0	1
Large Sample Size Disparity	1	944	952	957	947	976	958	974	954
	2	43	36	41	42	19	29	25	31
	3	0	0	0	0	0	0	0	0
	4	12	11	1	8	4	9	1	9
	5	0	0	1	3	0	1	0	6
	6	0	0	0	0	0	0	0	0
	7	1	0	0	0	1	0	0	0
	8	0	0	0	0	0	3	0	0
	9		1	0	0				

PARTITION I, MEAN VARIABILITY (SMALL)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	937	938	917	938	894	940	863	949
	2	47	49	52	43	69	41	70	24
	3	0	0	0	0	0	0	0	0
	4	15	13	27	15	20	15	28	21
	5	1	0	4	4	15	3	39	6
	6	0	0	0	0	0	0	0	0
	7	0	0	0	0	2	1	0	0
	8	0	0	0	0	0	0	0	0
	9	0	0	0	0	0	0	0	0
Medium Sample Size Disparity	1	945	949	932	956	886	949	800	956
	2	45	39	49	33	56	32	74	24
	3	0	0	0	0	0	0	0	0
	4	10	11	14	10	39	14	55	13
	5	0	1	4	1	19	5	67	6
	6	0	0	0	0	0	0	0	0
	7	0	0	1	0	0	0	1	0
	8	0	0	0	0	0	0	0	0
	9	0	0	0	0	0	0	0	1
	10							3	0
Large Sample Size Disparity	1	936	938	895	917	835	943	751	953
	2	47	43	69	55	79	36	77	24
	3	0	0	0	0	0	0	0	0
	4	15	18	27	20	53	13	70	15
	5	2	1	7	6	32	8	87	6
	6	0	0	0	0	0	0	0	0
	7	0	0	1	2	1	0	1	0
	8	0	0	1	0	0	0	6	0
	9	0	0	0	0	0	0	3	0
	10							5	2

PARTITION I, MEAN VARIABILITY (MEDIUM)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	865	865	851	873	824	914	807	936
	2	93	93	94	82	88	47	82	23
	3	0	0	0	0	0	0	0	0
	4	32	32	39	30	49	27	47	33
	5	7	7	16	15	38	12	64	8
	6	0	0	0	0	0	0	0	0
	7	3	3	0	0	1	0	0	0
	8	0	0	0	0	0	0	0	0
	9	0	0	0	0	0	0	0	0
Medium Sample Size Disparity	1	903	892	875	912	846	932	778	951
	2	67	82	85	64	71	48	78	23
	3	0	0	0	0	0	0	0	0
	4	24	22	33	20	46	14	58	16
	5	5	3	6	3	36	6	83	10
	6	0	0	0	0	0	0	0	0
	7	1	1	0	0	1	0	0	0
	8	0	0	0	0	0	0	1	0
	9	0	0	1	1	0	0	1	0
	10							1	
Large Sample Size Disparity	1	900	906	868	908	800	921	708	940
	2	70	64	78	61	86	45	83	24
	3	0	0	0	0	0	0	0	0
	4	22	21	37	19	55	23	61	24
	5	8	9	16	11	51	8	132	12
	6	0	0	0	0	0	0	0	0
	7	0	0	1	1	6	2	2	0
	8	0	0	0	0	0	0	7	0
	9	0	0	0	0	2	1	2	0
	10	0	0	0	0	0	0	5	0

PARTITION I, MEAN VARIABILITY (LARGE)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	771	772	747	787	763	871	711	874
	2	153	155	137	134	100	66	74	36
	3	0	0	0	0	0	0	0	0
	4	58	57	82	58	63	35	74	54
	5	17	15	30	20	71	24	136	33
	6	0	0	0	0	0	0	0	1
	7	1	1	3	0	1	3	0	0
	8	0	0	0	0	0	0	3	0
	9	0	0	1	1	1	1	0	0
	10					1		2	2
Medium Sample Size Disparity	1	775	777	766	823	703	875	676	901
	2	135	123	121	98	125	64	87	44
	3	0	0	0	0	0	0	0	0
	4	62	68	68	50	80	35	79	32
	5	25	29	38	24	87	24	150	22
	6	0	0	0	0	0	0	0	0
	7	2	2	6	3	3	1	0	0
	8	1	1	0	1	1	1	1	0
	9	0	0	1	1	1	0	4	0
	10	0	0	0	0	0	0	3	1
Large Sample Size Disparity	1	811	826	785	864	679	883	594	918
	2	109	104	98	70	130	67	80	32
	3	0	0	0	0	0	0	0	0
	4	55	48	63	40	99	31	112	34
	5	22	22	48	23	91	19	192	15
	6	0	0	0	0	0	0	0	0
	7	3	0	6	3	1	0	1	0
	8	0	0	0	0	0	0	8	1
	9	0	0	0	0	0	0	3	0
	10	0	0	0	0	0	0	10	0

PARTITION II, MEAN VARIABILITY (SMALL)

INVERSE PAIRING

	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	939	940	915	933	892	937	848	943
	2	49	45	69	56	58	42	73	32
	3	0	0	0	0	0	0	0	0
	4	12	14	12	8	36	14	37	18
	5	0	1	4	3	12	6	38	5
	6	0	0	0	0	0	0	0	0
	7	0	0	0	0	2	1	2	2
	8	0	0	0	0	0	0	2	0
Medium Sample Size Disparity	1	944	947	912	929	882	940	807	949
	2	42	40	63	49	59	38	59	22
	3	0	0	0	0	0	0	0	0
	4	11	11	17	19	37	19	61	23
	5	1	0	8	3	19	3	67	6
	6	0	0	0	0	0	0	0	0
	7	0	0	0	0	2	0	4	0
	8	2	2	0	0	1	0	1	0
	9							0	
	10							1	
Large Sample Size Disparity	1	936	934	912	942	839	922	747	950
	2	49	52	68	44	91	57	66	23
	3	0	0	0	0	0	0	0	0
	4	12	13	14	12	46	14	80	21
	5	3	1	5	2	26	7	93	4
	6	0	0	0	0	0	0	0	0
	7	0	0	1	0	1	0	2	0
	8	0	0	0	0	2	0	5	2
	9							4	0
	10							3	0

PARTITION II, MEAN VARIABILITY (MEDIUM)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	873	875	837	851	819	872	808	898
	2	86	86	98	100	103	76	57	40
	3	0	0	0	0	0	0	0	0
	4	33	29	53	42	57	45	88	58
	5	3	7	8	5	18	6	37	3
	6	0	0	0	0	0	0	0	0
	7	5	3	3	2	2	1	8	1
	8	0	0	1	0	1	0	2	0
Medium Sample Size Disparity	1	849	853	820	854	789	871	722	891
	2	106	99	120	97	107	84	70	38
	3	0	0	0	0	0	0	0	0
	4	39	46	44	39	69	37	110	60
	5	5	2	11	7	26	6	78	8
	6	0	0	0	0	0	0	0	0
	7	0	0	5	3	5	2	7	3
	8	1	0	0	0	3	0	12	0
	9					1		1	
Large Sample Size Disparity	1	867	870	819	861	755	884	656	921
	2	92	86	110	94	128	84	84	32
	3	0	0	0	0	0	0	0	0
	4	32	34	55	36	77	28	152	42
	5	8	9	11	6	34	3	89	4
	6	0	0	0	0	0	0	0	1
	7	1	1	2	2	3	0	3	0
	8	0	0	3	1	3	1	13	0
	9							2	
	10							1	

PARTITION II, MEAN VARIABILITY (LARGE)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	685	686	670	707	629	699	621	778
	2	170	182	183	175	175	158	59	40
	3	0	0	0	0	0	0	0	0
	4	120	107	101	86	148	119	213	162
	5	6	6	13	5	19	5	63	11
	6	0	0	0	1	0	1	0	0
	7	16	15	21	15	19	12	9	2
	8	3	4	11	11	9	6	32	7
	9			1		1		0	
	10							3	
Medium Sample Size Disparity	1	705	705	655	717	626	746	543	780
	2	166	177	177	151	141	136	68	50
	3	0	0	0	0	0	0	0	0
	4	99	94	122	101	161	99	237	153
	5	11	7	9	5	37	11	88	9
	6	0	0	0	0	0	0	0	1
	7	16	15	21	16	27	7	14	2
	8	3	2	16	10	8	1	49	5
	9							1	
Large Sample Size Disparity	1	718	734	681	741	597	747	494	833
	2	163	151	150	140	148	130	66	32
	3	0	0	0	0	0	0	0	0
	4	93	95	125	100	177	105	248	124
	5	9	7	16	8	32	9	108	5
	6	0	0	1	0	0	1	0	1
	7	9	9	20	6	27	2	17	0
	8	8	4	7	4	17	6	56	4
	9				1	2		4	1
	10							7	

PARTITION III, MEAN VARIABILITY (SMALL)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	910	904	885	901	868	931	846	948
	2	63	72	84	73	77	41	65	27
	3	0	0	0	0	0	0	0	0
	4	24	23	27	24	31	23	46	20
	5	3	1	3	1	24	4	40	5
	6	0	0	0	0	0	0	0	0
	7	0	0	1	1	0	1	1	0
	8							2	
Medium Sample Size Disparity	1	908	907	905	923	853	940	787	937
	2	67	69	60	52	72	42	62	30
	3	0	0	0	0	0	0	0	0
	4	20	18	22	17	49	14	75	27
	5	2	3	11	7	26	4	69	5
	6	0	0	0	0	0	0	0	0
	7	2	2	1	1	0	0	1	1
	8	1	1	1	0	0	0	6	0
Large Sample Size Disparity	1	923	924	878	937	924	937	699	937
	2	53	54	75	36	87	40	78	31
	3	0	0	0	0	0	0	0	0
	4	19	17	33	23	38	15	108	26
	5	5	4	10	3	45	6	100	5
	6	0	0	0	0	0	0	0	0
	7	0	1	3	0	4	1	2	0
	8	0	0	1	1	1	0	7	1
	9					1	1	2	
	10							4	

PARTITION III, MEAN VARIABILITY (MEDIUM)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
	1	663	679	660	713	651	763	622	818
	2	189	185	162	153	135	109	55	39
	3	0	0	0	0	0	0	0	0
Small Sample Size Disparity	4	108	96	134	103	131	99	181	117
	5	23	23	29	20	65	19	110	14
	6	0	0	0	0	0	0	0	2
	7	11	9	11	6	11	7	11	3
	8	6	8	3	4	6	1	21	5
	9			1	1	1	1	0	2
	10						1		
	1	736	732	652	737	630	799	584	851
	2	141	155	167	131	127	102	57	31
	3	0	0	0	0	0	0	0	0
Medium Sample Size Disparity	4	88	81	120	94	153	78	193	100
	5	22	20	42	27	75	18	131	14
	6	0	0	0	0	0	0	0	1
	7	8	5	12	5	12	2	8	1
	8	5	6	5	5	3	1	25	1
	9	0	1	2	1	0	0	1	0
	10							1	1
	1	726	738	676	771	635	830	502	848
	2	149	145	152	113	125	82	60	31
	3	0	0	0	0	0	0	0	0
Large Sample Size Disparity	4	102	87	112	81	130	62	191	99
	5	12	16	47	26	85	17	176	15
	6	0	0	0	0	0	1	0	3
	7	3	9	9	7	18	6	17	1
	8	8	5	4	1	4	2	49	3
	9	0	0	0	1	3	0	1	0
	10							4	

PARTITION III, MEAN VARIABILITY (LARGE)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	257	267	280	336	293	434	292	499
	2	234	229	204	197	177	193	43	51
	3	0	0	0	0	0	0	0	0
	4	313	330	290	285	309	252	327	373
	5	86	73	109	80	143	56	221	42
	6	0	0	0	0	0	0	0	1
	7	73	67	77	67	53	39	37	9
	8	32	31	39	33	21	20	77	24
	9	5	3	1	2	2	5	1	1
	10					2	1	2	0
Medium Sample Size Disparity	1	321	333	270	375	284	505	263	602
	2	223	233	198	203	144	154	41	50
	3	0	0	0	0	0	0	0	0
	4	278	272	313	273	274	226	305	243
	5	85	88	99	66	177	52	240	52
	6	0	0	0	0	0	0	0	1
	7	64	52	69	47	70	33	26	6
	8	24	20	45	31	44	27	113	44
	9	4	1	6	5	5	2	3	2
	10	1	1	0	0	2	1	9	0
Large Sample Size Disparity	1	396	425	329	450	313	588	224	667
	2	220	217	179	183	144	162	43	51
	3	0	0	0	0	0	0	0	0
	4	225	231	292	251	280	179	276	209
	5	75	61	102	54	156	40	264	40
	6	0	0	0	0	1	1	0	4
	7	56	40	54	41	60	22	35	7
	8	27	25	40	19	35	7	135	22
	9	1	1	4	2	7	0	6	0
	10	0	0	0	0	4	1	17	0

PARTITION IV, MEAN VARIABILITY (SMALL)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	845	842	856	885	849	889	785	667
	2	113	118	98	80	88	80	82	263
	3	0	0	0	0	0	0	0	0
	4	37	31	36	28	45	28	67	47
	5	3	4	7	5	15	2	53	10
	6	0	0	0	0	0	0	0	4
	7	1	4	2	1	2	1	11	4
	8	1	1	1	1	1	0	1	1
	9							1	4
	10								
Medium Sample Size Disparity	1	878	883	839	868	815	902	730	682
	2	83	82	104	88	103	70	93	258
	3	0	0	0	0	0	0	0	0
	4	33	27	46	38	50	22	88	48
	5	4	5	9	6	25	3	80	2
	6	0	0	0	0	0	0	0	8
	7	2	3	2	0	5	3	5	2
	8	0	0	0	0	2	0	3	0
	9							1	
Large Sample Size Disparity	1	894	898	863	907	786	897	641	657
	2	75	69	91	66	99	73	135	293
	3	0	0	0	0	0	0	0	0
	4	26	30	36	21	66	25	102	36
	5	4	2	9	5	42	3	94	3
	6	0	0	0	0	0	0	0	6
	7	1	1	1	1	4	1	10	2
	8	0	0	0	0	2	1	11	0
	9					1		7	3

PARTITION IV, MEAN VARIABILITY (MEDIUM)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	401	415	369	406	388	442	326	6
	2	287	283	275	293	263	297	283	541
	3	0	0	0	0	0	0	0	1
	4	213	213	237	212	231	190	190	279
	5	47	44	58	39	65	37	102	34
	6	0	0	0	0	0	0	0	84
	7	32	27	47	39	38	23	76	5
	8	15	11	13	11	13	8	18	0
	9	4	6	1	0	2	2	5	44
	10	1	1	0	0	0	1	0	6
Medium Sample Size Disparity	1	427	444	392	456	367	476	295	5
	2	274	263	250	296	242	291	273	654
	3	0	0	0	0	0	0	0	0
	4	202	209	238	170	234	166	198	193
	5	46	44	55	33	77	32	112	29
	6	0	0	0	0	0	0	0	78
	7	40	27	47	34	54	25	83	6
	8	10	12	15	10	20	8	29	0
	9	1	1	3	1	4	2	10	32
	10					2			3
Large Sample Size Disparity	1	493	505	391	496	372	535	234	7
	2	248	263	283	275	241	281	237	667
	3	0	0	0	0	0	0	0	0
	4	179	168	197	168	222	140	201	195
	5	49	38	63	26	84	21	146	12
	6	0	0	0	1	0	0	0	72
	7	22	17	47	26	54	17	119	3
	8	7	7	17	8	17	3	44	1
	9	1	1	2	0	7	2	15	39
	10	1	1			3	1	4	3

PARTITION IV, MEAN VARIABILITY (LARGE)

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
	1	38	46	36	51	27	36	27	0
	2	121	129	146	170	127	138	135	96
	3	0	0	0	0	0	0	0	0
Small Sample Size Disparity	4	314	335	306	323	298	329	248	299
	5	155	144	145	137	160	182	150	192
	6	0	0	0	1	0	2	0	144
	7	233	218	216	197	229	171	284	1
	8	83	80	78	62	92	67	82	0
	9	42	36	61	52	52	65	68	221
	10	14	12	12	7	15	10	6	47
	1	47	56	39	68	45	83	24	0
	2	159	164	146	173	124	189	95	148
	3	0	0	0	0	0	0	0	0
Medium Sample Size Disparity	4	294	326	271	301	271	339	222	327
	5	144	138	159	158	155	170	122	158
	6	0	1	0	1	0	3	0	123
	7	216	187	194	160	234	113	295	1
	8	89	78	97	73	91	54	126	0
	9	42	40	76	57	68	45	93	199
	10	9	10	18	9	12	4	22	43
	11							1	1
	1	67	84	57	100	38	112	25	0
	2	166	189	139	201	126	265	103	250
	3	0	0	0	0	0	0	0	0
Large Sample Size Disparity	4	316	332	305	332	288	329	193	322
	5	163	153	146	137	165	115	113	128
	6	1	1	0	3	0	8	1	125
	7	172	145	196	138	196	90	245	3
	8	72	60	83	49	79	30	141	0
	9	37	27	60	33	79	39	121	148
	10	6	9	14	7	29	12	55	23
	11							3	1

PARTITION V

INVERSE PAIRING

		Coefficient of Variation							
		0		.2		.6		1.0	
	Pattern Type	K	GH	K	GH	K	GH	K	GH
Small Sample Size Disparity	1	949	954	926	943	931	959	881	957
	2	39	33	53	41	51	30	58	25
	3	0	0	0	0	0	0	0	0
	4	12	13	15	13	12	8	22	14
	5	0	0	5	2	6	3	39	3
	6	0	0	0	0	0	0	0	1
	7	0	0	1	1	0	0	0	0
	8	0	0	0	0	0	0	0	0
Medium Sample Size Disparity	1	948	946	927	947	899	958	840	957
	2	43	45	53	38	57	33	71	16
	3	0	0	0	0	0	0	0	0
	4	6	7	18	14	30	6	34	19
	5	2	2	2	1	12	2	44	8
	6	0	0	0	0	0	0	0	0
	7	1	0	0	0	1	1	0	0
	8	0	0	0	0	0	0	3	0
	9					0		4	0
	10					1		4	0
Large Sample Size Disparity	1	939	945	927	952	861	947	778	973
	2	49	41	51	36	69	39	85	12
	3	0	0	0	0	0	0	0	1
	4	11	11	17	11	44	10	70	12
	5	1	3	5	1	24	4	60	3
	6	0	0	0	0	0	0	0	0
	7	0	0	0	0	1	0	3	0
	8	0	0	0	0	1	0	3	0
	9							0	
	10							1	